We saw in Chapter 8 that if the true value of ρ is less than 1 in absolute value, then

$$\sqrt{T}(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, (1 - \rho^2)). \qquad [17.1.3]$$

If [17.1.3] were also valid for the case when $\rho = 1$, it would seem to claim that $\sqrt{T}(\hat{\rho}_T - \rho)$ has zero variance, or that the distribution collapses to a point mass at zero:

$$\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{p} 0. \qquad [17.1.4]$$

As we shall see shortly, [17.1.4] is indeed a valid statement for unit root processes, but it obviously is not very helpful for hypothesis tests. To obtain a nondegenerate asymptotic distribution for $\hat{\rho}_T$ in the unit root case, it turns out that we have to multiply $\hat{\rho}_T$ by T rather than by $\sqrt{T}$. Thus, the unit root coefficient converges at a faster rate (T) than a coefficient for a stationary regression (which converges at $\sqrt{T}$), but at a slower rate than the coefficient on a time trend in the regressions analyzed in the previous chapter (which converged at $T^{3/2}$).

To get a better sense of why scaling by T is necessary when the true value of ρ is unity, recall that the difference between the estimate $\hat{\rho}_T$ and the true value can be expressed as in equation [8.2.3]:[1]

$$(\hat{\rho}_T - 1) = \frac{\sum_{t=1}^{T} y_{t-1}u_t}{\sum_{t=1}^{T} y_{t-1}^2}, \qquad [17.1.5]$$

so that

$$T(\hat{\rho}_T - 1) = \frac{(1/T)\sum_{t=1}^{T} y_{t-1}u_t}{(1/T^2)\sum_{t=1}^{T} y_{t-1}^2}. \qquad [17.1.6]$$

Consider first the numerator in [17.1.6]. When the true value of ρ is unity, equation [17.1.1] describes a random walk with

$$y_t = u_t + u_{t-1} + \cdots + u_1, \qquad [17.1.7]$$

since $y_0 = 0$. It follows from [17.1.7] that

$$y_t \sim N(0, \sigma^2 t). \qquad [17.1.8]$$

Note further that for a random walk,

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2,$$

implying that

$$y_{t-1}u_t = (1/2)\{y_t^2 - y_{t-1}^2 - u_t^2\}. \qquad [17.1.9]$$

If [17.1.9] is summed over $t = 1, 2, \ldots, T$, the result is

$$\sum_{t=1}^{T} y_{t-1}u_t = (1/2)\{y_T^2 - y_0^2\} - (1/2)\sum_{t=1}^{T} u_t^2. \qquad [17.1.10]$$

Recalling that $y_0 = 0$, equation [17.1.10] establishes that

$$(1/T)\sum_{t=1}^{T} y_{t-1}u_t = (1/2)\cdot(1/T)y_T^2 - (1/2)\cdot(1/T)\sum_{t=1}^{T} u_t^2, \qquad [17.1.11]$$

[1] This discussion is based on Fuller (1976, p. 369).

and if each side of [17.1.11] is divided by σ^2, the result is

$$\left(\frac{1}{\sigma^2 T}\right)\sum_{t=1}^{T} y_{t-1}u_t = \left(\frac{1}{2}\right)\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2 - \left(\frac{1}{2\sigma^2}\right)\left(\frac{1}{T}\right)\cdot\sum_{t=1}^{T} u_t^2. \qquad [17.1.12]$$

But [17.1.8] implies that $y_T/(\sigma\sqrt{T})$ is a $N(0, 1)$ variable, so that its square is $\chi^2(1)$:

$$[y_T/(\sigma\sqrt{T})]^2 \sim \chi^2(1). \qquad [17.1.13]$$

Also, $\Sigma_{t=1}^{T} u_t^2$ is the sum of T i.i.d. random variables, each with mean σ^2, and so, by the law of large numbers,

$$(1/T)\cdot\sum_{t=1}^{T} u_t^2 \xrightarrow{p} \sigma^2. \qquad [17.1.14]$$

Using [17.1.13] and [17.1.14], it follows from [17.1.12] that

$$[1/(\sigma^2 T)]\sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{L} (1/2)\cdot(X - 1), \qquad [17.1.15]$$

where $X \sim \chi^2(1)$.

Turning next to the denominator of [17.1.6], consider

$$\sum_{t=1}^{T} y_{t-1}^2. \qquad [17.1.16]$$

Recall from [17.1.8] that $y_{t-1} \sim N(0, \sigma^2(t-1))$, so $E(y_{t-1}^2) = \sigma^2(t-1)$. Consider the mean of [17.1.16],

$$E\left[\sum_{t=1}^{T} y_{t-1}^2\right] = \sigma^2\sum_{t=1}^{T}(t-1) = \sigma^2(T-1)T/2.$$

In order to construct a random variable that could have a convergent distribution, the quantity in [17.1.16] will have to be divided by T^2, as was done in the denominator of [17.1.6].

To summarize, if the true process is a random walk, then the deviation of the *OLS* estimate from the true value $(\hat{\rho}_T - 1)$ must be multiplied by T rather than $\sqrt{T}$ to obtain a variable with a useful asymptotic distribution. Moreover, this asymptotic distribution is not the usual Gaussian distribution but instead is a ratio involving a $\chi^2(1)$ variable in the numerator and a separate, nonstandard distribution in the denominator.

The asymptotic distribution of $T(\hat{\rho}_T - 1)$ will be fully characterized in Section 17.4. In preparation for this, the idea of Brownian motion is introduced in Section 17.2, followed by a discussion of the functional central limit theorem in Section 17.3.

17.2. *Brownian Motion*

Consider a random walk,

$$y_t = y_{t-1} + \varepsilon_t, \qquad [17.2.1]$$

in which the innovations are standard Normal variables:

$$\varepsilon_t \sim \text{i.i.d } N(0, 1).$$

If the process is started with $y_0 = 0$, then it follows as in [17.1.7] and [17.1.8] that

$$y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$$
$$y_t \sim N(0, t).$$

Moreover, the change in the value of y between dates t and s,

$$y_s - y_t = \varepsilon_{t+1} + \varepsilon_{t+2} + \cdots + \varepsilon_s,$$

is itself $N(0, (s - t))$ and is independent of the change between dates r and q for any dates $t < s < r < q$.

Consider the change between y_{t-1} and y_t. This innovation ε_t was taken to be $N(0, 1)$. Suppose we view ε_t as the sum of two independent Gaussian variables:

$$\varepsilon_t = e_{1t} + e_{2t},$$

with $e_{it} \sim$ i.i.d. $N(0, \frac{1}{2})$. We might then associate e_{1t} with the change between y_{t-1} and the value of y at some interim point (say, $y_{t-(1/2)}$),

$$y_{t-(1/2)} - y_{t-1} = e_{1t}, \tag{17.2.2}$$

and e_{2t} with the change between $y_{t-(1/2)}$ and y_t:

$$y_t - y_{t-(1/2)} = e_{2t}. \tag{17.2.3}$$

Sampled at integer dates $t = 1, 2, \ldots$, the process of [17.2.2] and [17.2.3] will have exactly the same properties as [17.2.1], since

$$y_t - y_{t-1} = e_{1t} + e_{2t} \sim \text{i.i.d. } N(0, 1).$$

In addition, the process of [17.2.2] and [17.2.3] is defined also at the noninteger dates $\{t + \frac{1}{2}\}_{t=0}^{\infty}$ and retains the property for both integer and noninteger dates that $y_s - y_t \sim N(0, s - t)$ with $y_s - y_t$ independent of the change over any other nonoverlapping interval.

By the same reasoning, we could imagine partitioning the change between $t - 1$ and t into N separate subperiods:

$$y_t - y_{t-1} = e_{1t} + e_{2t} + \cdots + e_{Nt},$$

with $e_{it} \sim$ i.i.d. $N(0, 1/N)$. The result would be a process with all the same properties as [17.2.1], defined at a finer and finer grid of dates as we increase N. The limit as $N \to \infty$ is a *continuous-time* process known as *standard Brownian motion*. The value of this process at date t is denoted $W(t)$.[2] A continuous-time process is a random variable that takes on a value for any nonnegative real number t, as distinct from a discrete-time process, which is only defined at integer values of t. To emphasize the distinction, we will put the date in parentheses when describing the value of a continuous-time variable at date t (as in $W(t)$) and use subscripts for a discrete-time variable (as in y_t). A discrete-time process was represented as a countable sequence of random variables, denoted $\{y_t\}_{t=1}^{\infty}$. A realization of a continuous-time process can be viewed as a stochastic function, denoted $W(\cdot)$, where $W: t \in [0, \infty) \to \mathbb{R}^1$.

A particular realization of Brownian motion turns out to be a continuous function of t. To see why it would be continuous, recall that the change between t and $t + \Delta$ is distributed $N(0, \Delta)$. Such a change is essentially certain to be arbitrarily small as the interval Δ goes to zero.

Definition: *Standard Brownian motion $W(\cdot)$ is a continuous-time stochastic process, associating each date $t \in [0, 1]$ with the scalar $W(t)$ such that:*

(*a*) $W(0) = 0$;

[2]Brownian motion is sometimes also referred to as a *Wiener process*.

(*b*) *For any dates* $0 \leq t_1 < t_2 < \cdots < t_k \leq 1$, *the changes* $[W(t_2) - W(t_1)]$, $[W(t_3) - W(t_2)], \ldots, [W(t_k) - W(t_{k-1})]$ *are independent multivariate Gaussian with* $[W(s) - W(t)] \sim N(0, s - t)$;

(*c*) *For any given realization,* $W(t)$ *is continuous in* t *with probability 1.*

There are advantages to restricting the analysis to dates t within a closed interval. All of the results in this text relate to the behavior of Brownian motion for dates within the unit interval ($t \in [0, 1]$), and in anticipation of this we have simply defined $W(\cdot)$ to be a function mapping $t \in [0, 1]$ into $\mathbb{R}^1$.

Other continuous-time processes can be generated from standard Brownian motion. For example, the process

$$Z(t) = \sigma \cdot W(t)$$

has independent increments and is distributed $N(0, \sigma^2 t)$ across realizations. Such a process is described as *Brownian motion with variance* σ^2. Thus, standard Brownian motion could also be described as Brownian motion with unit variance.

As another example,

$$Z(t) = [W(t)]^2 \qquad [17.2.4]$$

would be distributed as t times a $\chi^2(1)$ variable across realizations.

Although $W(t)$ is continuous in t, it cannot be differentiated using standard calculus; the direction of change at t is likely to be completely different from that at $t + \Delta$, no matter how small we make Δ.[3]

17.3. *The Functional Central Limit Theorem*

One of the uses of Brownian motion is to permit more general statements of the central limit theorem than those in Chapter 7. Recall the simplest version of the central limit theorem: if $u_t \sim$ i.i.d. with mean zero and variance σ^2, then the sample mean $\bar{u}_T \equiv (1/T)\Sigma_{t=1}^T u_t$ satisfies

$$\sqrt{T}\bar{u}_T \xrightarrow{L} N(0, \sigma^2).$$

Consider now an estimator based on the following principle: When given a sample of size T, we calculate the mean of the first half of the sample and throw out the rest of the observations:

$$\bar{u}_{[T/2]^*} = (1/[T/2]^*) \sum_{t=1}^{[T/2]^*} u_t.$$

Here $[T/2]^*$ denotes the largest integer that is less than or equal to $T/2$; that is, $[T/2]^* = T/2$ for T even and $[T/2]^* = (T - 1)/2$ for T odd. This strange estimator would also satisfy the central limit theorem:

$$\sqrt{[T/2]^*}\,\bar{u}_{[T/2]^*} \xrightarrow[T\to\infty]{L} N(0, \sigma^2). \qquad [17.3.1]$$

Moreover, this estimator would be independent of an estimator that uses only the second half of the sample.

More generally, we can construct a variable $X_T(r)$ from the sample mean of

[3]For an introduction to differentiation and integration of Brownian motion, see Malliaris and Brock (1982, Chapter 2).

the first rth fraction of observations, $r \in [0, 1]$, defined by

$$X_T(r) \equiv (1/T) \sum_{t=1}^{[Tr]^*} u_t. \qquad [17.3.2]$$

For any given realization, $X_T(r)$ is a step function in r, with

$$X_T(r) = \begin{cases} 0 & \text{for } 0 \le r < 1/T \\ u_1/T & \text{for } 1/T \le r < 2/T \\ (u_1 + u_2)/T & \text{for } 2/T \le r < 3/T \\ \vdots & \\ (u_1 + u_2 + \cdots + u_T)/T & \text{for } r = 1. \end{cases} \qquad [17.3.3]$$

Then

$$\sqrt{T}\cdot X_T(r) = (1/\sqrt{T}) \sum_{t=1}^{[Tr]^*} u_t = (\sqrt{[Tr]^*}/\sqrt{T})(1/\sqrt{[Tr]^*}) \sum_{t=1}^{[Tr]^*} u_t. \qquad [17.3.4]$$

But

$$(1/\sqrt{[Tr]^*}) \sum_{t=1}^{[Tr]^*} u_t \xrightarrow{L} N(0, \sigma^2),$$

by the central limit theorem as in [17.3.1], while $(\sqrt{[Tr]^*}/\sqrt{T}) \to \sqrt{r}$. Hence, the asymptotic distribution of $\sqrt{T}\cdot X_T(r)$ in [17.3.4] is that of $\sqrt{r}$ times a $N(0, \sigma^2)$ random variable, or

$$\sqrt{T}\cdot X_T(r) \xrightarrow{L} N(0, r\sigma^2)$$

and

$$\sqrt{T}\cdot [X_T(r)/\sigma] \xrightarrow{L} N(0, r). \qquad [17.3.5]$$

If we were similarly to consider the behavior of a sample mean based on observations $[Tr_1]^*$ through $[Tr_2]^*$ for $r_2 > r_1$, we would conclude that this too is asymptotically Normal,

$$\sqrt{T}\cdot [X_T(r_2) - X_T(r_1)]/\sigma \xrightarrow{L} N(0, r_2 - r_1),$$

and is independent of the estimator in [17.3.5], provided that $r < r_1$. It thus should not be surprising that the sequence of stochastic functions $\{\sqrt{T}\cdot X_T(\cdot)/\sigma\}_{T=1}^{\infty}$ has an asymptotic probability law that is described by standard Brownian motion $W(\cdot)$:

$$\sqrt{T}\cdot X_T(\cdot)/\sigma \xrightarrow{L} W(\cdot). \qquad [17.3.6]$$

Note the difference between the claims in [17.3.5] and [17.3.6]. The expression $X_T(\cdot)$ denotes a random function while $X_T(r)$ denotes the value that function assumes at date r; thus, $X_T(\cdot)$ is a function, while $X_T(r)$ is a random variable.

Result [17.3.6] is known as the *functional central limit theorem*. The derivation here assumed that u_t was i.i.d. A more general statement will be provided in Section 17.5.

Evaluated at $r = 1$, the function $X_T(r)$ in [17.3.2] is just the sample mean:

$$X_T(1) = (1/T) \sum_{t=1}^{T} u_t.$$

Thus, when the functions in [17.3.6] are evaluated at $r = 1$, the conventional central limit theorem [7.1.6] obtains as a special case of [17.3.6]:

$$\sqrt{T}X_T(1)/\sigma = [1/(\sigma\sqrt{T})] \sum_{t=1}^{T} u_t \xrightarrow{L} W(1) \sim N(0, 1). \qquad [17.3.7]$$

We earlier defined convergence in law for random variables, and now we need to extend the definition to cover random functions. Let $S(\cdot)$ represent a continuous-time stochastic process with $S(r)$ representing its value at some date r for $r \in [0, 1]$. Suppose, further, that for any given realization, $S(\cdot)$ is a continuous function of r with probability 1. For $\{S_T(\cdot)\}_{T=1}^{\infty}$ a sequence of such continuous functions, we say that $S_T(\cdot) \xrightarrow{L} S(\cdot)$ if all of the following hold:[4]

(a) For any finite collection of k particular dates,

$$0 \leq r_1 < r_2 < \cdots < r_k \leq 1,$$

the sequence of k-dimensional random vectors $\{\mathbf{y}_T\}_{T=1}^{\infty}$ converges in distribution to the vector $\mathbf{y}$, where

$$\mathbf{y}_T \equiv \begin{bmatrix} S_T(r_1) \\ S_T(r_2) \\ \vdots \\ S_T(r_k) \end{bmatrix} \qquad \mathbf{y} \equiv \begin{bmatrix} S(r_1) \\ S(r_2) \\ \vdots \\ S(r_k) \end{bmatrix};$$

(b) For each $\varepsilon > 0$, the probability that $S_T(r_1)$ differs from $S_T(r_2)$ for any dates r_1 and r_2 within δ of each other goes to zero uniformly in T as $\delta \to 0$;

(c) $P\{|S_T(0)| > \lambda\} \to 0$ uniformly in T as $\lambda \to \infty$.

This definition applies to sequences of continuous functions, though the function in [17.3.2] is a discontinuous step function. Fortunately, the discontinuities occur at a countable set of points. Formally, $S_T(\cdot)$ can be replaced with a similar continuous function, interpolating between the steps (as in Hall and Heyde, 1980). Alternatively, the definition of convergence of random functions can be generalized to allow for discontinuities of the type in [17.3.2] (as in Chapter 3 of Billingsley, 1968).

It will also be helpful to extend the earlier definition of convergence in probability to sequences of random functions. Let $\{S_T(\cdot)\}_{T=1}^{\infty}$ and $\{V_T(\cdot)\}_{T=1}^{\infty}$ denote sequences of random continuous functions with S_T: $r \in [0, 1] \to \mathbb{R}^1$ and V_T: $r \in [0, 1] \to \mathbb{R}^1$. Let the scalar Y_T represent the largest amount by which $S_T(r)$ differs from $V_T(r)$ for any r:

$$Y_T \equiv \sup_{r \in [0,1]} |S_T(r) - V_T(r)|.$$

Thus, $\{Y_T\}_{T=1}^{\infty}$ is a sequence of random variables, and we could talk about its probability limit using the standard definition given in [7.1.2]. If the sequence of scalars $\{Y_T\}_{T=1}^{\infty}$ converges in probability to zero, then we say that the sequence of functions $S_T(\cdot)$ converges in probability to $V_T(\cdot)$. That is, the expression

$$S_T(\cdot) \xrightarrow{p} V_T(\cdot)$$

is interpreted to mean that

$$\sup_{r \in [0,1]} |S_T(r) - V_T(r)| \xrightarrow{p} 0.$$

With this definition, result (a) of Proposition 7.3 can be generalized to apply

[4]The sequence of probability measures induced by $\{S_T(\cdot)\}_{T=1}^{\infty}$ weakly converges (in the sense of Billingsley, 1968) to the probability measure induced by $S(\cdot)$ if and only if conditions (a) through (c) hold; see Theorem A.2, p. 275, in Hall and Heyde (1980).

to sequences of functions. Specifically, if $\{S_T(\cdot)\}_{T=1}^{\infty}$ and $\{V_T(\cdot)\}_{T=1}^{\infty}$ are sequences of continuous functions with $V_T(\cdot) \xrightarrow{p} S_T(\cdot)$ and $S_T(\cdot) \xrightarrow{L} S(\cdot)$ for $S(\cdot)$ a continuous function, then $V_T(\cdot) \xrightarrow{L} S(\cdot)$; see, for example, Stinchcombe and White (1993).

Example 17.1

Let $\{x_T\}_{T=1}^{\infty}$ be a sequence of random scalars with $x_T \xrightarrow{p} 0$, and let $\{S_T(\cdot)\}_{T=1}^{\infty}$ be a sequence of random continuous functions, $S_T: r \in [0, 1] \rightarrow \mathbb{R}^1$ with $S_T(\cdot) \xrightarrow{L} S(\cdot)$. Then the sequence of functions $\{V_T(\cdot)\}_{T=1}^{\infty}$ defined by $V_T(r) \equiv S_T(r) + x_T$ has the property that $V_T(\cdot) \xrightarrow{L} S(\cdot)$. To see this, note that $V_T(r) - S_T(r) = x_T$ for all r, so that

$$\sup_{r\in[0,1]} |S_T(r) - V_T(r)| = |x_T|,$$

which converges in probability to zero. Hence, $V_T(\cdot) \xrightarrow{p} S_T(\cdot)$, and therefore $V_T(\cdot) \xrightarrow{L} S(\cdot)$.

Example 17.2

Let η_t be a strictly stationary time series with finite fourth moment, and let $S_T(r) = (1/\sqrt{T})\cdot\eta_{[Tr]^*}$. Then $S_T(\cdot) \xrightarrow{p} 0$. To see this, note that

$$\begin{aligned}
&P\left\{\sup_{r\in[0,1]} |S_T(r)| > \delta\right\} \\
&\quad = P\{[|(1/\sqrt{T})\cdot\eta_1| > \delta] \quad \text{or} \quad [|(1/\sqrt{T})\cdot\eta_2| > \delta] \quad \text{or} \quad \cdots \\
&\qquad \text{or} \quad [|(1/\sqrt{T})\cdot\eta_T| > \delta]\} \\
&\quad \le T\cdot P\{|(1/\sqrt{T})\cdot\eta_t| > \delta]\} \\
&\quad \le T\cdot\frac{E\{(1/\sqrt{T})\cdot\eta_t\}^4}{\delta^4} \\
&\quad = \frac{E(\eta_t^4)}{T\delta^4},
\end{aligned}$$

where the next-to-last line follows from Chebyshev's inequality. Since $E(\eta_t^4)$ is finite, this probability goes to zero as $T \rightarrow \infty$, establishing that $S_T(\cdot) \xrightarrow{p} 0$, as claimed.

Continuous Mapping Theorem

In Chapter 7 we saw that if $\{x_T\}_{T=1}^{\infty}$ is a sequence of random variables with $x_T \xrightarrow{L} x$ and if $g: \mathbb{R}^1 \rightarrow \mathbb{R}^1$ is a continuous function, then $g(x_T) \xrightarrow{L} g(x)$. A similar result holds for sequences of random functions. Here, the analog to the function $g(\cdot)$ is a continuous *functional*, which could associate a real random variable y with the stochastic function $S(\cdot)$. For example, $y = \int_0^1 S(r)\,dr$ and $y = \int_0^1 [S(r)]^2\,dr$ represent continuous functionals.[5] The *continuous mapping theorem*[6] states that if $S_T(\cdot) \xrightarrow{L} S(\cdot)$ and $g(\cdot)$ is a continuous functional, then $g(S_T(\cdot)) \xrightarrow{L} g(S(\cdot))$.

[5]Continuity of a functional $g(\cdot)$ in this context means that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $h(r)$ and $k(r)$ are any continuous bounded functions on $[0, 1]$, $h: [0, 1] \rightarrow \mathbb{R}^1$ and $k: [0, 1] \rightarrow \mathbb{R}^1$, such that $|h(r) - k(r)| < \delta$ for all $r \in [0, 1]$, then

$$|g[h(\cdot)] - g[k(\cdot)]| < \varepsilon.$$

[6]See, for example, Theorem A.3 on p. 276 in Hall and Heyde (1980).

The continuous mapping theorem also applies to a continuous functional $g(\cdot)$ that maps a continuous bounded function on [0, 1] into another continuous bounded function on [0, 1]. For example, the function whose value at r is a positive constant σ times $h(r)$ represents the result of applying the continuous functional $g[h(\cdot)] \equiv \sigma \cdot h(\cdot)$ to $h(\cdot)$.[7] Thus, it follows from [17.3.6] that

$$\sqrt{T}\cdot X_T(\cdot) \xrightarrow{L} \sigma \cdot W(\cdot). \quad [17.3.8]$$

Recalling that $W(r) \sim N(0, r)$, result [17.3.8] implies that $\sqrt{T}\cdot X_T(r) \approx N(0, \sigma^2 r)$.

As another example, consider the function $S_T(\cdot)$ whose value at r is given by

$$S_T(r) \equiv [\sqrt{T}\cdot X_T(r)]^2. \quad [17.3.9]$$

Since $\sqrt{T}\cdot X_T(\cdot) \xrightarrow{L} \sigma \cdot W(\cdot)$, it follows that

$$S_T(\cdot) \xrightarrow{L} \sigma^2[W(\cdot)]^2. \quad [17.3.10]$$

In other words, if the value $W(r)$ from a realization of standard Brownian motion at every date r is squared and then multiplied by σ^2, the resulting continuous-time process would follow essentially the same probability law as does the continuous-time process defined by $S_T(r)$ in [17.3.9] for T sufficiently large.

Applications to Unit Root Processes

The use of the functional central limit theorem to calculate the asymptotic distribution of statistics constructed from unit root processes was pioneered by Phillips (1986, 1987).[8] The simplest illustration of Phillips's approach is provided by a random walk,

$$y_t = y_{t-1} + u_t, \quad [17.3.11]$$

where $\{u_t\}$ is an i.i.d. sequence with mean zero and variance σ^2. If $y_0 = 0$, then [17.3.11] implies that

$$y_t = u_1 + u_2 + \cdots + u_t. \quad [17.3.12]$$

Equation [17.3.12] can be used to express the stochastic function $X_T(r)$ defined in [17.3.3] as

$$X_T(r) = \begin{cases} 0 & \text{for } 0 \le r < 1/T \\ y_1/T & \text{for } 1/T \le r < 2/T \\ y_2/T & \text{for } 2/T \le r < 3/T \\ \vdots & \\ y_T/T & \text{for } r = 1. \end{cases} \quad [17.3.13]$$

Figure 17.1 plots $X_T(r)$ as a function of r. Note that the area under this step function

[7]Here continuity of the functional $g(\cdot)$ means that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $h(r)$ and $k(r)$ are any continuous bounded functions on [0, 1], h: $[0, 1] \to \mathbb{R}^1$ and k: $[0, 1] \to \mathbb{R}^1$, such that $|h(r) - k(r)| < \delta$ for all $r \in [0, 1]$, then

$$|g[h(r)] - g[k(r)]| < \varepsilon$$

for all $r \in [0, 1]$.

[8]Result [17.4.7] in the next section for the case with i.i.d. errors was first derived by White (1958). Phillips (1986, 1987) developed the general derivation presented here based on the functional central limit theorem and the continuous mapping theorem. Other important contributions include Dickey and Fuller (1979), Chan and Wei (1988), Park and Phillips (1988, 1989), Sims, Stock, and Watson (1990), and Phillips and Solo (1992).

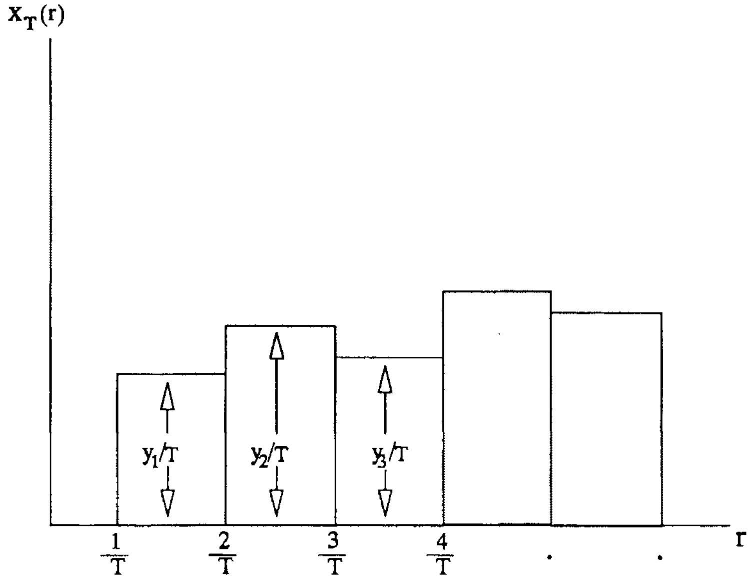

FIGURE 17.1 Plot of $X_T(r)$ as a function of r.

is the sum of T rectangles. The tth rectangle has width $1/T$ and height y_{t-1}/T, and therefore has area y_{t-1}/T^2. The integral of $X_T(r)$ is thus equivalent to

$$\int_0^1 X_T(r)\, dr = y_1/T^2 + y_2/T^2 + \cdots + y_{T-1}/T^2. \qquad [17.3.14]$$

Multiplying both sides of [17.3.14] by $\sqrt{T}$ establishes that

$$\int_0^1 \sqrt{T}\cdot X_T(r)\, dr = T^{-3/2} \sum_{t=1}^{T} y_{t-1}. \qquad [17.3.15]$$

But we know from [17.3.8] and the continuous mapping theorem that as $T \to \infty$,

$$\int_0^1 \sqrt{T}\cdot X_T(r)\, dr \xrightarrow{L} \sigma\cdot\int_0^1 W(r)\, dr,$$

implying from [17.3.15] that

$$T^{-3/2} \sum_{t=1}^{T} y_{t-1} \xrightarrow{L} \sigma\cdot\int_0^1 W(r)\, dr. \qquad [17.3.16]$$

It is also instructive to derive [17.3.16] from first principles. From [17.3.12], we can write

$$\begin{aligned} T^{-3/2} \sum_{t=1}^{T} y_{t-1} &= T^{-3/2}[u_1 + (u_1 + u_2) + (u_1 + u_2 + u_3) + \cdots \\ &\qquad + (u_1 + u_2 + u_3 + \cdots + u_{T-1})] \\ &= T^{-3/2}[(T-1)u_1 + (T-2)u_2 + (T-3)u_3 + \cdots \\ &\qquad + [T-(T-1)]u_{T-1}] \qquad [17.3.17] \\ &= T^{-3/2} \sum_{t=1}^{T} (T-t)u_t \\ &= T^{-1/2} \sum_{t=1}^{T} u_t - T^{-3/2} \sum_{t=1}^{T} tu_t. \end{aligned}$$

Recall from [16.1.24] that

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} u_t \\ T^{-3/2}\sum_{t=1}^{T} tu_t \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}\right). \qquad [17.3.18]$$

Thus, [17.3.17] implies that $T^{-3/2}\sum_{t=1}^{T} y_{t-1}$ is asymptotically Gaussian with mean zero and variance equal to

$$\sigma^2\{1 - 2\cdot(1/2) + 1/3\} = \sigma^2/3.$$

Evidently, $\sigma\cdot\int_0^1 W(r)\,dr$ in [17.3.16] describes a random variable that has a $N(0, \sigma^2/3)$ distribution.

Thus, if y_t is a driftless random walk, the sample mean $T^{-1}\sum_{t=1}^{T} y_t$ diverges but $T^{-3/2}\sum_{t=1}^{T} y_t$ converges to a Gaussian random variable whose distribution can be described as the integral of the realization of Brownian motion with variance σ^2.

Expression [17.3.17] also gives us a way to describe the asymptotic distribution of $T^{-3/2}\sum_{t=1}^{T} tu_t$ in terms of functionals on Brownian motion:

$$\begin{aligned} T^{-3/2}\sum_{t=1}^{T} tu_t &= T^{-1/2}\sum_{t=1}^{T} u_t - T^{-3/2}\sum_{t=1}^{T} y_{t-1} \\ &\xrightarrow{L} \sigma\cdot W(1) - \sigma\cdot\int_0^1 W(r)\,dr, \end{aligned} \qquad [17.3.19]$$

with the last line following from [17.3.7] and [17.3.16]. Recalling [17.3.18], the random variable on the right side of [17.3.19] evidently has a $N(0, \sigma^2/3)$ distribution.

A similar argument to that in [17.3.15] can be used to describe the asymptotic distribution of the sum of squares of a random walk. The statistic $S_T(r)$ defined in [17.3.9],

$$S_T(r) \equiv T\cdot[X_T(r)]^2, \qquad [17.3.20]$$

can be written using [17.3.13] as

$$S_T(r) = \begin{cases} 0 & \text{for } 0 \le r < 1/T \\ y_1^2/T & \text{for } 1/T \le r < 2/T \\ y_2^2/T & \text{for } 2/T \le r < 3/T \\ \vdots & \\ y_T^2/T & \text{for } r = 1. \end{cases} \qquad [17.3.21]$$

It follows that

$$\int_0^1 S_T(r)\,dr = y_1^2/T^2 + y_2^2/T^2 + \cdots + y_{T-1}^2/T^2.$$

Thus, from [17.3.10] and the continuous mapping theorem,

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{L} \sigma^2\cdot\int_0^1 [W(r)]^2\,dr. \qquad [17.3.22]$$

Two other useful results are

$$T^{-5/2}\sum_{t=1}^{T} ty_{t-1} = T^{-3/2}\sum_{t=1}^{T} (t/T)y_{t-1} \xrightarrow{L} \sigma\cdot\int_0^1 rW(r)\,dr \qquad [17.3.23]$$

for $r = t/T$ and

$$T^{-3}\sum_{t=1}^{T} t y_{t-1}^2 = T^{-2}\sum_{t=1}^{T}(t/T)y_{t-1}^2 \xrightarrow{L} \sigma^2\cdot\int_0^1 r\cdot[W(r)]^2\,dr. \qquad [17.3.24]$$

As yet another useful application, consider the statistic in [17.1.11]:

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t = (1/2)\cdot(1/T)y_T^2 - (1/2)\cdot(1/T)\sum_{t=1}^{T}u_t^2.$$

Recalling [17.3.21], this can be written

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t = (1/2)S_T(1) - (1/2)(1/T)\sum_{t=1}^{T}u_t^2. \qquad [17.3.25]$$

But $(1/T)\Sigma_{t=1}^{T}u_t^2 \xrightarrow{p} \sigma^2$, by the law of large numbers, and $S_T(1) \xrightarrow{L} \sigma^2[W(1)]^2$, by [17.3.10]. It thus follows from [17.3.25] that

$$T^{-1}\sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{L} (1/2)\sigma^2[W(1)]^2 - (1/2)\sigma^2. \qquad [17.3.26]$$

Recall that $W(1)$, the value of standard Brownian motion at date $r = 1$, has a $N(0, 1)$ distribution, meaning that $[W(1)]^2$ has a $\chi^2(1)$ distribution. Result [17.3.26] is therefore just another way to express the earlier result [17.1.15] using a functional on Brownian motion instead of the χ^2 distribution.

17.4. *Asymptotic Properties of a First-Order Autoregression when the True Coefficient Is Unity*

We are now in a position to calculate the asymptotic distribution of some simple regressions involving unit roots. For convenience, the results from Section 17.3 are collected in the form of a proposition.

Proposition 17.1: *Suppose that ξ_t follows a random walk without drift,*

$$\xi_t = \xi_{t-1} + u_t,$$

where $\xi_0 = 0$ and $\{u_t\}$ is an i.i.d. sequence with mean zero and variance σ^2. Then

(*a*) $T^{-1/2}\sum_{t=1}^{T} u_t \xrightarrow{L} \sigma\cdot W(1)$ [17.3.7];

(*b*) $T^{-1}\sum_{t=1}^{T} \xi_{t-1}u_t \xrightarrow{L} (1/2)\sigma^2\{[W(1)]^2 - 1\}$ [17.3.26];

(c) $T^{-3/2}\sum_{t=1}^{T} t u_t \xrightarrow{L} \sigma\cdot W(1) - \sigma\cdot\int_0^1 W(r)\,dr$ [17.3.19];

(*d*) $T^{-3/2}\sum_{t=1}^{T} \xi_{t-1} \xrightarrow{L} \sigma\cdot\int_0^1 W(r)\,dr$ [17.3.16];

(*e*) $T^{-2}\sum_{t=1}^{T} \xi_{t-1}^2 \xrightarrow{L} \sigma^2\cdot\int_0^1 [W(r)]^2\,dr$ [17.3.22];

(*f*) $T^{-5/2}\sum_{t=1}^{T} t\xi_{t-1} \xrightarrow{L} \sigma\cdot\int_0^1 rW(r)\,dr$ [17.3.23];

(*g*) $T^{-3}\sum_{t=1}^{T} t\xi_{t-1}^2 \xrightarrow{L} \sigma^2\cdot\int_0^1 r\cdot[W(r)]^2\,dr$ [17.3.24];

(*h*) $T^{-(v+1)}\sum_{t=1}^{T} t^v \rightarrow 1/(v+1)$ *for* $v = 0, 1, \ldots$ [16.1.15].

The expressions in brackets indicate where the stated result was earlier derived. Though the earlier derivations assumed that the initial value ξ_0 was equal to zero, the same results are obtained when ξ_0 is any fixed value or drawn from a specified distribution as in Phillips (1987).

The asymptotic distributions in Proposition 17.1 are all written in terms of functionals on standard Brownian motion, denoted $W(r)$. Note that this is the *same* Brownian motion $W(r)$ in each result (a) through (g), so that in general the magnitudes in Proposition 17.1 are all correlated. If we are not interested in capturing these correlations, then there are simpler ways to describe the asymptotic distributions. For example, we have seen that (a) is just a $N(0, \sigma^2)$ distribution, (b) is $(1/2)\sigma^2 \cdot [\chi^2(1) - 1]$, and (c) and (d) are $N(0, \sigma^2/3)$. Exercise 17.1 gives an example of one approach to calculating the covariances among random variables described by these functionals on Brownian motion.

Proposition 17.1 can be used to calculate the asymptotic distributions of statistics from a number of simple regressions involving unit roots. This section discusses several key cases.

Case 1. No Constant Term or Time Trend in the Regression; True Process Is a Random Walk

Consider first *OLS* estimation of ρ based on an $AR(1)$ regression,

$$y_t = \rho y_{t-1} + u_t, \qquad [17.4.1]$$

where u_t is i.i.d. with mean zero and variance σ^2. We are interested in the properties of the *OLS* estimate

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2} \qquad [17.4.2]$$

when the true value of ρ is unity. From [17.1.6], the deviation of the *OLS* estimate from the true value is characterized by

$$T(\hat{\rho}_T - 1) = \frac{T^{-1} \sum_{t=1}^{T} y_{t-1} u_t}{T^{-2} \sum_{t=1}^{T} y_{t-1}^2}. \qquad [17.4.3]$$

If the true value of ρ is unity, then

$$y_t = y_0 + u_1 + u_2 + \cdots + u_t. \qquad [17.4.4]$$

Apart from the initial term y_0 (which does not affect any of the asymptotic distributions), the variable y_t is the same as the quantity labeled ξ_t in Proposition 17.1. From result (b) of that proposition,

$$T^{-1} \sum_{t=1}^{T} y_{t-1} u_t \xrightarrow{L} (1/2)\sigma^2\{[W(1)]^2 - 1\}, \qquad [17.4.5]$$

while from result (e),

$$T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{L} \sigma^2 \cdot \int_0^1 [W(r)]^2\, dr. \qquad [17.4.6]$$

Since [17.4.3] is a continuous function of [17.4.5] and [17.4.6], it follows from Proposition 7.3(c) that under the null hypothesis that $\rho = 1$, the *OLS* estimate

$\hat{\rho}_T$ is characterized by

$$T(\hat{\rho}_T - 1) \xrightarrow{L} \frac{(1/2)\{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2\,dr}. \qquad [17.4.7]$$

Recall that $[W(1)]^2$ is a $\chi^2(1)$ variable. The probability that a $\chi^2(1)$ variable is less than unity is 0.68, and since the denominator of [17.4.7] must be positive, the probability that $\hat{\rho}_T - 1$ is negative approaches 0.68 as T becomes large. In other words, in two-thirds of the samples generated by a random walk, the estimate $\hat{\rho}_T$ will be less than the true value of unity. Moreover, in those samples for which $[W(1)]^2$ is large, the denominator of [17.4.7] will be large as well. The result is that the limiting distribution of $T(\hat{\rho}_T - 1)$ is skewed to the left.

Recall that in the stationary case when $|\rho| < 1$, the estimate $\hat{\rho}_T$ is downward-biased in small samples. Even so, in the stationary case the limiting distribution of $\sqrt{T}(\hat{\rho}_T - \rho)$ is symmetric around zero. By contrast, when the true value of ρ is unity, even the limiting distribution of $T(\hat{\rho}_T - 1)$ is asymmetric, with negative values twice as likely as positive values.

In practice, critical values for the random variable in [17.4.7] are found by calculating the exact small-sample distribution of $T(\hat{\rho}_T - 1)$ for given T, assuming that the innovations $\{u_t\}$ are Gaussian. This can be done either by Monte Carlo, as in the critical values reported in Fuller (1976), or by using exact numerical procedures described in Evans and Savin (1981). Sample percentiles for $T(\hat{\rho}_T - 1)$ are reported in the section labeled Case 1 in Table B.5 of Appendix B. For finite T, these are exact only under the assumption of Gaussian innovations. As T becomes large, these values also describe the asymptotic distribution for non-Gaussian innovations.

It follows from [17.4.7] that $\hat{\rho}_T$ is a superconsistent estimate of the true value ($\rho = 1$). This is easily seen by dividing [17.4.3] by $\sqrt{T}$:

$$\sqrt{T}(\hat{\rho}_T - 1) = \frac{T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}. \qquad [17.4.8]$$

From Proposition 17.1(b), the numerator in [17.4.8] converges to $T^{-1/2}(1/2)\sigma^2$ times $(X - 1)$, where X is a $\chi^2(1)$ random variable. Since a $\chi^2(1)$ variable has finite variance, the variance of the numerator in [17.4.8] is of order $1/T$, meaning that the numerator converges in probability to zero. Hence,

$$\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{p} 0.$$

Result [17.4.7] allows the point estimate $\hat{\rho}_T$ to be used by itself to test the null hypothesis of a unit root, without needing to calculate its standard error. Another popular statistic for testing the null hypothesis that $\rho = 1$ is based on the usual *OLS* t test of this hypothesis,

$$t_T = \frac{(\hat{\rho}_T - 1)}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{(\hat{\rho}_T - 1)}{\left\{s_T^2 \div \sum_{t=1}^{T} y_{t-1}^2\right\}^{1/2}}, \qquad [17.4.9]$$

where $\hat{\sigma}_{\hat{\rho}_T}$ is the usual *OLS* standard error for the estimated coefficient,

$$\hat{\sigma}_{\hat{\rho}_T} = \left\{s_T^2 \div \sum_{t=1}^{T} y_{t-1}^2\right\}^{1/2},$$

and s_T^2 denotes the OLS estimate of the residual variance:

$$s_T^2 = \sum_{t=1}^{T} (y_t - \hat{\rho}_T y_{t-1})^2/(T - 1).$$

Although the t statistic [17.4.9] is calculated in the usual way, it does not have a limiting Gaussian distribution when the true process is characterized by $\rho = 1$. To find the appropriate limiting distribution, note that [17.4.9] can equivalently be expressed as

$$t_T = T(\hat{\rho}_T - 1)\left\{T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right\}^{1/2} \div \{s_T^2\}^{1/2}, \qquad [17.4.10]$$

or, substituting from [17.4.3],

$$t_T = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}u_t}{\left\{T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right\}^{1/2}\{s_T^2\}^{1/2}}. \qquad [17.4.11]$$

As in Section 8.2, consistency of $\hat{\rho}_T$ implies $s_T^2 \xrightarrow{p} \sigma^2$. It follows from [17.4.5] and [17.4.6] that as $T \to \infty$,

$$t_T \xrightarrow{L} \frac{(1/2)\sigma^2\{[W(1)]^2 - 1\}}{\left\{\sigma^2\int_0^1 [W(r)]^2\,dr\right\}^{1/2}\{\sigma^2\}^{1/2}} = \frac{(1/2)\{[W(1)]^2 - 1\}}{\left\{\int_0^1 [W(r)]^2\,dr\right\}^{1/2}}. \qquad [17.4.12]$$

Statistical tables for the distribution of [17.4.11] for various sample sizes T are reported in the section labeled Case 1 in Table B.6; again, the small-sample results assume Gaussian innovations.

Example 17.3

The following $AR(1)$ process for the nominal three-month U.S. Treasury bill rate was fitted by OLS regression to quarterly data, t = 1947:II to 1989:I:

$$i_t = \underset{(0.010592)}{0.99694}\, i_{t-1}, \qquad [17.4.13]$$

with the standard error of $\hat{\rho}$ in parentheses. Here $T = 168$ and

$$T(\hat{\rho} - 1) = (168)(0.99694 - 1) = -0.51.$$

The distribution of this statistic was calculated in [17.4.7] under the assumption that the true value of ρ is unity. The null hypothesis is therefore that $\rho = 1$, and the alternative is that $\rho < 1$. From Table B.5, in a sample of this size, 95% of the time when there really is a unit root, the statistic $T(\hat{\rho} - 1)$ will be above -7.9. The observed value (-0.51) is well above this, and so the null hypothesis is accepted at the 5% level and we should conclude that these data might well be described by a random walk.

In order to have rejected the null hypothesis for a sample of this size, the estimated autoregressive coefficient $\hat{\rho}$ would have to be less than 0.95:

$$168(0.95 - 1) = -8.4.$$

The OLS t test of H_0: $\rho = 1$ is

$$t = (0.99694 - 1)/0.010592 = -0.29.$$

This is well above the 5% critical value from Table B.6 of -1.95, so the null hypothesis that the Treasury bill rate follows a random walk is also accepted by this test.

The test statistics [17.4.7] and [17.4.12] are examples of the *Dickey-Fuller test* for unit roots, named for the general battery of tests proposed by Dickey and Fuller (1979).

Case 2. Constant Term but No Time Trend Included in the Regression; True Process Is a Random Walk

For case 2, we continue to assume, as in case 1, that the data are generated by a random walk:

$$y_t = y_{t-1} + u_t,$$

with u_t i.i.d. with mean zero and variance σ^2. Although the true model is the same as in case 1, suppose now that a constant term is included in the $AR(1)$ specification that is to be estimated by OLS:

$$y_t = \alpha + \rho y_{t-1} + u_t. \qquad [17.4.14]$$

The task now is to describe the properties of the OLS estimates,

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T \end{bmatrix} = \begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_t \\ \Sigma y_{t-1} y_t \end{bmatrix}, \qquad [17.4.15]$$

under the null hypothesis that $\alpha = 0$ and $\rho = 1$ (here Σ indicates summation over $t = 1, 2, \ldots, T$). Recall the familiar characterization in [8.2.3] of the deviation of an estimated OLS coefficient vector ($\mathbf{b}_T$) from the true value ($\boldsymbol{\beta}$),

$$\mathbf{b}_T - \boldsymbol{\beta} = \left[\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right]^{-1} \left[\sum_{t=1}^{T} \mathbf{x}_t u_t\right], \qquad [17.4.16]$$

or, in this case,

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{bmatrix} = \begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma u_t \\ \Sigma y_{t-1} u_t \end{bmatrix}. \qquad [17.4.17]$$

As in case 1, y_t has the same properties as the variable ξ_t described in Proposition 17.1 under the maintained hypothesis. Thus, result (d) of that proposition establishes that the sum Σy_{t-1} must be divided by $T^{3/2}$ before obtaining a random variable that converges in distribution:

$$T^{-3/2} \Sigma y_{t-1} \xrightarrow{L} \sigma \cdot \int_0^1 W(r)\, dr. \qquad [17.4.18]$$

In other words,

$$\Sigma y_{t-1} = O_p(T^{3/2}).$$

Similarly, results [17.4.5] and [17.4.6] establish that

$$\Sigma y_{t-1} u_t = O_p(T)$$
$$\Sigma y_{t-1}^2 = O_p(T^2),$$

and from Proposition 17.1(a),

$$\Sigma u_t = O_p(T^{1/2}).$$

Thus, the order in probability of the individual terms in [17.4.17] is as follows:

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{bmatrix} = \begin{bmatrix} O_p(T) & O_p(T^{3/2}) \\ O_p(T^{3/2}) & O_p(T^2) \end{bmatrix}^{-1} \begin{bmatrix} O_p(T^{1/2}) \\ O_p(T) \end{bmatrix}. \qquad [17.4.19]$$

It is clear from [17.4.19] that the estimates $\hat{\alpha}_T$ and $\hat{\rho}_T$ have different rates of convergence, and as in the previous chapter, a scaling matrix $\mathbf{Y}_T$ is helpful in describing their limiting distributions. Recall from [16.1.18] that this rescaling is achieved by premultiplying [17.4.16] by $\mathbf{Y}_T$ and writing the result as

$$\begin{aligned} \mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}) &= \mathbf{Y}_T\left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \mathbf{Y}_T\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t u_t\right] \\ &= \left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]\mathbf{Y}_T^{-1}\right\}^{-1} \left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t u_t\right]\right\}. \end{aligned} \qquad [17.4.20]$$

From [17.4.19], for this application $\mathbf{Y}_T$ should be specified to be the following matrix:

$$\mathbf{Y}_T \equiv \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}, \qquad [17.4.21]$$

for which [17.4.20] becomes

$$\begin{aligned} \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{bmatrix} &= \left\{ \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{bmatrix} \begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{bmatrix} \right\}^{-1} \\ &\quad \times \left\{ \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{bmatrix} \begin{bmatrix} \Sigma u_t \\ \Sigma y_{t-1} u_t \end{bmatrix} \right\} \end{aligned}$$

or

$$\begin{bmatrix} T^{1/2}\hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma y_{t-1} u_t \end{bmatrix}. \qquad [17.4.22]$$

Consider the first term on the right side of [17.4.22]. Results [17.4.6] and [17.4.18] establish that

$$\begin{aligned} &\begin{bmatrix} 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix} \\ &\xrightarrow{L} \begin{bmatrix} 1 & \sigma\cdot\int W(r)\,dr \\ \sigma\cdot\int W(r)\,dr & \sigma^2\cdot\int [W(r)]^2\,dr \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}, \end{aligned} \qquad [17.4.23]$$

where the integral sign denotes integration over r from 0 to 1. Similarly, result (a) of Proposition 17.1 along with [17.4.5] determines the asymptotic distribution of the second term in [17.4.22]:

$$\begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma y_{t-1}u_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma\cdot W(1) \\ (1/2)\sigma^2\{[W(1)]^2 - 1\} \end{bmatrix} = \sigma\begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}\begin{bmatrix} W(1) \\ (1/2)\{[W(1)]^2 - 1\} \end{bmatrix}. \tag{17.4.24}$$

Substituting [17.4.23] and [17.4.24] into [17.4.22] establishes

$$\begin{aligned} \begin{bmatrix} T^{1/2}\hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{bmatrix} &\xrightarrow{L} \sigma\cdot\begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}^{-1}\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \\ &\quad\times \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}\begin{bmatrix} W(1) \\ (1/2)\{[W(1)]^2 - 1\} \end{bmatrix} \\ &= \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \\ &\quad\times \begin{bmatrix} W(1) \\ (1/2)\{[W(1)]^2 - 1\} \end{bmatrix}. \end{aligned} \tag{17.4.25}$$

Notice that

$$\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} = \Delta^{-1}\begin{bmatrix} \int [W(r)]^2\,dr & -\int W(r)\,dr \\ -\int W(r)\,dr & 1 \end{bmatrix}, \tag{17.4.26}$$

where

$$\Delta \equiv \int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2. \tag{17.4.27}$$

Thus, the second element in the vector expression in [17.4.25] states that

$$T(\hat{\rho}_T - 1) \xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\cdot\int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2}. \tag{17.4.28}$$

Neither estimate $\hat{\alpha}_T$ nor $\hat{\rho}_T$ has a limiting Gaussian distribution. Moreover, the asymptotic distribution of the estimate of ρ in [17.4.28] is not the same as the asymptotic distribution in [17.4.7]—when a constant term is included in the distribution, a different table of critical values must be used.

The second section of Table B.5 records percentiles for the distribution of $T(\hat{\rho}_T - 1)$ for case 2. As in case 1, the calculations assume Gaussian innovations, though as T becomes large, these are valid for non-Gaussian innovations as well.

Notice that this distribution is even more strongly skewed than that for case 1, so that when a constant term is included in the regression, the estimated coefficient on y_{t-1} must be farther from unity in order to reject the null hypothesis of a unit root. Indeed, for $T > 25$, 95% of the time the estimated value $\hat{\rho}_T$ will be less than unity. For example, if the estimated value $\hat{\rho}_T$ is 0.999 in a sample of size $T = 100$, the null hypothesis of $\rho = 1$ would be rejected in favor of the alternative that $\rho > 1$! If the true value of ρ is unity, we would not expect to obtain an estimate as large as 0.999.

Dickey and Fuller also proposed an alternative test based on the *OLS t* test of the null hypothesis that $\rho = 1$:

$$t_T = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}}, \qquad [17.4.29]$$

where

$$\hat{\sigma}^2_{\hat{\rho}_T} = s_T^2[0 \quad 1]\begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad [17.4.30]$$

$$s_T^2 = (T-2)^{-1}\sum_{t=1}^{T}(y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2.$$

Notice that if both sides of [17.4.30] are multiplied by T^2, the result can be written as

$$\begin{aligned} T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} &= s_T^2[0 \quad T]\begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ T \end{bmatrix} \\ &= s_T^2[0 \quad 1]\mathbf{Y}_T\begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1}\mathbf{Y}_T\begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned} \qquad [17.4.31]$$

for $\mathbf{Y}_T$ the matrix in [17.4.21]. Recall from [17.4.23] that

$$\begin{aligned} &\mathbf{Y}_T\begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1}\mathbf{Y}_T \\ &= \left\{\mathbf{Y}_T^{-1}\begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}\mathbf{Y}_T^{-1}\right\}^{-1} \\ &= \begin{bmatrix} 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix}^{-1} \\ &\xrightarrow{L} \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}^{-1}\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}^{-1}. \end{aligned} \qquad [17.4.32]$$

Thus, from [17.4.31],

$$T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \xrightarrow{p} s_T^2[0 \quad \sigma^{-1}]\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ \sigma^{-1} \end{bmatrix}. \qquad [17.4.33]$$

It is also easy to show that

$$s_T^2 \xrightarrow{p} \sigma^2, \qquad [17.4.34]$$

from which [17.4.33] becomes

$$T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \xrightarrow{L} [0 \quad 1] \begin{bmatrix} 1 & \int W(r)\, dr \\ \int W(r)\, dr & \int [W(r)]^2\, dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$= \frac{1}{\int [W(r)]^2\, dr - \left[\int W(r)\, dr\right]^2}. \qquad [17.4.35]$$

Thus, the asymptotic distribution of the *OLS t* test in [17.4.29] is

$$t_T = \frac{T(\hat{\rho}_T - 1)}{\{T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T}\}^{1/2}} \xrightarrow{p} T(\hat{\rho}_T - 1) \times \left\{\int [W(r)]^2\, dr - \left[\int W(r)\, dr\right]^2\right\}^{1/2}$$
$$\xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1) \cdot \int W(r)\, dr}{\left\{\int [W(r)]^2\, dr - \left[\int W(r)\, dr\right]^2\right\}^{1/2}}. \qquad [17.4.36]$$

Sample percentiles for the *OLS t* test of $\rho = 1$ are reported for case 2 in the second section of Table B.6. As T grows large, these approach the distribution in the last line of [17.4.36].

Example 17.4

When a constant term is included in the estimated autoregression for the interest rate data from Example 17.3, the result is

$$i_t = \underset{(0.112)}{0.211} + \underset{(0.019133)}{0.96691}\, i_{t-1}, \qquad [17.4.37]$$

with standard errors reported in parentheses. The Dickey-Fuller test based on the estimated value of ρ for this specification is

$$T(\hat{\rho} - 1) = (168)(0.96691 - 1) = -5.56.$$

From Table B.5, the 5% critical value is found by interpolation to be -13.8. Since $-5.56 > -13.8$, the null hypothesis of a unit root ($\rho = 1$) is accepted at the 5% level based on the Dickey-Fuller $\hat{\rho}$ test. The *OLS t* statistic is

$$(0.96691 - 1)/0.019133 = -1.73,$$

which from Table B.6 is to be compared with -2.89. Since $-1.73 > -2.89$, the null hypothesis of a unit root is again accepted.

These statistics test the null hypothesis that $\rho = 1$. However, a maintained assumption on which the derivation of [17.4.25] was based is that the true value of α is zero. Thus, it might seem more natural to test for a unit root in this specification by testing the joint hypothesis that $\alpha = 0$ and $\rho = 1$. Dickey and Fuller (1981) used Monte Carlo to calculate the distribution of the Wald form of the *OLS F* test of this hypothesis (expression [8.1.32] or [8.1.37]). Their values are reported under the heading Case 2 in Table B.7.

Example 17.5

The *OLS* Wald *F* statistic for testing the joint hypothesis that $\alpha = 0$ and $\rho = 1$ for the regression in [17.4.37] is 1.81. Under the classical regression assumptions, this would have an $F(2, 166)$ distribution. In this case, however, the usual statistic is to be compared with the values under Case 2 in Table B.7, for which the 5% critical value is found by interpolation to be 4.67. Since $1.81 < 4.67$, the joint null hypothesis that $\alpha = 0$ and $\rho = 1$ is accepted at the 5% level.

Case 3. Constant Term but No Time Trend Included in the Regression; True Process Is Random Walk with Drift

In case 3, the same regression [17.4.14] is estimated as in case 2, though now it is supposed that the true process is a random walk with drift:

$$y_t = \alpha + y_{t-1} + u_t, \qquad [17.4.38]$$

where the true value of α is not zero. Although this might seem like a minor change, it has a radical effect on the asymptotic distribution of $\hat{\alpha}$ and $\hat{\rho}$. To see why, note that [17.4.38] implies that

$$y_t = y_0 + \alpha t + (u_1 + u_2 + \cdots + u_t) = y_0 + \alpha t + \xi_t, \qquad [17.4.39]$$

where

$$\xi_t \equiv u_1 + u_2 + \cdots + u_t \qquad \text{for } t = 1, 2, \ldots, T$$

with $\xi_0 \equiv 0$.

Consider the behavior of the sum

$$\sum_{t=1}^{T} y_{t-1} = \sum_{t=1}^{T} [y_0 + \alpha(t-1) + \xi_{t-1}]. \qquad [17.4.40]$$

The first term in [17.4.40] is just Ty_0, and if this is divided by T, the result will be a fixed value. The second term, $\Sigma\alpha(t-1)$, must be divided by T^2 in order to converge:

$$T^{-2}\sum_{t=1}^{T} \alpha(t-1) \to \alpha/2,$$

by virtue of Proposition 17.1(h). The third term converges when divided by $T^{3/2}$:

$$T^{-3/2}\sum_{t=1}^{T} \xi_{t-1} \xrightarrow{L} \sigma\cdot\int_0^1 W(r)\,dr,$$

from Proposition 17.1(d). The order in probability of the three individual terms in [17.4.40] is thus

$$\sum_{t=1}^{T} y_{t-1} = \underbrace{\sum_{t=1}^{T} y_0}_{O_p(T)} + \underbrace{\sum_{t=1}^{T} \alpha(t-1)}_{O_p(T^2)} + \underbrace{\sum_{t=1}^{T} \xi_{t-1}}_{O_p(T^{3/2})}.$$

The time trend $\alpha(t-1)$ asymptotically dominates the other two components:

$$\begin{aligned} T^{-2}\sum_{t=1}^{T} y_{t-1} &= T^{-1}y_0 + T^{-2}\sum_{t=1}^{T} \alpha(t-1) + T^{-1/2}\left\{T^{-3/2}\sum_{t=1}^{T} \xi_{t-1}\right\} \\ &\xrightarrow{p} 0 + \alpha/2 + 0. \end{aligned} \qquad [17.4.41]$$

Similarly, we have that

$$\sum_{t=1}^{T} y_{t-1}^2 = \sum_{t=1}^{T} [y_0 + \alpha(t-1) + \xi_{t-1}]^2$$

$$= \underbrace{\sum_{t=1}^{T} y_0^2}_{O_p(T)} + \underbrace{\sum_{t=1}^{T} \alpha^2(t-1)^2}_{O_p(T^3)} + \underbrace{\sum_{t=1}^{T} \xi_{t-1}^2}_{O_p(T^2)}$$

$$+ \underbrace{\sum_{t=1}^{T} 2y_0\alpha(t-1)}_{O_p(T^2)} + \underbrace{\sum_{t=1}^{T} 2y_0\xi_{t-1}}_{O_p(T^{3/2})} + \underbrace{\sum_{t=1}^{T} 2\alpha(t-1)\xi_{t-1}}_{O_p(T^{5/2})}.$$

When divided by T^3, the only term that does not vanish asymptotically is that due to the time trend $\alpha^2(t-1)^2$:

$$T^{-3} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} \alpha^2/3. \qquad [17.4.42]$$

Finally, observe that

$$\sum_{t=1}^{T} y_{t-1}u_t = \sum_{t=1}^{T} [y_0 + \alpha(t-1) + \xi_{t-1}]u_t$$

$$= \underbrace{y_0 \sum_{t=1}^{T} u_t}_{O_p(T^{1/2})} + \underbrace{\sum_{t=1}^{T} \alpha(t-1)u_t}_{O_p(T^{3/2})} + \underbrace{\sum_{t=1}^{T} \xi_{t-1}u_t}_{O_p(T)},$$

from which

$$T^{-3/2} \sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{p} T^{-3/2} \sum_{t=1}^{T} \alpha(t-1)u_t. \qquad [17.4.43]$$

Results [17.4.41] through [17.4.43] imply that when the true process is a random walk with drift, the estimated OLS coefficients in [17.4.15] satisfy

$$\begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{bmatrix} = \begin{bmatrix} O_p(T) & O_p(T^2) \\ O_p(T^2) & O_p(T^3) \end{bmatrix}^{-1} \begin{bmatrix} O_p(T^{1/2}) \\ O_p(T^{3/2}) \end{bmatrix}.$$

Thus, for this case, the Sims, Stock, and Watson scaling matrix would be

$$\mathbf{Y}_T \equiv \begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{bmatrix},$$

for which [17.4.20] becomes

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{bmatrix}$$

$$= \left\{ \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{bmatrix} \begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{bmatrix} \right\}^{-1}$$

$$\times \left\{ \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{bmatrix} \begin{bmatrix} \Sigma u_t \\ \Sigma y_{t-1}u_t \end{bmatrix} \right\}$$

or

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\rho}_T - 1) \end{bmatrix} = \begin{bmatrix} 1 & T^{-2}\Sigma y_{t-1} \\ T^{-2}\Sigma y_{t-1} & T^{-3}\Sigma y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-3/2}\Sigma y_{t-1}u_t \end{bmatrix}. \qquad [17.4.44]$$

From [17.4.41] and [17.4.42], the first term in [17.4.44] converges to

$$\begin{bmatrix} 1 & T^{-2}\Sigma y_{t-1} \\ T^{-2}\Sigma y_{t-1} & T^{-3}\Sigma y_{t-1}^2 \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{bmatrix} \equiv \mathbf{Q}. \qquad [17.4.45]$$

From [17.4.43] and [17.3.18], the second term in [17.4.44] satisfies

$$\begin{aligned} \begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-3/2}\Sigma y_{t-1}u_t \end{bmatrix} &\xrightarrow{p} \begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-3/2}\Sigma\alpha(t-1)u_t \end{bmatrix} \\ &\xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2\begin{bmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{bmatrix}\right) \qquad [17.4.46] \\ &= N(\mathbf{0}, \sigma^2\mathbf{Q}). \end{aligned}$$

Combining [17.4.44] through [17.4.46], it follows that

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} N(\mathbf{0}, \mathbf{Q}^{-1}\cdot\sigma^2\mathbf{Q}\cdot\mathbf{Q}^{-1}) = N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1}). \qquad [17.4.47]$$

Thus, for case 3, both estimated coefficients are asymptotically Gaussian. In fact, the asymptotic properties of $\hat{\alpha}_T$ and $\hat{\rho}_T$ are exactly the same as those for $\hat{\alpha}_T$ and $\hat{\delta}_T$ in the deterministic time trend regression analyzed in Chapter 16. The reason for this correspondence is very simple: the regressor y_{t-1} is asymptotically dominated by the time trend $\alpha\cdot(t-1)$. In large samples, it is as if the explanatory variable y_{t-1} were replaced by the time trend $\alpha\cdot(t-1)$. Recalling the analysis of Section 16.2, it follows that for case 3, the standard *OLS* t and F statistics can be calculated in the usual way and compared with the standard tables (Tables B.3 and B.4, respectively).

Case 4. Constant Term and Time Trend Included in the Regression; True Process Is Random Walk With or Without Drift

Suppose, as in the previous case, that the true model is

$$y_t = \alpha + y_{t-1} + u_t,$$

where u_t is i.i.d. with mean zero and variance σ^2. For this case, the true value of α turns out not to matter for the asymptotic distribution. In contrast to the previous case, we now assume that a time trend is included in the regression that is actually estimated by *OLS*:

$$y_t = \alpha + \rho y_{t-1} + \delta t + u_t. \qquad [17.4.48]$$

If $\alpha \neq 0$, y_{t-1} would be asymptotically equivalent to a time trend. Since a time trend is already included as a separate variable in the regression, this would make the explanatory variables collinear in large samples. Describing the asymptotic distribution of the estimates therefore requires not just a rescaling of variables but also a rotation of the kind introduced in Section 16.3.

Note that the regression model of [17.4.48] can equivalently be written as

$$\begin{aligned} y_t &= (1-\rho)\alpha + \rho[y_{t-1} - \alpha(t-1)] + (\delta + \rho\alpha)t + u_t \\ &\equiv \alpha^* + \rho^*\xi_{t-1} + \delta^* t + u_t, \end{aligned} \qquad [17.4.49]$$

where $\alpha^* \equiv (1-\rho)\alpha$, $\rho^* \equiv \rho$, $\delta^* \equiv (\delta + \rho\alpha)$, and $\xi_t \equiv y_t - \alpha t$. Moreover, under the null hypothesis that $\rho = 1$ and $\delta = 0$,

$$\xi_t = y_0 + u_1 + u_2 + \cdots + u_t;$$

that is, ξ_t is the random walk described in Proposition 17.1. Consider, as in Section 16.3, a hypothetical regression of y_t on a constant, ξ_{t-1}, and a time trend, producing the *OLS* estimates

$$\begin{bmatrix} \hat{\alpha}_T^* \\ \hat{\rho}_T^* \\ \hat{\delta}_T^* \end{bmatrix} = \begin{bmatrix} T & \Sigma\xi_{t-1} & \Sigma t \\ \Sigma\xi_{t-1} & \Sigma\xi_{t-1}^2 & \Sigma\xi_{t-1}t \\ \Sigma t & \Sigma t\xi_{t-1} & \Sigma t^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_t \\ \Sigma\xi_{t-1}y_t \\ \Sigma t y_t \end{bmatrix}. \qquad [17.4.50]$$

The maintained hypothesis is that $\alpha = \alpha_0$, $\rho = 1$, and $\delta = 0$, which in the transformed system would mean $\alpha^* = 0$, $\rho^* = 1$, and $\delta^* = \alpha_0$. The deviations of the *OLS* estimates from these true values are given by

$$\begin{bmatrix} \hat{\alpha}_T^* \\ \hat{\rho}_T^* - 1 \\ \hat{\delta}_T^* - \alpha_0 \end{bmatrix} = \begin{bmatrix} T & \Sigma\xi_{t-1} & \Sigma t \\ \Sigma\xi_{t-1} & \Sigma\xi_{t-1}^2 & \Sigma\xi_{t-1}t \\ \Sigma t & \Sigma t\xi_{t-1} & \Sigma t^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma u_t \\ \Sigma\xi_{t-1}u_t \\ \Sigma t u_t \end{bmatrix}. \qquad [17.4.51]$$

Consulting the rates of convergence in Proposition 17.1, in this case the scaling matrix should be

$$\mathbf{Y}_T = \begin{bmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{bmatrix},$$

and [17.4.20] would be

$$\begin{aligned} &\begin{bmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_T^* \\ \hat{\rho}_T^* - 1 \\ \hat{\delta}_T^* - \alpha_0 \end{bmatrix} \\ &= \left\{ \begin{bmatrix} T^{-1/2} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & T^{-3/2} \end{bmatrix} \begin{bmatrix} T & \Sigma\xi_{t-1} & \Sigma t \\ \Sigma\xi_{t-1} & \Sigma\xi_{t-1}^2 & \Sigma\xi_{t-1}t \\ \Sigma t & \Sigma t\xi_{t-1} & \Sigma t^2 \end{bmatrix} \right. \\ &\qquad \left. \times \begin{bmatrix} T^{-1/2} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & T^{-3/2} \end{bmatrix} \right\}^{-1} \\ &\qquad \times \left\{ \begin{bmatrix} T^{-1/2} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & T^{-3/2} \end{bmatrix} \begin{bmatrix} \Sigma u_t \\ \Sigma\xi_{t-1}u_t \\ \Sigma t u_t \end{bmatrix} \right\} \end{aligned}$$

or

$$\begin{bmatrix} T^{1/2}\hat{\alpha}_T^* \\ T(\hat{\rho}_T^* - 1) \\ T^{3/2}(\hat{\delta}_T^* - \alpha_0) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma t \\ T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma\xi_{t-1}^2 & T^{-5/2}\Sigma\xi_{t-1}t \\ T^{-2}\Sigma t & T^{-5/2}\Sigma t\xi_{t-1} & T^{-3}\Sigma t^2 \end{bmatrix}^{-1} \times \begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma\xi_{t-1}u_t \\ T^{-3/2}\Sigma tu_t \end{bmatrix} \qquad [17.4.52]$$

The asymptotic distribution can then be found from Proposition 17.1:

$$\begin{bmatrix} T^{1/2}\hat{\alpha}_T^* \\ T(\hat{\rho}_T^* - 1) \\ T^{3/2}(\hat{\delta}_T^* - \alpha_0) \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} 1 & \sigma\cdot\int W(r)\,dr & \frac{1}{2} \\ \sigma\cdot\int W(r)\,dr & \sigma^2\cdot\int [W(r)]^2\,dr & \sigma\cdot\int rW(r)\,dr \\ \frac{1}{2} & \sigma\cdot\int rW(r)\,dr & \frac{1}{3} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \sigma\cdot W(1) \\ \frac{1}{2}\sigma^2\{[W(1)]^2 - 1\} \\ \sigma\cdot\{W(1) - \int W(r)\,dr\} \end{bmatrix}$$

$$= \sigma\begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & \int W(r)\,dr & \frac{1}{2} \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ \frac{1}{2} & \int rW(r)\,dr & \frac{1}{3} \end{bmatrix}^{-1} \qquad [17.4.53]$$

$$\times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \\ W(1) - \int W(r)\,dr \end{bmatrix}$$

$$= \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\,dr & \frac{1}{2} \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ \frac{1}{2} & \int rW(r)\,dr & \frac{1}{3} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \\ W(1) - \int W(r)\,dr \end{bmatrix}.$$

Note that $\hat{\rho}_T^*$, the *OLS* estimate of ρ based on [17.4.49], is identical to $\hat{\rho}_T$, the *OLS* estimate of ρ based on [17.4.48]. Thus, the asymptotic distribution of $T(\hat{\rho}_T - 1)$ is given by the middle row of [17.4.53]. Note that this distribution does not depend on either σ or α; in particular, it does not matter whether or not the true value of α is zero.

The asymptotic distribution of $\hat{\sigma}_{\hat{\rho}_T}$, the *OLS* standard error for $\hat{\rho}_T$, can be found using similar calculations to those in [17.4.31] and [17.4.32]. Notice that

$$
\begin{aligned}
T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} &= T^2 \cdot s_T^2 [0 \quad 1 \quad 0] \begin{bmatrix} T & \Sigma\xi_{t-1} & \Sigma t \\ \Sigma\xi_{t-1} & \Sigma\xi_{t-1}^2 & \Sigma\xi_{t-1}t \\ \Sigma t & \Sigma t\xi_{t-1} & \Sigma t^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
&= s_T^2 [0 \quad 1 \quad 0] \begin{bmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{bmatrix} \\
&\quad \times \begin{bmatrix} T & \Sigma\xi_{t-1} & \Sigma t \\ \Sigma\xi_{t-1} & \Sigma\xi_{t-1}^2 & \Sigma\xi_{t-1}t \\ \Sigma t & \Sigma t\xi_{t-1} & \Sigma t^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
&= s_T^2 [0 \quad 1 \quad 0] \\
&\quad \times \begin{bmatrix} 1 & T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma t \\ T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma\xi_{t-1}^2 & T^{-5/2}\Sigma\xi_{t-1}t \\ T^{-2}\Sigma t & T^{-5/2}\Sigma t\xi_{t-1} & T^{-3}\Sigma t^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
&\xrightarrow{L} \sigma^2 [0 \quad 1 \quad 0] \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \\
&\quad \times \begin{bmatrix} 1 & \int W(r)\,dr & \frac{1}{2} \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ \frac{1}{2} & \int rW(r)\,dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
&= [0 \quad 1 \quad 0] \begin{bmatrix} 1 & \int W(r)\,dr & \frac{1}{2} \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ \frac{1}{2} & \int rW(r)\,dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
&\equiv Q.
\end{aligned}
\qquad [17.4.54]
$$

From this result it follows that the asymptotic distribution of the *OLS* t test of the hypothesis that $\rho = 1$ is given by

$$t_T = T(\hat{\rho}_T - 1) \div (T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T})^{1/2} \xrightarrow{p} T(\hat{\rho}_T - 1) \div \sqrt{Q}. \qquad [17.4.55]$$

Again, this distribution does not depend on α or σ. The small-sample distribution of the *OLS* t statistic under the assumption of Gaussian disturbances is presented under case 4 in Table B.6. If this distribution were truly t, then a value below -2.0 would be sufficient to reject the null hypothesis. However, Table B.6 reveals that, because of the nonstandard distribution, the t statistic must be below -3.4 before the null hypothesis of a unit root could be rejected.

The assumption that the true value of δ is equal to zero is again an auxiliary hypothesis upon which the asymptotic properties of the test depend. Thus, as in case 2, it is natural to consider the *OLS* F test of the joint null hypothesis that $\delta = 0$ and $\rho = 1$. Though this F test is calculated in the usual way, its asymptotic distribution is nonstandard, and the calculated F statistic should be compared with the value under case 4 in Table B.7.

Summary of Dickey-Fuller Tests in the Absence of Serial Correlation

We have seen that the asymptotic properties of the *OLS* estimate $\hat{\rho}_T$ when the true value of ρ is unity depend on whether or not a constant term or a time trend is included in the regression that is estimated and on whether or not the random walk that describes the true process for y_t includes a drift term. These results are summarized in Table 17.1.

Which is the "correct" case to use to test the null hypothesis of a unit root? The answer depends on why we are interested in testing for a unit root. If the analyst has a specific null hypothesis about the process that generated the data, then obviously this would guide the choice of test. In the absence of such guidance, one general principle would be to fit a specification that is a plausible description of the data under both the null hypothesis and the alternative. This principle would suggest using the case 4 test for a series with an obvious trend and the case 2 test for series without a significant trend.

For example, Figure 17.2 plots the nominal interest rate series used in the examples in this section. Although this series has tended upward over this sample period, there is nothing in economic theory to suggest that nominal interest rates should exhibit a deterministic time trend, and so a natural null hypothesis is that the true process is a random walk without trend. In terms of framing a plausible alternative, it is difficult to maintain that these data could have been generated by $i_t = \rho i_{t-1} + u_t$ with $|\rho|$ significantly less than 1. If these data were to be described by a stationary process, surely the process would have a positive mean. This argues for including a constant term in the estimated regression, even though under the null hypothesis the true process does not contain a constant term. Thus, case 2 is a sensible approach for these data, as analyzed in Examples 17.4 and 17.5.

As a second example, Figure 17.3 plots quarterly real GNP for the United States from 1947:I to 1989:I. Given a growing population and technological improvements, such a series would certainly be expected to exhibit a persistent upward trend, and this trend is unmistakable in the figure. The question is whether this trend arises from the positive drift term of a random walk:

$$H_0\colon y_t = \alpha + y_{t-1} + u_t \qquad \alpha > 0,$$

or from a deterministic time trend added to a stationary $AR(1)$:

$$H_A\colon y_t = \alpha + \delta t + \rho y_{t-1} + u_t \qquad |\rho| < 1.$$

Thus, the recommended test statistics for this case are those described in case 4.

The following model for 100 times the log of real GNP (denoted y_t) was estimated by *OLS* regression:

$$y_t = \underset{(13.53)}{27.24} + \underset{(0.019304)}{0.96252}\, y_{t-1} + \underset{(0.01521)}{0.02753}\, t. \qquad [17.4.56]$$

(standard errors in parentheses). The sample size is $T = 168$. The Dickey-Fuller

ρ test is

$$T(\hat{\rho} - 1) = 168(0.96252 - 1.0) = -6.3.$$

Since $-6.3 > -21.0$, the null hypothesis that GNP is characterized by a random walk with possible drift is accepted at the 5% level. The Dickey-Fuller t test,

$$t = \frac{0.96252 - 1.0}{0.019304} = -1.94,$$

exceeds the 5% critical value of -3.44, so that the null hypothesis of a unit root is accepted by this test as well. Finally, the F test of the joint null hypothesis that $\delta = 0$ and $\rho = 1$ is 2.44. Since this is less than the 5% critical value of 6.42 from Table B.7, this null hypothesis is again accepted.

TABLE 17.1
Summary of Dickey-Fuller Tests for Unit Roots in the Abseuce of Serial Correlation

Case 1:

Estimated regression: $y_t = \rho y_{t-1} + u_t$
True process: $y_t = y_{t-1} + u_t \qquad u_t \sim \text{i.i.d. } N(0, \sigma^2)$
$T(\hat{\rho}_T - 1)$ has the distribution described under the heading Case 1 in Table B.5.
$(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T}$ has the distribution described under Case 1 in Table B.6.

Case 2:

Estimated regression: $y_t = \alpha + \rho y_{t-1} + u_t$
True process: $y_t = y_{t-1} + u_t \qquad u_t \sim \text{i.i.d. } N(0, \sigma^2)$
$T(\hat{\rho}_T - 1)$ has the distribution described under Case 2 in Table B.5.
$(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T}$ has the distribution described under Case 2 in Table B.6.
OLS F test of joint hypothesis that $\alpha = 0$ and $\rho = 1$ has the distribution described under Case 2 in Table B.7.

Case 3:

Estimated regression: $y_t = \alpha + \rho y_{t-1} + u_t$
True process: $y_t = \alpha + y_{t-1} + u_t \qquad \alpha \neq 0, u_t \sim \text{i.i.d. } (0, \sigma^2)$
$(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T} \xrightarrow{L} N(0, 1)$

Case 4:

Estimated regression: $y_t = \alpha + \rho y_{t-1} + \delta t + u_t$
True process: $y_t = \alpha + y_{t-1} + u_t \qquad \alpha$ any, $u_t \sim \text{i.i.d. } N(0, \sigma^2)$
$T(\hat{\rho}_T - 1)$ has the distribution described under Case 4 in Table B.5.
$(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T}$ has the distribution described under Case 4 in Table B.6.
OLS F test of joint hypothesis that $\rho = 1$ and $\delta = 0$ has the distribution described under Case 4 in Table B.7.

Notes to Table 17.1

Estimated regression indicates the form in which the regression is estimated, using observations $t = 1, 2, \ldots, T$ and conditioning on observation $t = 0$.

True process describes the null hypothesis under which the distribution is calculated.

$\hat{\rho}_T$ is the *OLS* estimate of ρ from the indicated regression based on a sample of size T.

$(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T}$ is the *OLS t* test of $\rho = 1$.

OLS F test of a hypothesis involving two restrictions is given by expression [17.7.39].

If $u_t \sim$ i.i.d. $N(0, \sigma^2)$, then Tables B.5 through B.7 give Monte Carlo estimates of the exact small-sample distribution. The tables are also valid for large T when u_t is non-Gaussian i.i.d. as well as for certain heterogeneously distributed serially uncorrelated processes. For serially correlated u_t, see Table 17.2 or 17.3.

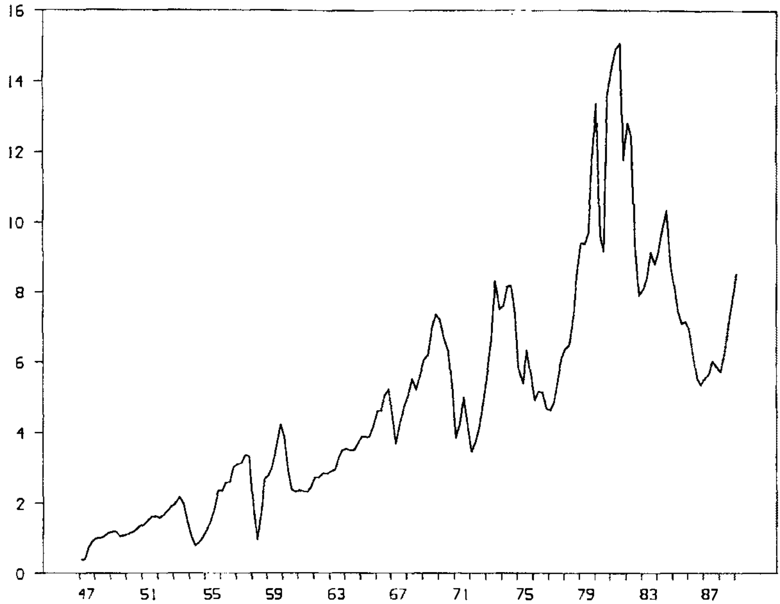

FIGURE 17.2 U.S. nominal interest rate on 3-month Treasury bills, data sampled quarterly but quoted at an annual rate, 1947:I to 1989:I.

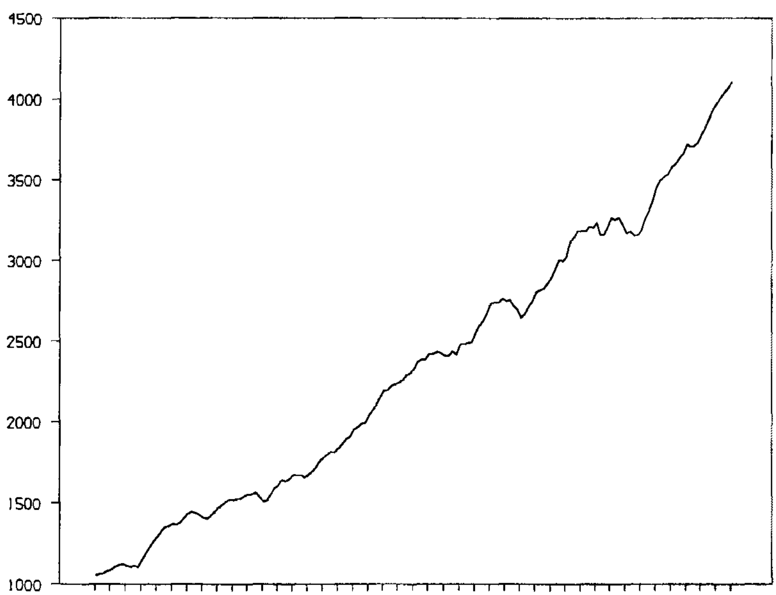

FIGURE 17.3 U.S. real GNP, data sampled quarterly but quoted at an annual rate in billions of 1982 dollars, 1947:I to 1989:I.

Of the tests discussed so far, those developed for case 2 seem appropriate for the interest rate data and the tests developed for case 4 seem best for the GNP data. However, more general tests presented in Sections 17.6 and 17.7 are to be preferred for describing either of these series. This is because the maintained assumption throughout this section has been that the disturbance term u_t in the regression is i.i.d. There is no strong reason to expect this for either of these time series. The next section develops results that can be used to test for unit roots in serially correlated processes.

17.5. *Asymptotic Results for Unit Root Processes with General Serial Correlation*

This section generalizes Proposition 17.1 to allow for serial correlation. The following preliminary result is quite helpful.

Proposition 17.2: *Let*

$$u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad [17.5.1]$$

where

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t \varepsilon_\tau) = \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{j=0}^{\infty} j \cdot |\psi_j| < \infty. \qquad [17.5.2]$$

Then

$$u_1 + u_2 + \cdots + u_t = \psi(1) \cdot (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + \eta_t - \eta_0, \qquad [17.5.3]$$

where $\psi(1) \equiv \Sigma_{j=0}^{\infty} \psi_j$, $\eta_t = \Sigma_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$, $\alpha_j = -(\psi_{j+1} + \psi_{j+2} + \psi_{j+3} + \cdots)$, *and* $\Sigma_{j=0}^{\infty} |\alpha_j| < \infty$.

The condition in [17.5.2] is slightly stronger than absolute summability, though it is satisfied by any stationary *ARMA* process.

Notice that if y_t is an $I(1)$ process y_t whose first difference is given by u_t, or

$$\Delta y_t = u_t,$$

then

$$y_t = u_1 + u_2 + \cdots + u_t + y_0 = \psi(1) \cdot (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + \eta_t - \eta_0 + y_0.$$

Proposition 17.2 thus states that any $I(1)$ process whose first difference satisfies [17.5.1] and [17.5.2] can be written as the sum of a random walk $(\psi(1) \cdot (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t))$, initial conditions $(y_0 - \eta_0)$, and a stationary process (η_t). This observation was first made by Beveridge and Nelson (1981), and [17.5.3] is sometimes referred to as the *Beveridge-Nelson decomposition*.

Notice that η_t is a stationary process. An important implication of this is that if [17.5.3] is divided by $\sqrt{t}$, only the first term $(1/\sqrt{t})\psi(1) \cdot (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t)$ should matter for the distribution of $(1/\sqrt{t}) \cdot (u_1 + u_2 + \cdots + u_t)$ as $t \to \infty$.

As an example of how this result can be used, suppose that $X_T(r)$ is defined as in [17.3.2]:

$$X_T(r) \equiv (1/T) \sum_{t=1}^{[Tr]^*} u_t, \qquad [17.5.4]$$

where u_t satisfies the conditions of Proposition 17.2 with ε_t i.i.d. and $E(\varepsilon_t^4) < \infty$. Then the continuous-time process $\sqrt{T}\cdot X_T(r)$ converges to $\sigma\cdot\psi(1)$ times standard Brownian motion:

$$\sqrt{T}\cdot X_T(\cdot) \xrightarrow{L} \sigma\cdot\psi(1)\cdot W(\cdot). \qquad [17.5.5]$$

To derive [17.5.5], note from Proposition 17.2 that

$$\begin{aligned}\sqrt{T}\cdot X_T(r) &= (1/\sqrt{T})\cdot\sum_{t=1}^{[Tr]^*} u_t \\ &= \psi(1)\cdot(1/\sqrt{T})\cdot\sum_{t=1}^{[Tr]^*} \varepsilon_t + (1/\sqrt{T})\cdot(\eta_{[Tr]^*} - \eta_0) \qquad [17.5.6] \\ &\equiv \psi(1)\cdot(1/\sqrt{T})\cdot\sum_{t=1}^{[Tr]^*} \varepsilon_t + S_T(r),\end{aligned}$$

where we have defined $S_T(r) \equiv (1/\sqrt{T})\cdot(\eta_{[Tr]^*} - \eta_0)$. Notice as in Example 17.2 that

$$S_T(\cdot) \xrightarrow{p} 0 \qquad [17.5.7]$$

as $T \to \infty$. Furthermore, from [17.3.8],

$$(1/\sqrt{T})\cdot\sum_{t=1}^{[Tr]^*} \varepsilon_t \xrightarrow{L} \sigma\cdot W(r). \qquad [17.5.8]$$

Substituting [17.5.7] and [17.5.8] into [17.5.6] produces [17.5.5].

Another implication is found by evaluating the functions in [17.5.5] at $r = 1$:

$$(1/\sqrt{T}) \sum_{t=1}^{T} u_t \xrightarrow{L} \sigma\cdot\psi(1)\cdot W(1). \qquad [17.5.9]$$

Since $W(1)$ is distributed $N(0, 1)$, result [17.5.9] states that

$$(1/\sqrt{T}) \sum_{t=1}^{T} u_t \xrightarrow{L} N(0, \sigma^2[\psi(1)]^2),$$

which is the usual central limit theorem of Proposition 7.11.

The following proposition uses this basic idea to generalize the other results from Proposition 17.1; for details on the proofs, see Appendix 17.A.

Proposition 17.3: *Let $u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} j\cdot|\psi_j| < \infty$ and $\{\varepsilon_t\}$ is an i.i.d. sequence with mean zero, variance σ^2, and finite fourth moment. Define*

$$\gamma_j \equiv E(u_t u_{t-j}) = \sigma^2 \sum_{s=0}^{\infty} \psi_s\psi_{s+j} \qquad \textit{for } j = 0, 1, 2, \ldots \qquad [17.5.10]$$

$$\lambda \equiv \sigma \sum_{j=0}^{\infty} \psi_j = \sigma\cdot\psi(1)$$

$$\xi_t \equiv u_1 + u_2 + \cdots + u_t \qquad \textit{for } t = 1, 2, \ldots, T \qquad [17.5.11]$$

with $\xi_0 \equiv 0$. Then

(a) $T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{L} \lambda \cdot W(1);$

(b) $T^{-1/2} \sum_{t=1}^{T} u_{t-j}\varepsilon_t \xrightarrow{L} N(0, \sigma^2\gamma_0) \qquad \text{for } j = 1, 2, \ldots;$

(c) $T^{-1} \sum_{t=1}^{T} u_t u_{t-j} \xrightarrow{p} \gamma_j \qquad \text{for } j = 0, 1, 2, \ldots;$

(d) $T^{-1} \sum_{t=1}^{T} \xi_{t-1}\varepsilon_t \xrightarrow{L} (1/2)\sigma\cdot\lambda\cdot\{[W(1)]^2 - 1\};$

(e) $T^{-1} \sum_{t=1}^{T} \xi_{t-1}u_{t-j}$

$$\xrightarrow{L} \begin{cases} (1/2)\{\lambda^2\cdot[W(1)]^2 - \gamma_0\} & \text{for } j = 0 \\ (1/2)\{\lambda^2\cdot[W(1)]^2 - \gamma_0\} + \gamma_0 + \gamma_1 + \gamma_2 + \cdots + \gamma_{j-1} & \\ & \text{for } j = 1, 2, \ldots; \end{cases}$$

(f) $T^{-3/2} \sum_{t=1}^{T} \xi_{t-1} \xrightarrow{L} \lambda\cdot\int_0^1 W(r)\, dr;$

(g) $T^{-3/2} \sum_{t=1}^{T} tu_{t-j} \xrightarrow{L} \lambda\cdot\left\{W(1) - \int_0^1 W(r)\, dr\right\} \qquad \text{for } j = 0, 1, 2, \ldots;$

(h) $T^{-2} \sum_{t=1}^{T} \xi_{t-1}^2 \xrightarrow{L} \lambda^2\cdot\int_0^1 [W(r)]^2\, dr;$

(i) $T^{-5/2} \sum_{t=1}^{T} t\xi_{t-1} \xrightarrow{L} \lambda\cdot\int_0^1 rW(r)\, dr;$

(j) $T^{-3} \sum_{t=1}^{T} t\xi_{t-1}^2 \xrightarrow{L} \lambda^2\cdot\int_0^1 r\cdot[W(r)]^2\, dr;$

(k) $T^{-(v+1)} \sum_{t=1}^{T} t^v \rightarrow 1/(v + 1) \qquad \text{for } v = 0, 1, \ldots.$

Again, there are simpler ways to describe individual results; for example, (a) is a $N(0, \lambda^2)$ distribution, (d) is $(1/2)\sigma\lambda\cdot[\chi^2(1) - 1]$, and (f) and (g) are both $N(0, \lambda^2/3)$ distributions.

These results can be used to construct unit root tests for serially correlated observations in two ways. One approach, due to Phillips (1987) and Phillips and Perron (1988), is to continue to estimate the regressions in exactly the form indicated in Table 17.1, but to adjust the test statistics to take account of serial correlation and potential heteroskedasticity in the disturbances. This approach is described in Section 17.6. The second approach, due to Dickey and Fuller (1979), is to add lagged changes of y as explanatory variables in the regressions in Table 17.1. This is described in Section 17.7.

17.6. *Phillips-Perron Tests for Unit Roots*

Asymptotic Distribution for Case 2 Assumptions with Serially Correlated Disturbances

To illustrate the basic idea behind the Phillips (1987) and Phillips and Perron (1988) tests for unit roots, we will discuss in detail the treatment they propose for the analog of case 2 of Section 17.4. After this case has been reviewed, similar results will be stated for case 1 and case 4, with details developed in exercises at the end of the chapter.

Case 2 of Section 17.4 considered *OLS* estimation of α and ρ in the regression model

$$y_t = \alpha + \rho y_{t-1} + u_t \qquad [17.6.1]$$

under the assumption that the true $\alpha = 0$, $\rho = 1$, and u_t is i.i.d. Phillips and Perron (1988) generalized these results to the case when u_t is serially correlated and possibly heteroskedastic as well. For now we will assume that the true process is

$$y_t - y_{t-1} = u_t = \psi(L)\varepsilon_t,$$

where $\psi(L)$ and ε_t satisfy the conditions of Proposition 17.3. More general conditions under which the same techniques are valid will be discussed at the end of this section.

If [17.6.1] were a stationary autoregression with $|\rho| < 1$, the *OLS* estimate $\hat{\rho}_T$ in [17.4.15] would not give a consistent estimate of ρ when u_t is serially correlated. However, if ρ is equal to 1, the rate T convergence of $\hat{\rho}_T$ turns out to ensure that $\hat{\rho}_T \xrightarrow{p} 1$ even when u_t is serially correlated. Phillips and Perron (1988) therefore proposed estimating [17.6.1] by *OLS* even when u_t is serially correlated and then modifying the statistics in Section 17.4 to take account of the serial correlation.

Let $\hat{\alpha}_T$ and $\hat{\rho}_T$ be the *OLS* estimates based on [17.6.1] without any correction for serial correlation; that is, $\hat{\alpha}_T$ and $\hat{\rho}_T$ are the magnitudes defined in [17.4.15]. If the true values are $\alpha = 0$ and $\rho = 1$, then, as in [17.4.22],

$$\begin{bmatrix} T^{1/2}\hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma y_{t-1}u_t \end{bmatrix}, \qquad [17.6.2]$$

where Σ denotes summation over t from 1 to T. Also, under the null hypothesis that $\alpha = 0$ and $\rho = 1$, it follows as in [17.4.4] that

$$y_t = y_0 + u_1 + u_2 + \cdots + u_t.$$

If $u_t = \psi(L)\varepsilon_t$ as in Proposition 17.3, then y_t is the variable labeled ξ_t in Proposition 17.3, plus the inconsequential value y_0. Using results (f) and (h) of that proposition,

$$\begin{aligned} &\begin{bmatrix} 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix}^{-1} \\ &\xrightarrow{L} \begin{bmatrix} 1 & \lambda\cdot\int W(r)\,dr \\ \lambda\cdot\int W(r)\,dr & \lambda^2\cdot\int [W(r)]^2\,dr \end{bmatrix}^{-1} \qquad [17.6.3] \\ &= \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1} \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1}, \end{aligned}$$

where the integral sign denotes integration over r from 0 to 1. Similarly, results (a) and (e) of Proposition 17.3 give

$$\begin{aligned} \begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma y_{t-1}u_t \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} \lambda\cdot W(1) \\ \tfrac{1}{2}\{\lambda^2[W(1)]^2 - \gamma_0\} \end{bmatrix} \\ &= \begin{bmatrix} \lambda\cdot W(1) \\ \tfrac{1}{2}\lambda^2\{[W(1)]^2 - 1\} \end{bmatrix} + \begin{bmatrix} 0 \\ \tfrac{1}{2}\{\lambda^2 - \gamma_0\} \end{bmatrix} \qquad [17.6.4] \\ &= \lambda\begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} W(1) \\ \tfrac{1}{2}\{[W(1)]^2 - 1\} \end{bmatrix} + \begin{bmatrix} 0 \\ \tfrac{1}{2}\{\lambda^2 - \gamma_0\} \end{bmatrix}. \end{aligned}$$

Substituting [17.6.3] and [17.6.4] into [17.6.2] produces

$$
\begin{aligned}
\begin{bmatrix} T^{1/2}\hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{bmatrix} &\xrightarrow{L} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1} \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \\
&\quad \times \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1} \left\{ \lambda \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{2}\{\lambda^2 - \gamma_0\} \end{bmatrix} \right\} \\
&= \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \end{bmatrix} \\
&\quad + \left\{ \begin{bmatrix} 1 & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \frac{1}{2}\{\lambda^2 - \gamma_0\}/\lambda \end{bmatrix} \right\}.
\end{aligned}
$$

[17.6.5]

The second element of this vector states that

$$
\begin{aligned}
&T(\hat{\rho}_T - 1) \\
&\xrightarrow{L} [0 \quad 1] \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \end{bmatrix} \\
&\quad + (1/2)\frac{\{\lambda^2 - \gamma_0\}}{\lambda^2} [0 \quad 1] \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad [17.6.6] \\
&= \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2} + \frac{(1/2)\cdot(\lambda^2 - \gamma_0)}{\lambda^2\left\{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2\right\}}.
\end{aligned}
$$

The first term of the last equality in [17.6.6] is the same as [17.4.28], which described the asymptotic distribution that $T(\hat{\rho}_T - 1)$ would have if u_t were i.i.d. The final term in [17.6.6] is a correction for serial correlation. Notice that if u_t is serially uncorrelated, then $\psi_0 = 1$ and $\psi_j = 0$ for $j = 1, 2, \ldots$. Thus, if u_t is serially uncorrelated, then $\lambda^2 = \sigma^2 \cdot [\psi(1)]^2 = \sigma^2$ and $\gamma_0 = E(u_t^2) = \sigma^2$. Hence [17.6.6] includes the earlier result [17.4.28] as a special case when u_t is serially uncorrelated.

It is easy to use $\hat{\sigma}_{\hat{\rho}_T}$, the *OLS* standard error for $\hat{\rho}_T$, to construct a sample statistic that can be used to estimate the correction for serial correlation. Let $\mathbf{Y}_T$ be the matrix defined in [17.4.21], and let s_T^2 be the *OLS* estimate of the variance of u_t:

$$
s_T^2 = (T - 2)^{-1} \sum_{t=1}^{T} (y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2.
$$

Then the asymptotic distribution of $T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T}$ can be found using the same approach as in [17.4.31] through [17.4.33]:

$$
\begin{aligned}
&T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T}\\
&= s_T^2[0 \quad 1]\mathbf{Y}_T\begin{bmatrix} T & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2\end{bmatrix}^{-1}\mathbf{Y}_T\begin{bmatrix}0\\1\end{bmatrix}\\
&\xrightarrow{p} s_T^2[0 \quad 1]\begin{bmatrix}1 & 0\\ 0 & \lambda\end{bmatrix}^{-1}\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr\end{bmatrix}^{-1}\begin{bmatrix}1 & 0\\ 0 & \lambda\end{bmatrix}^{-1}\begin{bmatrix}0\\1\end{bmatrix} \qquad [17.6.7]\\
&= (s_T^2/\lambda^2)[0 \quad 1]\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr\end{bmatrix}^{-1}\begin{bmatrix}0\\1\end{bmatrix}\\
&= (s_T^2/\lambda^2)\frac{1}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2}.
\end{aligned}
$$

It follows from [17.6.6] that

$$
\begin{aligned}
&T(\hat{\rho}_T - 1) - \tfrac{1}{2}(T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2)(\lambda^2 - \gamma_0)\\
&\xrightarrow{p} T(\hat{\rho}_T - 1) - \frac{1}{2}\left(\frac{1}{\lambda^2}\right)\frac{1}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2}(\lambda^2 - \gamma_0) \qquad [17.6.8]\\
&\xrightarrow{L} \frac{\tfrac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2}.
\end{aligned}
$$

Thus, the statistic in [17.6.8] has the same asymptotic distribution [17.4.28] as the variable tabulated under the heading Case 2 in Table B.5.

Result [17.6.8] can also be used to find the asymptotic distribution of the *OLS t* test of $\rho = 1$:

$$
\begin{aligned}
t_T &= \frac{(\hat{\rho}_T - 1)}{\hat{\sigma}_{\hat{\rho}_T}}\\
&= \frac{T(\hat{\rho}_T - 1)}{\{T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T}\}^{1/2}}\\
&\xrightarrow{p} \left\{\frac{\tfrac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2} + \tfrac{1}{2}(T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T}/s_T^2)\cdot(\lambda^2 - \gamma_0)\right\}\\
&\qquad \div \{T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T}\}^{1/2}\\
&= \frac{\tfrac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2} \div \{T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T}\}^{1/2}\\
&\qquad + \{(1/2)(1/s_T)(\lambda^2 - \gamma_0)\} \times \{T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2\}^{1/2}
\end{aligned}
$$

$$\xrightarrow{p} \left\{ \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2} \right\} \left(\frac{\lambda^2}{s_T^2}\right)^{1/2}$$

$$\times \left\{ \int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2 \right\}^{1/2}$$

$$+ \{(1/2)(1/s_T)(\lambda^2 - \gamma_0)\} \times \{T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2\}^{1/2} \qquad [17.6.9]$$

with the last convergence following from [17.6.7]. Moreover,

$$s_T^2 = (T-2)^{-1} \sum_{t=1}^{T} (y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2 \xrightarrow{p} E(u_t^2) = \gamma_0. \qquad [17.6.10]$$

Hence, [17.6.9] implies that

$$(\gamma_0/\lambda^2)^{1/2} \cdot t_T \xrightarrow{p} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)\,dr}{\left\{ \int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2 \right\}^{1/2}} \qquad [17.6.11]$$

$$+ \{\tfrac{1}{2}(\lambda^2 - \gamma_0)/\lambda\} \times \{T \cdot \hat{\sigma}_{\hat{\rho}_T} \div s_T\}.$$

Thus,

$$(\gamma_0/\lambda^2)^{1/2} \cdot t_T - \{\tfrac{1}{2}(\lambda^2 - \gamma_0)/\lambda\} \times \{T \cdot \hat{\sigma}_{\hat{\rho}_T} \div s_T\}$$

$$\xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1)\int W(r)\,dr}{\left\{ \int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2 \right\}^{1/2}}, \qquad [17.6.12]$$

which is the same limiting distribution [17.4.36] obtained for the random variable tabulated for case 2 in Table B.6.

The statistics in [17.6.8] and [17.6.12] require knowledge of the population parameters γ_0 and λ^2. Although these moments are unknown, they are easy to estimate consistently. Since $\gamma_0 = E(u_t^2)$, one consistent estimate is given by

$$\hat{\gamma}_0 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2, \qquad [17.6.13]$$

where $\hat{u}_t = y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1}$ is the *OLS* sample residual. Phillips and Perron (1988) instead used the standard *OLS* estimate $\hat{\gamma}_0 = (T-2)^{-1}\Sigma\hat{u}_t^2 = s_T^2$. Similarly, from result (a) of Proposition 17.3, λ^2 is the asymptotic variance of the sample mean of u:

$$\sqrt{T} \cdot \bar{u} = T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{L} N(0, \lambda^2). \qquad [17.6.14]$$

Recalling the discussion of the variance of the sample mean in Sections 7.2 and 10.5, this magnitude can equivalently be described as

$$\lambda^2 = \sigma^2 \cdot [\psi(1)]^2 = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j = 2\pi s_u(0), \qquad [17.6.15]$$

where γ_j is the jth autocovariance of u_t and $s_u(0)$ is the population spectrum of u_t at frequency zero. Thus, any of the estimates of this magnitude proposed in Section

10.5 might be used. For example, if only the first q autocovariances are deemed relevant, the Newey-West estimator could be used:

$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{q} [1 - j/(q + 1)]\hat{\gamma}_j, \qquad [17.6.16]$$

where

$$\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j} \qquad [17.6.17]$$

and $\hat{u}_t = y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1}$.

To summarize, under the null hypothesis that the first difference of y_t is a zero-mean covariance-stationary process, the Phillips and Perron[9] approach is to estimate equation [17.6.1] by *OLS* and use the standard *OLS* formulas to calculate $\hat{\rho}$ and its standard error $\hat{\sigma}_{\hat{\rho}}$ along with the standard error of the regression s. The jth autocovariance of $\hat{u}_t = y_t - \hat{\alpha} - \hat{\rho} y_{t-1}$ is then calculated from [17.6.17]. The resulting estimates $\hat{\gamma}_0$ and $\hat{\lambda}^2$ are then used in [17.6.8] to construct a statistic that has the same asymptotic distribution as does the variable tabulated in the case 2 section of Table B.5. The analogous adjustments to the standard *OLS* t test of $\rho = 1$ described in [17.6.12] produce a statistic that can be compared with the case 2 section of Table B.6

Example 17.6

Let $\hat{u}_t$ denote the *OLS* sample residual for the interest rate regression [17.4.37] of Example 17.4:

$$\hat{u}_t \equiv i_t - \underset{(0.112)}{0.211} - \underset{(0.019133)}{0.96691}\, i_{t-1} \qquad \text{for } t = 1, 2, \ldots, 168.$$

The estimated autocovariances of these *OLS* residuals are

$$\hat{\gamma}_0 = (1/T) \sum_{t=1}^{T} \hat{u}_t^2 = 0.630 \qquad \hat{\gamma}_1 = (1/T) \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} = 0.114$$

$$\hat{\gamma}_2 = (1/T) \sum_{t=3}^{T} \hat{u}_t \hat{u}_{t-2} = -0.162 \qquad \hat{\gamma}_3 = (1/T) \sum_{t=4}^{T} \hat{u}_t \hat{u}_{t-3} = 0.064$$

$$\hat{\gamma}_4 = (1/T) \sum_{t=5}^{T} \hat{u}_t \hat{u}_{t-4} = 0.047.$$

[9]The procedure recommended by Phillips and Perron differs slightly from that in the text. To see the relation, write the first line of [17.6.7] as

$$T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2 = [0 \quad 1] \begin{bmatrix} 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{T^{-2}\Sigma y_{t-1}^2 - T^{-3}(\Sigma y_{t-1})^2} = \frac{1}{T^{-1}[T^{-1}\Sigma y_{t-1}^2 - (T^{-1}\Sigma y_{t-1})^2]}$$

$$= \frac{1}{T^{-1}[T^{-1}\Sigma(y_{t-1} - \bar{y}_{-1})^2]},$$

where $\bar{y}_{-1} \equiv T^{-1}\Sigma y_{t-1}$ and the last equality follows from [4.A.5]. Instead of this expression, Phillips and Perron used

$$\frac{1}{T^{-2}\Sigma(y_t - \bar{y})^2}.$$

The advantage of the formula in the text is that it is trivial to calculate from the output produced by standard regression packages and the identical formula can be used for cases 1, 2, and 4.

Thus, if the serial correlation of u_t is to be described with $q = 4$ autocovariances,

$$\begin{aligned}\hat{\lambda}^2 &= 0.630 + 2(\tfrac{4}{5})(0.114) + 2(\tfrac{3}{5})(-0.162) + 2(\tfrac{2}{5})(0.064) + 2(\tfrac{1}{5})(0.047)\\ &= 0.688.\end{aligned}$$

The usual OLS formula for the variance of the residuals from this regression is

$$s^2 = (T-2)^{-1}\sum_{t=1}^{T}\hat{u}_t^2 = 0.63760.$$

Hence, the Phillips-Perron ρ statistic is

$$\begin{aligned}&T(\hat{\rho}-1) - (1/2)\cdot(T^2\cdot\hat{\sigma}_{\hat{\rho}}^2/s^2)\cdot(\hat{\lambda}^2-\hat{\gamma}_0)\\ &\quad = 168(0.96691-1) - \tfrac{1}{2}\{[(168)(0.019133)]^2/(0.63760)\}(0.688-0.630)\\ &\quad = -6.03.\end{aligned}$$

Comparing this with the 5% critical value for case 2 of Table B.5, we see that $-6.03 > -13.8$. We thus accept the null hypothesis that the interest rate data could plausibly have been generated by a simple unit root process.

Similarly, the adjustment to the t statistic from Example 17.4 described in [17.6.12] is

$$\begin{aligned}&(\hat{\gamma}_0/\hat{\lambda}^2)^{1/2}t - \{\tfrac{1}{2}(\hat{\lambda}^2-\hat{\gamma}_0)(T\cdot\hat{\sigma}_{\hat{\rho}}/s) \div \hat{\lambda}\}\\ &\quad = \{(0.630)/(0.688)\}^{1/2}(0.96691-1)/0.019133\\ &\qquad -\{(1/2)(0.688-0.630)[(168)(0.019133)/\sqrt{(0.63760)}] \div \sqrt{(0.688)}\}\\ &\quad = -1.80.\end{aligned}$$

Since $-1.80 > -2.89$, the null hypothesis of a unit root is again accepted at the 5% level.

Phillips-Perron Tests for Cases 1 and 4

The asymptotic distributions in [17.6.8] and [17.6.12] were derived under the assumption that the true process for the first difference of y_t is serially correlated with mean zero. Even though the true unit root process exhibited no drift, it was assumed that the estimated OLS regression included a constant term as in case 2 of Section 17.4.

The same ideas can be used to generalize case 1 or case 4 of Section 17.4, and the statistics [17.6.8] and [17.6.12] can be compared in each case with the corresponding values in Tables B.5 and B.6. These results are summarized in Table 17.2. The reader is invited to confirm these claims in exercises at the end of the chapter.

Example 17.7

The residuals from the GNP regression [17.4.56] have the following estimated autocovariances:

$$\begin{array}{lll}\hat{\gamma}_0 = 1.136 & \hat{\gamma}_1 = 0.424 & \hat{\gamma}_2 = 0.285\\ \hat{\gamma}_3 = 0.006 & \hat{\gamma}_4 = -0.110, & \end{array}$$

from which

$$\hat{\lambda}^2 = 1.136 + 2\{\tfrac{4}{5}(0.424) + \tfrac{3}{5}(0.285) + \tfrac{2}{5}(0.006) - \tfrac{1}{5}(0.110)\} = 2.117.$$

Also, $s^2 = 1.15627$. Thus, for these data the Phillips-Perron ρ test is

$$\begin{aligned}
&T(\hat{\rho} - 1) - \tfrac{1}{2}(T^2 \cdot \hat{\sigma}_{\hat{\rho}}^2/s^2)(\hat{\lambda}^2 - \hat{\gamma}_0) \\
&= 168(0.96252 - 1) - \tfrac{1}{2}\left\{\frac{[(168)(0.019304)]^2}{1.15627}\right\}(2.117 - 1.136) \\
&= -10.76.
\end{aligned}$$

Since $-10.76 > -21.0$, the null hypothesis that log GNP follows a unit root process with or without drift is accepted at the 5% level.

The Phillips-Perron t test is

$$\begin{aligned}
&(\hat{\gamma}_0/\hat{\lambda}^2)^{1/2} t - \{\tfrac{1}{2}(\hat{\lambda}^2 - \hat{\gamma}_0)(T \cdot \hat{\sigma}_{\hat{\rho}}/s) \div \hat{\lambda}\} \\
&= \{(1.136)/(2.117)\}^{1/2}(0.96252 - 1)/0.019304 \\
&\quad - \{\tfrac{1}{2}(2.117 - 1.136)[[(168)(0.019304)]/\sqrt{(1.15627)}] \div \sqrt{(2.117)}\} \\
&= -2.44.
\end{aligned}$$

Since $-2.44 > -3.44$, the null hypothesis of a unit root is again accepted.

More General Processes for u_t

The Newey-West estimator $\hat{\lambda}^2$ in [17.6.16] can provide a consistent estimate of λ^2 for an $MA(\infty)$ process, provided that q, the lag truncation parameter, goes to infinity as the sample size T grows, and provided that q grows sufficiently slowly relative to T. Phillips (1987) established such consistency assuming that $q_T \rightarrow \infty$ and $q_T/T^{1/4} \rightarrow 0$; for example, $q_T = A \cdot T^{1/5}$ satisfies this requirement. Phillips's results warrant using a larger value of q with a larger data set, though they do not tell us exactly how large to choose q in practice. Monte Carlo investigations have been provided by Phillips and Perron (1988), Schwert (1989), and Kim and Schmidt (1990), though no simple rule emerges from these studies. Andrews's (1991) procedures might be used in this context.

Asymptotic results can also be obtained under weaker assumptions about u_t than those in Proposition 17.3. For example, the reader may note from the proof of result 17.3(e) that the parameter γ_0 appears because it is the plim of $T^{-1} \times \Sigma_{t=1}^T u_t^2$. Under the conditions of the proposition, the law of large numbers ensures that this plim is just the expected value of u_t^2, which expected value was denoted γ_0. However, even if the data are heterogeneously distributed with $E(u_t^2) = \gamma_{0,t}$, it may still be the case that $T^{-1}\Sigma_{t=1}^T \gamma_{0,t}$ converges to some constant. If $T^{-1}\Sigma_{t=1}^T u_t^2$ also converges to this constant, then this constant plays the role of γ_0 in a generalization of result 17.3(e).

Similarly, let $\bar{u}_T$ denote the sample mean from some heterogeneously distributed process with population mean zero:

$$\bar{u}_T \equiv T^{-1} \sum_{t=1}^{T} u_t;$$

and let λ_T^2 denote T times the variance of $\bar{u}_T$:

$$\lambda_T^2 \equiv T \cdot \mathrm{Var}(\bar{u}_T) = T^{-1} \cdot E(u_1 + u_2 + \cdots + u_T)^2.$$

The sample mean $\bar{u}_T$ may still satisfy the central limit theorem:

$$T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{L} N(0, \lambda^2)$$

TABLE 17.2
Summary of Phillips-Perron Tests for Unit Roots

Case 1:

Estimated regression: $y_t = \rho y_{t-1} + u_t$

True process: $y_t = y_{t-1} + u_t$

Z_ρ has the same asymptotic distribution as the variable described under the heading Case 1 in Table B.5.

Z_t has the same asymptotic distribution as the variable described under Case 1 in Table B.6.

Case 2:

Estimated regression: $y_t = \alpha + \rho y_{t-1} + u_t$

True process: $y_t = y_{t-1} + u_t$

Z_ρ has the same asymptotic distribution as the variable described under Case 2 in Table B.5.

Z_t has the same asymptotic distribution as the variable described under Case 2 in Table B.6.

Case 4:

Estimated regression: $y_t = \alpha + \rho y_{t-1} + \delta t + u_t$

True process: $y_t = \alpha + y_{t-1} + u_t \qquad \alpha$ any

Z_ρ has the same asymptotic distribution as the variable described under Case 4 in Table B.5.

Z_t has the same asymptotic distribution as the variable described under Case 4 in Table B.6.

Notes to Table 17.2

Estimated regression indicates the form in which the regression is estimated, using observations $t = 1, 2, \ldots, T$ and conditioning on observation $t = 0$.

True process describes the null hypothesis under which the distribution is calculated. In each case, u_t is assumed to have mean zero but can be heterogeneously distributed and serially correlated with

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(u_t^2) = \gamma_0$$
$$\lim_{T\to\infty} T^{-1} E(u_1 + u_2 + \cdots + u_T)^2 = \lambda^2.$$

Z_ρ is the following statistic:

$$Z_\rho \equiv T(\hat{\rho}_T - 1) - (1/2)\{T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2\}(\hat{\lambda}_T^2 - \hat{\gamma}_{0,T}),$$

where

$$\hat{\gamma}_{j,T} = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}$$

$\hat{u}_t$ = OLS sample residual from the estimated regression

$$\hat{\lambda}_T^2 = \hat{\gamma}_{0,T} + 2 \cdot \sum_{j=1}^{q} [1 - j/(q+1)]\hat{\gamma}_{j,T}$$

$$s_T^2 = (T - k)^{-1} \sum_{t=1}^{T} \hat{u}_t^2$$

k = number of parameters in estimated regression

$\hat{\sigma}_{\hat{\rho}_T}$ = OLS standard error for $\hat{\rho}$.

Z_t is the following statistic:

$$Z_t \equiv (\hat{\gamma}_{0,T}/\hat{\lambda}_T^2)^{1/2} \cdot (\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T} - (1/2)(\hat{\lambda}_T^2 - \hat{\gamma}_{0,T})(1/\hat{\lambda}_T)\{T \cdot \hat{\sigma}_{\hat{\rho}_T} \div s_T\}.$$

or

$$T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{L} \lambda \cdot W(1),$$

where

$$\lambda^2 \equiv \lim_{T\to\infty} \lambda_T^2, \qquad [17.6.18]$$

providing a basis for generalizing result 17.3(a).

If u_t were a covariance-stationary process with absolutely summable autocovariances, then Proposition 7.5(b) would imply that $\lim_{T\to\infty} \lambda_T^2 = \Sigma_{j=-\infty}^{\infty} \gamma_j$. Recalling [7.2.8], expression [17.6.18] would in this case just be another way to describe the parameter λ^2 in Proposition 17.3.

Thus, the parameters γ_0 and λ^2 in [17.6.8] and [17.6.12] can more generally be defined as

$$\gamma_0 \equiv \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(u_t^2) \qquad [17.6.19]$$

$$\lambda^2 \equiv \lim_{T\to\infty} T^{-1} \cdot E(u_1 + u_2 + \cdots + u_T)^2. \qquad [17.6.20]$$

Phillips (1987) and Perron and Phillips (1988) derived [17.6.8] and [17.6.12] assuming that u_t is a zero-mean but otherwise heterogeneously distributed process satisfying certain restrictions on the serial dependence and higher moments. From this perspective, expressions [17.6.19] and [17.6.20] can be used as the definitions of the parameters γ_0 and λ^2. Clearly, the estimators [17.6.13] and [17.6.16] continue to be appropriate for this alternative interpretation.

On the Observational Equivalence of Unit Root and Covariance-Stationary Processes

We saw in Section 15.4 that given any $I(0)$ process for y_t and any finite sample size T, there exists an $I(1)$ process that will be impossible to distinguish from the $I(0)$ representation on the basis of the first and second sample moments of y. Yet the Phillips and Perron procedures seem to offer a way to test the null hypothesis that the sample was generated from an arbitrary $I(1)$ process. What does it mean if the test leads us to reject the null hypothesis that y_t is $I(1)$ when we know that there exists an $I(1)$ process that describes the sample arbitrarily well?

Some insight into this question can be gained by considering the example in equation [15.4.8],

$$(1 - L)y_t = (1 + \theta L)\varepsilon_t, \qquad [17.6.21]$$

where θ is slightly larger than -1 and ε_t is i.i.d. with mean zero and variance σ^2. The model [17.6.21] implies that

$$\begin{aligned} y_t &= (\varepsilon_t + \theta\varepsilon_{t-1}) + (\varepsilon_{t-1} + \theta\varepsilon_{t-2}) + \cdots + (\varepsilon_1 + \theta\varepsilon_0) + y_0 \\ &= \varepsilon_t + (1 + \theta)\varepsilon_{t-1} + (1 + \theta)\varepsilon_{t-2} + \cdots + (1 + \theta)\varepsilon_1 + \theta\varepsilon_0 + y_0 \\ &= \varepsilon_t + (1 + \theta)\xi_{t-1} + \theta\varepsilon_0 + y_0, \end{aligned}$$

where

$$\xi_{t-1} \equiv \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{t-1}.$$

For large t, the variable y_t is dominated by the unit root component, $(1 + \theta)\xi_{t-1}$, and the asymptotic results are all governed by this term. However, if θ is close to -1, then in a finite sample y_t would behave essentially like the white noise series ε_t plus a constant $(\theta\varepsilon_0 + y_0)$. In such a case the Phillips-Perron test is likely to reject the null hypothesis of a unit root in finite samples even though it is true.[10] For example, Schwert (1989) generated Monte Carlo samples of size $T = 1{,}000$ according to the unit root model [17.6.21] with $\theta = -0.8$. The Phillips-Perron test that is supposed to reject only 5% of the time actually rejected the null hypothesis in virtually every sample, even though the null hypothesis is true! Similar results were reported by Phillips and Perron (1988) and Kim and Schmidt (1990).

Campbell and Perron (1991) argued that such false rejections are not necessarily a bad thing. If θ is near -1, then for many purposes an $I(0)$ model may provide a more useful description of the process in [17.6.21] than does the true $I(1)$ model. In support of this claim, they generated samples from the process [17.6.21] and estimated by *OLS* both an autoregressive process in levels,

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

and an autoregressive process in differences,

$$\Delta y_t = \alpha + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_p \Delta y_{t-p} + \varepsilon_t.$$

They found that for θ close to -1, forecasts based on the levels y_t tended to perform better than those based on the differences Δy_t, even though the true data-generating process was $I(1)$.

A related issue, of course, arises with false acceptances. Clearly, if the true model is

$$y_t = \rho y_{t-1} + \varepsilon_t \qquad [17.6.22]$$

with ρ slightly below 1, then the null hypothesis that $\rho = 1$ is likely to be accepted in small samples, even though it is false. The value of accepting a false null hypothesis in this case is that imposing the condition $\rho = 1$ may produce a better forecast than one based on an estimated $\hat{\rho}_T$, particularly given the small-sample downward bias of $\hat{\rho}_T$. Furthermore, when ρ is close to 1, the values in Table B.6 might give a better small-sample approximation to the distribution of $(\hat{\rho}_T - 1) \div \hat{\sigma}_{\hat{\rho}_T}$ than the traditional t tables.[11]

This discussion underscores that the goal of unit root tests is to find a parsimonious representation that gives a reasonable approximation to the true process, as opposed to determining whether or not the true process is literally $I(1)$.

17.7. *Asymptotic Properties of a* p*th-Order Autoregression and the Augmented Dickey-Fuller Tests for Unit Roots*

The Phillips-Perron tests were based on simple *OLS* regressions of y_t on its own lagged value and possibly a constant or time trend as well. Corrections for serial correlation were then made to the standard *OLS* coefficient and t statistics. This section discusses an alternative approach, due to Dickey and Fuller (1979), which controls for serial correlation by including higher-order autoregressive terms in the regression.

[10] For more detailed discussion, see Phillips and Perron (1988, p. 344).

[11] See Evans and Savin (1981, 1984) for a description of the small-sample distributions.

An Alternative Representation of an AR(p) Process

Suppose that the data were really generated from an $AR(p)$ process,

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) y_t = \varepsilon_t, \qquad [17.7.1]$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence with mean zero, variance σ^2, and finite fourth moment. It is helpful to write the autoregression [17.7.1] in a slightly different form. To do so, define

$$\rho \equiv \phi_1 + \phi_2 + \cdots + \phi_p \qquad [17.7.2]$$

$$\zeta_j \equiv -[\phi_{j+1} + \phi_{j+2} + \cdots + \phi_p] \qquad \text{for } j = 1, 2, \ldots, p-1. \qquad [17.7.3]$$

Notice that for any values of $\phi_1, \phi_2, \ldots, \phi_p$, the following polynomials in L are equivalent:

$$\begin{aligned}
&(1 - \rho L) - (\zeta_1 L + \zeta_2 L^2 + \cdots + \zeta_{p-1} L^{p-1})(1 - L) \\
&= 1 - \rho L - \zeta_1 L + \zeta_1 L^2 - \zeta_2 L^2 + \zeta_2 L^3 - \cdots - \zeta_{p-1} L^{p-1} + \zeta_{p-1} L^p \\
&= 1 - (\rho + \zeta_1) L - (\zeta_2 - \zeta_1) L^2 - (\zeta_3 - \zeta_2) L^3 - \cdots \\
&\quad - (\zeta_{p-1} - \zeta_{p-2}) L^{p-1} - (-\zeta_{p-1}) L^p \\
&= 1 - [(\phi_1 + \phi_2 + \cdots + \phi_p) - (\phi_2 + \phi_3 + \cdots + \phi_p)] L \\
&\quad - [-(\phi_3 + \phi_4 + \cdots + \phi_p) + (\phi_2 + \phi_3 + \cdots + \phi_p)] L^2 - \cdots \\
&\quad - [-(\phi_p) + (\phi_{p-1} + \phi_p)] L^{p-1} - (\phi_p) L^p \\
&= 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_{p-1} L^{p-1} - \phi_p L^p.
\end{aligned} \qquad [17.7.4]$$

Thus, the autoregression [17.7.1] can equivalently be written

$$\{(1 - \rho L) - (\zeta_1 L + \zeta_2 L^2 + \cdots + \zeta_{p-1} L^{p-1})(1 - L)\} y_t = \varepsilon_t \qquad [17.7.5]$$

or

$$y_t = \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t. \qquad [17.7.6]$$

Suppose that the process that generated y_t contains a single unit root; that is, suppose one root of

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = 0 \qquad [17.7.7]$$

is unity,

$$1 - \phi_1 - \phi_2 - \cdots - \phi_p = 0, \qquad [17.7.8]$$

and all other roots of [17.7.7] are outside the unit circle. Notice that [17.7.8] implies that the coefficient ρ in [17.7.2] is unity. Moreover, when $\rho = 1$, expression [17.7.4] would imply

$$\begin{aligned}
&(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) \\
&\qquad = (1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1} z^{p-1})(1 - z).
\end{aligned} \qquad [17.7.9]$$

Of the p values of z that make the left side of [17.7.9] zero, one is $z = 1$ and all other roots are presumed to be outside the unit circle. The same must be true of the right side as well, meaning that all roots of

$$(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1} z^{p-1}) = 0$$

lie outside the unit circle. Under the null hypothesis that $\rho = 1$, expression [17.7.5] could then be written as

$$(1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_{p-1} L^{p-1})\, \Delta y_t = \varepsilon_t$$

or

$$\Delta y_t = u_t, \tag{17.7.10}$$

where

$$u_t = (1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_{p-1} L^{p-1})^{-1} \varepsilon_t.$$

Equation [17.7.10] indicates that y_t behaves like the variable ξ_t described in Proposition 17.3, with

$$\psi(L) = (1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_{p-1} L^{p-1})^{-1}.$$

One of the advantages of writing the autoregression of [17.7.1] in the equivalent form of [17.7.6] is that only one of the regressors in [17.7.6], namely, y_{t-1}, is $I(1)$, whereas all of the other regressors $(\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1})$ are stationary. Thus, [17.7.6] is the Sims, Stock, and Watson (1990) canonical form, originally proposed for this problem by Fuller (1976). Since no knowledge of any population parameters is needed to write the model in this canonical form, in this case it is convenient to estimate the parameters by direct *OLS* estimation of [17.7.6].

Results generalizing those for case 1 in Section 17.4 are obtained when the regression is estimated as written in [17.7.6] without a constant term. Cases 2 and 3 are generalized by including a constant term in [17.7.6], while case 4 is generalized by including a constant term and a time trend in [17.7.6]. For illustration, the case 2 regression is discussed in detail. Comparable results for case 1, case 3, and case 4 will be summarized in Table 17.3 later in this section, with details developed in exercises at the end of the chapter.

Case 2. The Estimated Autoregression Includes a Constant Term, but the Data Were Really Generated by a Unit Root Autoregression with No Drift

Following the usual notational convention for *OLS* estimation of autoregressions, we assume that the initial sample is of size $T + p$, with observations numbered $\{y_{-p+1}, y_{-p+2}, \ldots, y_T\}$, and condition on the first p observations. We are interested in the properties of *OLS* estimation of

$$\begin{aligned} y_t &= \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \varepsilon_t \\ &\equiv \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t, \end{aligned} \tag{17.7.11}$$

where $\boldsymbol{\beta} \equiv (\zeta_1, \zeta_2, \ldots, \zeta_{p-1}, \alpha, \rho)'$ and $\mathbf{x}_t \equiv (\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1}, 1, y_{t-1})'$. The deviation of the *OLS* estimate $\mathbf{b}_T$ from the true value $\boldsymbol{\beta}$ is given by

$$\mathbf{b}_T - \boldsymbol{\beta} = \left[\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right]^{-1} \left[\sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t\right]. \tag{17.7.12}$$

Letting $u_t \equiv y_t - y_{t-1}$, the individual terms in [17.7.12] are

$$\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' = \qquad [17.7.13]$$

$$\begin{bmatrix} \Sigma u_{t-1}^2 & \Sigma u_{t-1}u_{t-2} & \cdots & \Sigma u_{t-1}u_{t-p+1} & \Sigma u_{t-1} & \Sigma u_{t-1}y_{t-1} \\ \Sigma u_{t-2}u_{t-1} & \Sigma u_{t-2}^2 & \cdots & \Sigma u_{t-1}u_{t-p+1} & \Sigma u_{t-2} & \Sigma u_{t-2}y_{t-1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \Sigma u_{t-p+1}u_{t-1} & \Sigma u_{t-p+1}u_{t-2} & \cdots & \Sigma u_{t-p+1}^2 & \Sigma u_{t-p+1} & \Sigma u_{t-p+1}y_{t-1} \\ \Sigma u_{t-1} & \Sigma u_{t-2} & \cdots & \Sigma u_{t-p+1} & T & \Sigma y_{t-1} \\ \Sigma y_{t-1}u_{t-1} & \Sigma y_{t-1}u_{t-2} & \cdots & \Sigma y_{t-1}u_{t-p+1} & \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}$$

$$\sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t = \begin{bmatrix} \Sigma u_{t-1}\varepsilon_t \\ \Sigma u_{t-2}\varepsilon_t \\ \vdots \\ \Sigma u_{t-p+1}\varepsilon_t \\ \Sigma \varepsilon_t \\ \Sigma y_{t-1}\varepsilon_t \end{bmatrix} \qquad [17.7.14]$$

with Σ denoting summation over $t = 1, 2, \ldots, T$.

Under the null hypothesis that $\alpha = 0$ and $\rho = 1$, we saw in [17.7.10] that y_t behaves like $\xi_t = u_1 + u_2 + \cdots + u_t$ in Proposition 17.3. Consulting the rates of convergence in Proposition 17.3, for this case the scaling matrix should be

$$\underset{(p+1\times p+1)}{\mathbf{Y}_T} \equiv \begin{bmatrix} \sqrt{T} & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{T} & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{T} & 0 \\ 0 & 0 & \cdots & 0 & T \end{bmatrix}. \qquad [17.7.15]$$

Premultiplying [17.7.12] by $\mathbf{Y}_T$ as in [17.4.20] results in

$$\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}) = \left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]\mathbf{Y}_T^{-1}\right\}^{-1}\left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t\varepsilon_t\right]\right\}. \qquad [17.7.16]$$

Consider the matrix $\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t'\mathbf{Y}_T^{-1}$. Elements in the upper left $(p \times p)$ block of $\Sigma\mathbf{x}_t\mathbf{x}_t'$ are divided by T, the first p elements of the $(p + 1)$th row or $(p + 1)$th column are divided by $T^{3/2}$, and the row $(p + 1)$, column $(p + 1)$ element of $\Sigma\mathbf{x}_t\mathbf{x}_t'$ is divided by T^2. Moreover,

$$T^{-1}\Sigma u_{t-i}u_{t-j} \xrightarrow{p} \gamma_{|i-j|} \quad \text{from result (c) of Proposition 17.3}$$
$$T^{-1}\Sigma u_{t-j} \xrightarrow{p} E(u_{t-j}) = 0 \quad \text{from the law of large numbers}$$
$$T^{-3/2}\Sigma y_{t-1}u_{t-j} \xrightarrow{p} 0 \quad \text{from Proposition 17.3(e)}$$
$$T^{-3/2}\Sigma y_{t-1} \xrightarrow{L} \lambda\cdot\int W(r)\,dr \quad \text{from Proposition 17.3(f)}$$
$$T^{-2}\Sigma y_{t-1}^2 \xrightarrow{L} \lambda^2\cdot\int [W(r)]^2\,dr \quad \text{from Proposition 17.3(h),}$$

where

$$\begin{aligned} \gamma_j &= E\{(\Delta y_t)(\Delta y_{t-j})\} \\ \lambda &\equiv \sigma\cdot\psi(1) = \sigma/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1}) \qquad [17.7.17] \\ \sigma^2 &= E(\varepsilon_t^2) \end{aligned}$$

and the integral sign denotes integration over r from 0 to 1. Thus,

$$\mathbf{Y}_T^{-1}[\Sigma \mathbf{x}_t \mathbf{x}_t']\mathbf{Y}_T^{-1}$$

$$\xrightarrow{L} \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} & 0 & 0 \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & \lambda \cdot \int W(r)\,dr \\ 0 & 0 & \cdots & 0 & \lambda \cdot \int W(r)\,dr & \lambda^2 \cdot \int [W(r)]^2\,dr \end{bmatrix}$$

$$\equiv \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}, \qquad [17.7.18]$$

where

$$\mathbf{V} \equiv \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix} \qquad [17.7.19]$$

$$\mathbf{Q} \equiv \begin{bmatrix} 1 & \lambda \cdot \int W(r)\,dr \\ \lambda \cdot \int W(r)\,dr & \lambda^2 \cdot \int [W(r)]^2\,dr \end{bmatrix} \qquad [17.7.20]$$

Next, consider the second term in [17.7.16],

$$\mathbf{Y}_T^{-1}[\Sigma \mathbf{x}_t \varepsilon_t] = \begin{bmatrix} T^{-1/2}\Sigma u_{t-1}\varepsilon_t \\ T^{-1/2}\Sigma u_{t-2}\varepsilon_t \\ \vdots \\ T^{-1/2}\Sigma u_{t-p+1}\varepsilon_t \\ T^{-1/2}\Sigma \varepsilon_t \\ T^{-1}\Sigma y_{t-1}\varepsilon_t \end{bmatrix} \qquad [17.7.21]$$

The first $p - 1$ elements of this vector are $\sqrt{T}$ times the sample mean of a martingale difference sequence whose variance-covariance matrix is

$$E\begin{bmatrix} u_{t-1}\varepsilon_t \\ u_{t-2}\varepsilon_t \\ \vdots \\ u_{t-p+1}\varepsilon_t \end{bmatrix}[u_{t-1}\varepsilon_t \quad u_{t-2}\varepsilon_t \quad \cdots \quad u_{t-p+1}\varepsilon_t]$$

$$= \sigma^2 \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix} \qquad [17.7.22]$$

$$= \sigma^2 \mathbf{V}.$$

Thus, the first $p - 1$ terms in [17.7.21] satisfy the usual central limit theorem,

$$\begin{bmatrix} T^{-1/2}\Sigma u_{t-1}\varepsilon_t \\ T^{-1/2}\Sigma u_{t-2}\varepsilon_t \\ \vdots \\ T^{-1/2}\Sigma u_{t-p+1}\varepsilon_t \end{bmatrix} \xrightarrow{L} \mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V}). \qquad [17.7.23]$$

The distribution of the last two elements in [17.7.21] can be obtained from results (a) and (d) of Proposition 17.3:

$$\begin{bmatrix} T^{-1/2}\Sigma\varepsilon_t \\ T^{-1}\Sigma y_{t-1}\varepsilon_t \end{bmatrix} \xrightarrow{L} \mathbf{h}_2 \sim \begin{bmatrix} \sigma\cdot W(1) \\ \frac{1}{2}\sigma\lambda\cdot\{[W(1)]^2 - 1\} \end{bmatrix}. \qquad [17.7.24]$$

Substituting [17.7.18] through [17.7.24] into [17.7.16] results in

$$\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2 \end{bmatrix}. \qquad [17.7.25]$$

Coefficients on Δy_{t-j}

The first $p - 1$ elements of $\boldsymbol{\beta}$ are $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$, which are the coefficients on zero-mean stationary regressors $(\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1})$. The block consisting of the first $p - 1$ elements in [17.7.25] states that

$$\sqrt{T}\begin{bmatrix} \hat{\zeta}_{1,T} - \zeta_1 \\ \hat{\zeta}_{2,T} - \zeta_2 \\ \vdots \\ \hat{\zeta}_{p-1,T} - \zeta_{p-1} \end{bmatrix} \xrightarrow{L} \mathbf{V}^{-1}\mathbf{h}_1. \qquad [17.7.26]$$

Recalling from [17.7.23] that $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V})$, it follows that $\mathbf{V}^{-1}\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V}^{-1})$, or

$$\sqrt{T}\begin{bmatrix} \hat{\zeta}_{1,T} - \zeta_1 \\ \hat{\zeta}_{2,T} - \zeta_2 \\ \vdots \\ \hat{\zeta}_{p-1,T} - \zeta_{p-1} \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \sigma^2\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix}^{-1}\right), \qquad [17.7.27]$$

where $\gamma_j = E\{(\Delta y_t)(\Delta y_{t-j})\}$.

This means that a null hypothesis involving the coefficients on the stationary regressors $(\zeta_1, \zeta_2, \ldots, \zeta_{p-1})$ in [17.7.11] can be tested in the usual way, with the standard t and F statistics asymptotically valid. To see this, suppose that the null hypothesis is H_0: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ for $\mathbf{R}$ a known $[m \times (p + 1)]$ matrix where m is the number of restrictions. The Wald form of the *OLS* χ^2 test [8.2.23] is given by

$$\begin{aligned} \chi_T^2 &= (\mathbf{R}\mathbf{b}_T - \mathbf{r})'\left\{s_T^2\mathbf{R}\left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\mathbf{R}'\right\}^{-1}(\mathbf{R}\mathbf{b}_T - \mathbf{r}) \\ &= [\mathbf{R}\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})]'\left\{s_T^2\mathbf{R}\cdot\sqrt{T}\left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\sqrt{T}\cdot\mathbf{R}'\right\}^{-1} \qquad [17.7.28] \\ &\quad \times [\mathbf{R}\cdot\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})], \end{aligned}$$

where

$$
\begin{aligned}
s_T^2 = {} & [T - (p + 1)]^{-1} \sum_{t=1}^{T} (y_t - \hat{\zeta}_{1,T}\Delta y_{t-1} - \hat{\zeta}_{2,T}\Delta y_{t-2} - \cdots \\
& - \hat{\zeta}_{p-1,T}\Delta y_{t-p+1} - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2 \\
& \xrightarrow{p} E(\varepsilon_t^2) = \sigma^2.
\end{aligned} \tag{17.7.29}
$$

If none of the restrictions involves α or ρ, then the last two columns of $\mathbf{R}$ contain all zeros:

$$
\underset{[m \times (p+1)]}{\mathbf{R}} = \begin{bmatrix} \underset{[m \times (p-1)]}{\mathbf{R}_1} & \underset{(m \times 2)}{\mathbf{0}} \end{bmatrix}. \tag{17.7.30}
$$

In this case, $\mathbf{R}\sqrt{T} = \mathbf{R}\mathbf{Y}_T$ for $\mathbf{Y}_T$ the matrix in [17.7.15], so that [17.7.28] can be written as

$$
\chi_T^2 = [\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})]' \left\{ s_T^2 \mathbf{R}\mathbf{Y}_T \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t' \right]^{-1} \mathbf{Y}_T\mathbf{R}' \right\}^{-1} [\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})].
$$

From [17.7.18], [17.7.25], [17.7.29], and [17.7.30], this converges to

$$
\begin{aligned}
\chi_T^2 \xrightarrow{L} {} & \left\{ [\mathbf{R}_1 \quad \mathbf{0}] \begin{bmatrix} \mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2 \end{bmatrix} \right\}' \\
& \times \left\{ \sigma^2 [\mathbf{R}_1 \quad \mathbf{0}] \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_1' \\ \mathbf{0} \end{bmatrix} \right\}^{-1} \left\{ [\mathbf{R}_1 \quad \mathbf{0}] \begin{bmatrix} \mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2 \end{bmatrix} \right\} \\
= {} & [\mathbf{R}_1\mathbf{V}^{-1}\mathbf{h}_1]'[\sigma^2\mathbf{R}_1\mathbf{V}^{-1}\mathbf{R}_1']^{-1}[\mathbf{R}_1\mathbf{V}^{-1}\mathbf{h}_1].
\end{aligned} \tag{17.7.31}
$$

But since $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V})$, it follows that the $(m \times 1)$ vector $[\mathbf{R}_1\mathbf{V}^{-1}\mathbf{h}_1]$ is distributed $N(\mathbf{0}, [\sigma^2\mathbf{R}_1\mathbf{V}^{-1}\mathbf{R}_1'])$. Hence, expression [17.7.31] is a quadratic form in a Gaussian vector that satisfies the conditions of Proposition 8.1:

$$
\chi_T^2 \xrightarrow{L} \chi^2(m).
$$

This verifies that the usual t or F tests applied to any subset of the coefficients $\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_{p-1}$ have the standard limiting distributions.

Note, moreover, that [17.7.27] is exactly the same asymptotic distribution that would be obtained if the data were differenced before estimating the autoregression:

$$
\Delta y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1}\Delta y_{t-p+1} + \alpha + \varepsilon_t.
$$

Thus, if the goal is to estimate $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$ or test hypotheses about these coefficients, there is no need based on asymptotic distribution theory for differencing the data before estimating the autoregression. Many researchers do recommend differencing the data first, but the reason is to reduce the small-sample bias and small-sample mean squared errors of the estimates, not to change the asymptotic distribution.

Coefficients on Constant Term and y_{t-1}

The last two elements of $\boldsymbol{\beta}$ are α and ρ, which are coefficients on the constant term and the $I(1)$ regressor, y_{t-1}. From [17.7.25], [17.7.20], and [17.7.24], their limiting distribution is given by

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} 1 & \lambda \cdot \int W(r)\,dr \\ \lambda \cdot \int W(r)\,dr & \lambda^2 \cdot \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma \cdot W(1) \\ \frac{1}{2}\sigma\lambda \cdot \{[W(1)]^2 - 1\} \end{bmatrix}$$

$$= \sigma \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1} \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1} \qquad [17.7.32]$$

$$\times \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma & 0 \\ 0 & \sigma/\lambda \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \end{bmatrix}.$$

The second element of this vector implies that (λ/σ) times $T(\hat{\rho}_T - 1)$ has the same asymptotic distribution as [17.4.28], which described the estimate of ρ in a regression without lagged Δy and with serially uncorrelated disturbances:

$$T \cdot (\lambda/\sigma) \cdot (\hat{\rho}_T - 1) \xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1) \cdot \int W(r)\,dr}{\left\{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2\right\}}. \qquad [17.7.33]$$

Recall from [17.7.17] that

$$\lambda/\sigma = (1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})^{-1}. \qquad [17.7.34]$$

This magnitude is clearly estimated consistently by

$$(1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \cdots - \hat{\zeta}_{p-1,T})^{-1},$$

where $\hat{\zeta}_{j,T}$ denotes the estimate of ζ_j based on the *OLS* regression [17.7.11]. Thus, the generalization of the Dickey-Fuller ρ test when lagged changes in y are included in the regression is

$$\frac{T \cdot (\hat{\rho}_T - 1)}{1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \cdots - \hat{\zeta}_{p-1,T}} \xrightarrow{L} \frac{\frac{1}{2}\{[W(1)]^2 - 1\} - W(1) \cdot \int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2}. \qquad [17.7.35]$$

This is to be compared with the case 2 section of Table B.5.

Consider next an *OLS* t test of the null hypothesis that $\rho = 1$:

$$t_T = \frac{(\hat{\rho}_T - 1)}{\{s_T^2 \cdot \mathbf{e}_{p+1}' \cdot (\Sigma \mathbf{x}_t \mathbf{x}_t')^{-1} \cdot \mathbf{e}_{p+1}\}^{1/2}}, \qquad [17.7.36]$$

where $\mathbf{e}_{p+1}$ denotes a $[(p + 1) \times 1]$ vector with unity in the last position and zeros elsewhere. Multiplying the numerator and denominator of [17.7.36] by T results in

$$t_T = \frac{T(\hat{\rho}_T - 1)}{\left\{s_T^2 \cdot \mathbf{e}'_{p+1}\mathbf{Y}_T(\Sigma \mathbf{x}_t\mathbf{x}'_t)^{-1}\mathbf{Y}_T\mathbf{e}_{p+1}\right\}^{1/2}}. \qquad [17.7.37]$$

But

$$\begin{aligned}
\mathbf{e}'_{p+1}\mathbf{Y}_T(\Sigma \mathbf{x}_t\mathbf{x}'_t)^{-1}\mathbf{Y}_T\mathbf{e}_{p+1} &= \mathbf{e}'_{p+1}\left\{\mathbf{Y}_T^{-1}(\Sigma \mathbf{x}_t\mathbf{x}'_t)\mathbf{Y}_T^{-1}\right\}^{-1}\mathbf{e}_{p+1} \\
&\xrightarrow{L} \mathbf{e}'_{p+1}\begin{bmatrix} \mathbf{V}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1} \end{bmatrix}\mathbf{e}_{p+1} \\
&= \frac{1}{\lambda^2 \cdot \left\{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2\right\}}
\end{aligned}$$

by virtue of [17.7.18] and [17.7.20]. Hence, from [17.7.37] and [17.7.33],

$$\begin{aligned}
t_T &\xrightarrow{L} (\sigma/\lambda)\frac{\tfrac{1}{2}\{[W(1)]^2 - 1\} - W(1)\cdot\int W(r)\,dr}{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2} \\
&\quad \div \left\{\frac{\sigma^2}{\lambda^2\left\{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2\right\}}\right\}^{1/2} \qquad [17.7.38] \\
&= \frac{\tfrac{1}{2}\{[W(1)]^2 - 1\} - W(1)\cdot\int W(r)\,dr}{\left\{\int [W(r)]^2\,dr - \left[\int W(r)\,dr\right]^2\right\}^{1/2}}.
\end{aligned}$$

This is the same distribution as in [17.4.36]. Thus, the usual t test of $\rho = 1$ for *OLS* estimation of [17.7.11] can be compared with the case 2 section of Table B.6 without any corrections for the fact that lagged values of Δy are included in the regression.

A similar result applies to the Dickey-Fuller F test of the joint hypothesis that $\alpha = 0$ and $\rho = 1$. This null hypothesis can be represented as $\mathbf{R\beta} = \mathbf{r}$, where

$$\underset{[2\times(p+1)]}{\mathbf{R}} = \begin{bmatrix} \underset{[2\times(p-1)]}{\mathbf{0}} & \underset{(2\times 2)}{\mathbf{I}_2} \end{bmatrix}$$

and $\mathbf{r} = (0, 1)'$. The F test is then

$$F_T = (\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{R}'\left\{s_T^2 \cdot \mathbf{R}(\Sigma \mathbf{x}_t\mathbf{x}'_t)^{-1}\mathbf{R}'\right\}^{-1}\mathbf{R}(\mathbf{b}_T - \boldsymbol{\beta})/2. \qquad [17.7.39]$$

Define $\tilde{\mathbf{Y}}_T$ to be the following (2×2) matrix:

$$\tilde{\mathbf{Y}}_T \equiv \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}. \qquad [17.7.40]$$

Notice that [17.7.39] can be written

$$F_T = (\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{R}'\tilde{\mathbf{Y}}_T\left\{s_T^2\cdot\tilde{\mathbf{Y}}_T\mathbf{R}(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{R}'\tilde{\mathbf{Y}}_T\right\}^{-1} \times \tilde{\mathbf{Y}}_T\mathbf{R}(\mathbf{b}_T - \boldsymbol{\beta})/2. \qquad [17.7.41]$$

The matrix in [17.7.40] has the property that

$$\tilde{\mathbf{Y}}_T\mathbf{R} = \mathbf{R}\mathbf{Y}_T$$

for $\mathbf{R} = [\mathbf{0} \quad \mathbf{I}_2]$ and $\mathbf{Y}_T$ the $(p + 1) \times (p + 1)$ matrix in [17.7.15]. From [17.7.25], $\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} \mathbf{Q}^{-1}\mathbf{h}_2$. Thus, [17.7.41] implies that

$$\begin{aligned}
F_T &= (\mathbf{b}_T - \boldsymbol{\beta})'(\mathbf{R}\mathbf{Y}_T)'\left\{s_T^2\cdot\mathbf{R}\mathbf{Y}_T(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{Y}_T\mathbf{R}'\right\}^{-1}\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})/2 \\
&\xrightarrow{L} (\mathbf{Q}^{-1}\mathbf{h}_2)'\{\sigma^2\mathbf{Q}^{-1}\}^{-1}(\mathbf{Q}^{-1}\mathbf{h}_2)/2 \\
&= \mathbf{h}_2'\mathbf{Q}^{-1}\mathbf{h}_2/(2\sigma^2) \\
&= [1/(2\sigma^2)]\begin{bmatrix}\sigma\cdot W(1) & \tfrac{1}{2}\sigma\lambda\{[W(1)]^2 - 1\}\end{bmatrix} \\
&\quad \times \begin{bmatrix} 1 & \lambda\cdot\int W(r)\,dr \\ \lambda\cdot\int W(r)\,dr & \lambda^2\cdot\int [W(r)]^2\,dr\end{bmatrix}^{-1}\begin{bmatrix}\sigma\cdot W(1) \\ \tfrac{1}{2}\sigma\lambda\{[W(1)]^2 - 1\}\end{bmatrix} \\
&= \left(\frac{1}{2\sigma^2}\right)\sigma^2\begin{bmatrix}W(1) & \frac{1}{2}\{[W(1)]^2 - 1\}\end{bmatrix}\begin{bmatrix}1 & 0 \\ 0 & \lambda\end{bmatrix} \\
&\quad \times \begin{bmatrix}1 & 0 \\ 0 & \lambda\end{bmatrix}^{-1}\begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr\end{bmatrix}^{-1}\begin{bmatrix}1 & 0 \\ 0 & \lambda\end{bmatrix}^{-1} \\
&\quad \times \begin{bmatrix}1 & 0 \\ 0 & \lambda\end{bmatrix}\begin{bmatrix}W(1) \\ \tfrac{1}{2}\{[W(1)]^2 - 1\}\end{bmatrix} \\
&= \tfrac{1}{2}\begin{bmatrix}W(1) & \tfrac{1}{2}\{[W(1)]^2 - 1\}\end{bmatrix} \\
&\quad \times \begin{bmatrix} 1 & \int W(r)\,dr \\ \int W(r)\,dr & \int [W(r)]^2\,dr\end{bmatrix}^{-1}\begin{bmatrix}W(1) \\ \tfrac{1}{2}\{[W(1)]^2 - 1\}\end{bmatrix}.
\end{aligned} \qquad [17.7.42]$$

This is identical to the asymptotic distribution of the F test when the regression does not include lagged Δy and the disturbances are i.i.d. Thus, the F statistic in [17.7.41] based on OLS estimation of [17.7.11] can be compared with the case 2 section of Table B.7 without corrections.

Finally, consider a hypothesis test involving a restriction[12] across $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$ and ρ,

$$H_0\text{: } r_1\zeta_1 + r_2\zeta_2 + \cdots + r_{p-1}\zeta_{p-1} + 0\cdot\alpha + r_{p+1}\rho = r$$

[12]Since the maintained assumption is that $\rho = 1$, this is a slightly unnatural way to write a hypothesis. Nevertheless, framing the hypothesis this way will shortly prove useful in deriving the asymptotic distribution of an autoregression estimated in the usual form without the Dickey-Fuller transformation.

or

$$\mathbf{r}'\boldsymbol{\beta} = r. \qquad [17.7.43]$$

The distribution of the t test of this hypothesis will be dominated asymptotically by the parameters with the slowest rate of convergence, namely, $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$. Since these are asymptotically Gaussian, the test statistic is asymptotically Gaussian and so can be compared with the usual t tables. To demonstrate this formally, note that the usual t statistic for testing this hypothesis is

$$t_T = \frac{\mathbf{r}'\mathbf{b}_T - r}{\left\{s_T^2\mathbf{r}'(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{r}\right\}^{1/2}} = \frac{T^{1/2}(\mathbf{r}'\mathbf{b}_T - r)}{\left\{s_T^2 T^{1/2}\mathbf{r}'(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{r}T^{1/2}\right\}^{1/2}}. \qquad [17.7.44]$$

Define $\tilde{\mathbf{r}}_T$ to be the vector that results when the last element of $\mathbf{r}$ is replaced by $r_{p+1}/\sqrt{T}$,

$$\tilde{\mathbf{r}}_T' \equiv [r_1 \quad r_2 \quad \cdots \quad r_{p-1} \quad 0 \quad r_{p+1}/\sqrt{T}], \qquad [17.7.45]$$

and notice that

$$T^{1/2}\mathbf{r} = \mathbf{Y}_T\tilde{\mathbf{r}}_T \qquad [17.7.46]$$

for $\mathbf{Y}_T$ the matrix in [17.7.15]. Using [17.7.46] and the null hypothesis that $r = \mathbf{r}'\boldsymbol{\beta}$, expression [17.7.44] can be written

$$t_T = \frac{\tilde{\mathbf{r}}_T'\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})}{\left\{s_T^2\tilde{\mathbf{r}}_T'\mathbf{Y}_T(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{Y}_T\tilde{\mathbf{r}}_T\right\}^{1/2}}. \qquad [17.7.47]$$

Notice from [17.7.45] that

$$\tilde{\mathbf{r}}_T \to \tilde{\mathbf{r}},$$

where

$$\tilde{\mathbf{r}}' \equiv [r_1 \quad r_2 \quad \cdots \quad r_{p-1} \quad 0 \quad 0].$$

Using this result along with [17.7.18] and [17.7.25] in [17.7.47] produces

$$\begin{aligned} t_T \xrightarrow{L} & \frac{\tilde{\mathbf{r}}'\begin{bmatrix}\mathbf{V}^{-1}\mathbf{h}_1\\ \mathbf{Q}^{-1}\mathbf{h}_2\end{bmatrix}}{\left\{\sigma^2\tilde{\mathbf{r}}'\begin{bmatrix}\mathbf{V}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}^{-1}\end{bmatrix}\tilde{\mathbf{r}}\right\}^{1/2}} \qquad [17.7.48] \\ &= \frac{[r_1 \quad r_2 \quad \cdots \quad r_{p-1}]\mathbf{V}^{-1}\mathbf{h}_1}{\{\sigma^2[r_1 \quad r_2 \quad \cdots \quad r_{p-1}]\mathbf{V}^{-1}[r_1 \quad r_2 \quad \cdots \quad r_{p-1}]'\}^{1/2}}. \end{aligned}$$

Since $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V})$, it follows that

$$[r_1 \quad r_2 \quad \cdots \quad r_{p-1}]\mathbf{V}^{-1}\mathbf{h}_1 \sim N(0, h),$$

where

$$h = \sigma^2[r_1 \quad r_2 \quad \cdots \quad r_{p-1}]\mathbf{V}^{-1}[r_1 \quad r_2 \quad \cdots \quad r_{p-1}]'.$$

Thus, the limiting distribution in [17.7.48] is that of a Gaussian scalar divided by its standard deviation and is therefore $N(0, 1)$. This confirms the claim that the t test of $\mathbf{r}'\boldsymbol{\beta} = r$ can be compared with the usual t tables.

One interesting implication of this last result concerns the asymptotic properties of the estimated coefficients if the autoregression is estimated in the usual levels form rather than the transformed regression [17.7.11]. Thus, suppose that the following specification is estimated by OLS:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \qquad [17.7.49]$$

for some $p \geq 2$. Recalling [17.7.2] and [17.7.3], the relation between the estimates $(\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_{p-1}, \hat{\rho})$ investigated previously and estimates $(\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p)$ based on OLS estimation of [17.7.49] is

$$\begin{aligned} \hat{\phi}_p &= -\hat{\zeta}_{p-1} \\ \hat{\phi}_j &= \hat{\zeta}_j - \hat{\zeta}_{j-1} \qquad \text{for } j = 2, 3, \ldots, p-1 \\ \hat{\phi}_1 &= \hat{\rho} + \hat{\zeta}_1. \end{aligned}$$

Thus, each of the coefficients $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ is a linear combination of the elements of $(\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_{p-1}, \hat{\rho})$. The analysis of [17.7.43] establishes that any individual estimate $\hat{\phi}_j$ converges at rate $\sqrt{T}$ to a Gaussian random variable. Recalling the discussion of [16.3.20] and [16.3.21], an OLS t or F test based on [17.7.49] is numerically identical to the equivalent t or F test expressed in terms of the representation in [17.7.11]. Thus, the usual t tests associated with hypotheses about any individual coefficients $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ in [17.7.49] can be compared with standard t or $N(0, 1)$ tables. Indeed, any hypothesis about linear combinations of the ϕ's other than the sum $\phi_1 + \phi_2 + \cdots + \phi_p$ satisfies the standard conditions. The sum $\hat{\phi}_1 + \hat{\phi}_2 + \cdots + \hat{\phi}_p$, of course, has the nonstandard distribution of the estimate $\hat{\rho}$ described in [17.7.33].

Summary of Asymptotic Results for an Estimated Autoregression That Includes a Constant Term

The preceding analysis applies to OLS estimation of

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \varepsilon_t$$

under the assumption that the true value of α is zero and the true value of ρ is 1. The other maintained assumptions were that ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment and that roots of

$$(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1} z^{p-1}) = 0$$

are outside the unit circle. It was seen that the estimates $\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_{p-1}$ converge at rate $\sqrt{T}$ to Gaussian variates, and standard t or F tests for hypotheses about these coefficients have the usual limiting Gaussian or χ^2 distributions. The estimates $\hat{\alpha}$ and $\hat{\rho}$ converge at rates $\sqrt{T}$ and T, respectively, to nonstandard distributions. If the difference between the OLS estimate $\hat{\rho}$ and the hypothesized true value of unity is multiplied by the sample size and divided by $(1 - \hat{\zeta}_1 - \hat{\zeta}_2 - \cdots - \hat{\zeta}_{p-1})$, the resulting statistic has the same asymptotic distribution as the variable tabulated in the case 2 section of Table B.5. The usual t statistic of the hypothesis $\rho = 1$ does not need to be adjusted for sample size or serial correlation and has the same asymptotic distribution as the variable tabulated in the case 2 section of Table B.6. The usual F statistic of the joint hypothesis $\alpha = 0$ and $\rho = 1$ likewise does not have to be adjusted for sample size or serial correlation and has the same distribution as the variable tabulated in the case 2 section of Table B.7.

When the autoregression includes lagged changes as here, tests for a unit root based on the value of ρ, t tests, or F tests are described as *augmented Dickey-Fuller tests*.

Example 17.8

The following model was estimated by OLS for the interest rate data described in Example 17.3 (standard errors in parentheses):

$$i_t = \underset{(0.0788)}{0.335}\ \Delta i_{t-1} - \underset{(0.0808)}{0.388}\ \Delta i_{t-2} + \underset{(0.0800)}{0.276}\ \Delta i_{t-3}$$
$$- \underset{(0.0794)}{0.107}\ \Delta i_{t-4} + \underset{(0.109)}{0.195} + \underset{(0.018604)}{0.96904}\ i_{t-1}.$$

Dates t = 1948:II through 1989:I were used for estimation, so in this case the sample size is T = 164. For these estimates, the augmented Dickey-Fuller ρ test [17.7.35] would be

$$\frac{164}{1 - 0.335 + 0.388 - 0.276 + 0.107}(0.96904 - 1) = -5.74.$$

Since $-5.74 > -13.8$, the null hypothesis that the Treasury bill rate follows a fifth-order autoregression with no constant term, and a single unit root, is accepted at the 5% level. The OLS t test for this same hypothesis is

$$(0.96904 - 1)/(0.018604) = -1.66.$$

Since $-1.66 > -2.89$, the null hypothesis of a unit root is accepted by the augmented Dickey-Fuller t test as well. Finally, the OLS F test of the joint null hypothesis that $\rho = 1$ and $\alpha = 0$ is 1.65. Since this is less than 4.68, the null hypothesis is again accepted.

The null hypothesis that the autoregression in levels requires only four lags is based on the OLS t test of $\zeta_4 = 0$:

$$-0.107/0.0794 = -1.35.$$

From Table B.3, the 5% two-sided critical value for a t variable with 158 degrees of freedom is -1.98. Since $-1.35 > -1.98$, the null hypothesis that only four lags are needed for the autoregression in levels is accepted.

Asymptotic Results for Other Autoregressions

Up to this point in this section, we have considered an autoregression that is a generalization of case 2 of Section 17.4—a constant is included in the estimated regression, though the population process is presumed to exhibit no drift. Parallel generalizations for cases 1, 3, and 4 can be obtained in the same fashion. The reader is invited to derive these generalizations in exercises at the end of the chapter. The key results are summarized in Table 17.3.

TABLE 17.3
Summary of Asymptotic Results for Autoregressions Containing a Unit Root

Case 1:

Estimated regression:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1}\Delta y_{t-p+1} + \rho y_{t-1} + \varepsilon_t$$

True process: same specification as estimated regression with $\rho = 1$

Any t or F test involving $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$ can be compared with the usual t or F tables for an asymptotically valid test.

Z_{DF} has the same asymptotic distribution as the variable described under the heading Case 1 in Table B.5.

OLS t test of $\rho = 1$ has the same asymptotic distribution as the variable described under Case 1 in Table B.6.

Case 2:

Estimated regression:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \varepsilon_t$$

True process: same specification as estimated regression with $\alpha = 0$ and $\rho = 1$

Any t or F test involving $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$ can be compared with the usual t or F tables for an asymptotically valid test.

Z_{DF} has the same asymptotic distribution as the variable described under Case 2 in Table B.5.

OLS t test of $\rho = 1$ has the same asymptotic distribution as the variable described under Case 2 in Table B.6.

OLS F test of joint hypothesis that $\alpha = 0$ and $\rho = 1$ has the same asymptotic distribution as the variable described under Case 2 in Table B.7.

Case 3:

Estimated regression:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \varepsilon_t$$

True process: same specification as estimated regression with $\alpha \neq 0$ and $\rho = 1$

$\hat{\rho}_T$ converges at rate $T^{3/2}$ to a Gaussian variable; all other estimated coefficients converge at rate $T^{1/2}$ to Gaussian variables.

Any t or F test involving any coefficients from the regression can be compared with the usual t or F tables for an asymptotically valid test.

Case 4:

Estimated regression:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \delta t + \varepsilon_t$$

True process: same specification as estimated regression with α any value, $\rho = 1$, and $\delta = 0$

Any t or F test involving $\zeta_1, \zeta_2, \ldots, \zeta_{p-1}$ can be compared with the usual t or F tables for an asymptotically valid test.

Z_{DF} has the same asymptotic distribution as the variable described under Case 4 in Table B.5.

OLS t test of $\rho = 1$ has the same asymptotic distribution as the variable described under Case 4 in Table B.6.

OLS F test of joint hypothesis that $\rho = 1$ and $\delta = 0$ has the same asymptotic distribution as the variable described under Case 4 in Table B.7.

Notes to Table 17.3

Estimated regression indicates the form in which the regression is estimated, using observations $t = 1, 2, \ldots, T$ and conditioning on observations $t = 0, -1, \ldots, -p + 1$.

True process describes the null hypothesis under which the distribution is calculated. In each case it is assumed that roots of

$$(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1} z^{p-1}) = 0$$

are all outside the unit circle and that ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment.

Z_{DF} in each case is the following statistic:

$$Z_{DF} \equiv T(\hat{\rho}_T - 1)/(1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \cdots - \hat{\zeta}_{p-1,T}),$$

where $\hat{\rho}_T, \hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{p-1,T}$ are the *OLS* estimates from the indicated regression.

OLS t test of $\rho = 1$ is $(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T}$, where $\hat{\sigma}_{\hat{\rho}_T}$ is the *OLS* standard error of $\hat{\rho}_T$.

OLS F test of a hypothesis involving two restrictions is given by expression [17.7.39].

Example 17.9

The following autoregression was estimated by *OLS* for the GNP data in Figure 17.3 (standard errors in parentheses):

$$\begin{aligned} y_t = {} & \underset{(0.0777)}{0.329}\,\Delta y_{t-1} + \underset{(0.0813)}{0.209}\,\Delta y_{t-2} - \underset{(0.0818)}{0.084}\,\Delta y_{t-3} \\ & - \underset{(0.0788)}{0.075}\,\Delta y_{t-4} + \underset{(13.57)}{35.92} + \underset{(0.019386)}{0.94969}\,y_{t-1} + \underset{(0.0152)}{0.0378}\,t. \end{aligned}$$

Here, $T = 164$ and the augmented Dickey-Fuller ρ test is

$$\frac{164}{1 - 0.329 - 0.209 + 0.084 + 0.075}(0.94969 - 1) = -13.3.$$

Since $-13.3 > -21.0$, the null hypothesis that the log of GNP is $ARIMA(4, 1, 0)$ with possible drift is accepted at the 5% level. The augmented Dickey-Fuller t test also accepts this hypothesis:

$$(0.94969 - 1)/(0.019386) = -2.60 > -3.44.$$

The *OLS* F test of the joint null hypothesis that $\rho = 1$ and $\delta = 0$ is $3.74 < 6.42$, and so the augmented Dickey-Fuller F test is also consistent with the unit root specification.

Unit Root AR(p) *Processes with* p *Unknown*

Various suggestions have been proposed for how to proceed when the process is regarded as $ARIMA(p, 1, 0)$ with p unknown but finite. One simple approach is to estimate [17.7.11] with p taken to be some prespecified upper bound $\bar{p}$. The *OLS* t test of $\zeta_{\bar{p}-1} = 0$ can then be compared with the usual critical value for a t statistic from Table B.3. If the null hypothesis is accepted, the *OLS* F test of the joint null hypothesis that both $\zeta_{\bar{p}-1} = 0$ and $\zeta_{\bar{p}-2} = 0$ can be compared with the usual $F(2, T - k)$ distribution in Table B.4. The procedure continues sequentially until the joint null hypothesis that $\zeta_{\bar{p}-1} = 0, \zeta_{\bar{p}-2} = 0, \ldots, \zeta_{\bar{p}-\ell} = 0$ is rejected for some ℓ. The recommended regression is then

$$y_t = \zeta_1\Delta y_{t-1} + \zeta_2\Delta y_{t-2} + \cdots + \zeta_{\bar{p}-\ell}\Delta y_{t-\bar{p}+\ell} + \alpha + \rho y_{t-1} + \delta t.$$

If no value of ℓ leads to rejection, the simple Dickey-Fuller test of Table 17.1 is used. Hall (1991) discussed a variety of alternative strategies for estimating p.

Just as in the Phillips-Perron consideration of the $MA(\infty)$ case, the researcher might want to choose bigger values for p, the autoregressive lag length, the larger is the sample size T. Said and Dickey (1984) showed that as long as p goes to infinity sufficiently slowly relative to T, then the *OLS* t test of $\rho = 1$ can continue to be compared with the Dickey-Fuller values in Table B.6.

Again, it is worthwhile to keep in mind that there always exists a p such that an $ARIMA(p, 1, 0)$ representation can describe a stationary process arbitrarily well for a given sample. The Said-Dickey test of $\rho = 1$ might therefore best be viewed as follows. For a given fixed p, we can certainly ask whether an $ARIMA(p - 1, 1, 0)$ describes the data nearly as well as an $ARIMA(p, 0, 0)$. Imposing $\rho = 1$ when the true value of ρ is close to unity may improve forecasts and small-sample estimates of the other parameters. The Said-Dickey result permits the researcher to use a larger value of p on which to base this comparison the larger is the sample size T.

17.8. *Other Approaches to Testing for Unit Roots*

This section briefly describes some alternative approaches to testing for unit roots.

Variance Ratio Tests

Let

$$\Delta y_t = \alpha + u_t,$$

where

$$u_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \equiv \psi(L)\varepsilon_t$$

for ε_t a white noise sequence with variance σ^2. Recall from expression [15.3.10] that the permanent effect of ε_t on the level of y_{t+s} is given by

$$\lim_{s\to\infty} \frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi(1).$$

If y_t is stationary or stationary around a deterministic time trend, an innovation ε_t has no permanent effect on y, requiring $\psi(1) = 0$.

Cochrane (1988) and Lo and MacKinlay (1988) proposed a test for unit roots that exploits this property. Consider the change in y over s periods,

$$y_{t+s} - y_t = \alpha s + u_{t+s} + u_{t+s-1} + \cdots + u_{t+1}, \qquad [17.8.1]$$

and notice that

$$(y_{t+s} - y_t)/s = \alpha + s^{-1}(u_{t+s} + u_{t+s-1} + \cdots + u_{t+1}). \qquad [17.8.2]$$

The second term in [17.8.2] could be viewed as the sample mean of s observations drawn from the process followed by u. Thus, Proposition 7.5(b) and result [7.2.8] imply that

$$\lim_{s\to\infty} s\cdot \mathrm{Var}[s^{-1}(u_{t+s} + u_{t+s-1} + \cdots + u_{t+1})] = \sigma^2\cdot[\psi(1)]^2. \qquad [17.8.3]$$

Let $\hat{\alpha}_T$ denote the average change in y in a sample of T observations:

$$\hat{\alpha}_T = T^{-1} \sum_{t=1}^{T} (y_t - y_{t-1}).$$

Consider the following estimate of the variance of the change in y over its value s periods earlier:

$$\hat{J}_T(s) = T^{-1} \sum_{t=0}^{T-s} (y_{t+s} - y_t - \hat{\alpha}_T s)^2. \qquad [17.8.4]$$

This should converge in probability to

$$J(s) = E(y_{t+s} - y_t - \alpha s)^2 = E(u_{t+s} + u_{t+s-1} + \cdots + u_{t+1})^2 \qquad [17.8.5]$$

as the sample size T becomes large. Comparing this expression with [17.8.3],

$$\lim_{s\to\infty} s^{-1}\cdot J(s) = \sigma^2\cdot[\psi(1)]^2.$$

Cochrane (1988) therefore proposed calculating [17.8.4] as a function of s. If the true process for y_t is stationary or stationary around a deterministic time trend, this statistic should go to zero for large s. If the true process for y_t is $I(1)$, this statistic gives a measure of the quantitative importance of permanent effects of ε as reflected in the long-run multiplier $\psi(1)$. However, the statistic in [17.8.4] is not reliable unless s is much smaller than T.

If the data truly followed a random walk so that $\psi(L) = 1$, then $J(s)$ in [17.8.5] would equal $s \cdot \sigma^2$ for any s, where σ^2 is the variance of u_t. Lo and MacKinlay (1988) exploited this property to suggest tests of the random walk hypothesis based on alternative values of s. See Lo and MacKinlay (1989) and Cecchetti and Lam (1991) for evidence on the small-sample properties of these tests.

Other Tests for Unit Roots

The Phillips-Perron approach was based on an $MA(\infty)$ representation for Δy_t, while the Said-Dickey approach was based on an $AR(\infty)$ representation. Tests based on a finite $ARMA(p, q)$ representation for Δy_t have been explored by Said and Dickey (1985), Hall (1989), Said (1991), and Pantula and Hall (1991).

A number of other approaches to testing for unit roots have been proposed, including Sargan and Bhargava (1983), Solo (1984), Bhargava (1986), Dickey and Pantula (1987), Park and Choi (1988), Schmidt and Phillips (1992), Stock (1991), and Kwiatkowski, Phillips, Schmidt, and Shin (1992). See Stock (1993) for an excellent survey. Asymptotic inference for processes with near unit root behavior has been discussed by Chan and Wei (1987), Phillips (1988), and Sowell (1990).

17.9. *Bayesian Analysis and Unit Roots*

Up to this point in the chapter we have adopted a classical statistical perspective, calculating the distribution of $\hat{\rho}$ conditional on a particular value of ρ such as $\rho = 1$. This section considers the Bayesian perspective, in which the true value of ρ is regarded as a random variable and the goal is to describe the distribution of this random variable conditional on the data.

Recall from Proposition 12.3 that if the prior density for the vector of unknown coefficients $\boldsymbol{\beta}$ and innovation precision σ^{-2} is of the Normal-gamma form of [12.1.19] and [12.1.20], then the posterior distribution of $\boldsymbol{\beta}$ conditional on the data is multivariate t. This result holds exactly for any finite sample and holds regardless of whether the process is stationary. Thus, in the case of the diffuse prior distribution represented by $N = \lambda = 0$ and $\mathbf{M}^{-1} = \mathbf{0}$, a Bayesian would essentially use the usual t and F statistics in the standard way.

How can the classical distribution of $\hat{\rho}$ be strongly skewed while the Bayesian distribution of ρ is that of a symmetric t variable? Sims (1988) and Sims and Uhlig (1991) provided a detailed discussion of this question. The classical test of the null hypothesis $\rho = 1$ is based only on the distribution of $\hat{\rho}$ when the true value of ρ is unity. By contrast, the Bayesian inference is based on the distribution of $\hat{\rho}|\rho$ for all the possible values of ρ, with the distribution of $\hat{\rho}|\rho$ weighted according to the prior probability for ρ. If the distribution of $\hat{\rho}|\rho$ had the same skew and dispersion for every ρ as it does at $\rho = 1$, then we would conclude that, having observed any particular $\hat{\rho}$, the true value of ρ is probably somewhat higher. However, the distribution of $\hat{\rho}|\rho$ changes with ρ—the lower the true value of ρ, the smaller the skew and the greater the dispersion, since from [17.1.3] the variance of $\sqrt{T}(\hat{\rho} - \rho)$ is approximately $(1 - \rho^2)$. Because lower values of ρ imply greater dispersion for $\hat{\rho}$, in the absence of skew we would suspect that a given observation $\hat{\rho} = 0.95$ was more likely to have been generated by a distribution centered at $\rho = 0.90$ with large dispersion than by a distribution centered at $\rho = 1$ with small dispersion. The effects of skew and dispersion turn out to cancel, so that with a uniform prior distribution for the value of ρ, having observed $\hat{\rho} = 0.95$, it is just as likely that the true value of ρ is greater than 0.95 as that the true value of ρ is less than 0.95.

Example 17.10

For the GNP regression in Example 17.9, the probability that $\rho \geq 1$ conditional on the data is the probability that a t variable with $T = 164$ degrees of freedom[13] exceeds $(1 - 0.94969)/0.019386 = 2.60$. From Table B.3, this probability is around 0.005. Hence, although the value of ρ must be large, it is unlikely to be as big as unity.

The contrast between the Bayesian inference in Example 17.10 and the classical inference in Example 17.9 is one of the reasons given by Sims (1988) and Sims and Uhlig (1991) for preferring Bayesian methods. Note that the probability calculated in Example 17.10 will be less than 0.025 if and only if a classical 95% confidence interval around the point estimate $\hat{\rho}$ does not contain unity. Thus, an alternative way of describing the finding of Example 17.10 is that the standard asymptotic classical confidence region around $\hat{\rho}$ does not include $\rho = 1$. Even so, Example 17.9 showed that the null hypothesis of a unit root is accepted by the augmented Dickey-Fuller test. The classical asymptotic confidence region centered at $\rho = \hat{\rho}$ seems inconsistent with a unit root, while the classical asymptotic confidence region centered at $\rho = 1$ supports a unit root. Such disconnected confidence regions resulting from the classical approach may seem somewhat troublesome and counterintuitive.[14] By contrast, the Bayesian has a single, consistent summary of the plausibility of different values of ρ, which is that implied by the posterior distribution of ρ conditional on the data.

One could, of course, use a prior distribution that reflected more confidence in the prior information about the value of ρ. As long as the prior distribution was in the Normal-gamma class, this would cause us to shift the point estimate 0.94969 in the direction of the prior mean and reduce the standard error and increase the degrees of freedom as warranted by the prior information, but a t distribution would still be used to interpret the resulting statistic.

Although the Normal-gamma class is convenient to work with, it might not be sufficiently flexible to reflect the researcher's true prior beliefs. Sims (1988, p. 470) discussed Bayesian inference in which a point mass with positive probability is placed on the possibility that $\rho = 1$. DeJong and Whiteman (1991) used numerical methods to calculate posterior distributions under a range of prior distributions defined numerically and concluded that the evidence for unit roots in many key economic time series is quite weak.

Phillips (1991a) noted that there is a prior distribution for which the Bayesian inference mimics the classical approach. He argued that the diffuse prior distribution of Proposition 12.3 is actually highly informative in a time series regression and suggested instead a prior distribution due to Jeffreys (1946). Although this prior distribution has some theoretical arguments on its behalf, it has the unusual property in this application that the prior distribution is a function of the sample size T—Phillips would propose using a different prior distribution for $f(\rho)$ when

[13]Recall from Proposition 12.3(b) that the degrees of freedom are given by $N^* = N + T$. Thus, the Bayesian interpretation is not quite identical to the classical t statistic, whose degrees of freedom would be $T - k$.

[14]Stock (1991) has recently proposed a solution to this problem from the classical perspective. Another approach is to rely on the exact small-sample distribution, as advocated by Andrews (1993).

the analyst is going to obtain a sample of size 50 than when the analyst is going to obtain a sample of size 100. This would not be appropriate if the prior distribution is intended to represent the actual information available to the analyst before seeing the data. Phillips (1991b, pp. 468–69) argued that, in order to be truly uninformative, a prior distribution in this context would have this property, since the larger the true value of ρ, the more rapidly information about ρ contained in the sample $\{y_1, y_2, \ldots, y_T\}$ is going to accumulate with the sample size T. Certainly the concept of what it means for a prior distribution to be "uninformative" can be difficult and controversial.[15]

The potential difficulty in persuading others of the validity of one's prior beliefs has always been the key weakness of Bayesian statistics, and it seems unavoidable here. The best a Bayesian can do may be to take an explicit stand on the nature and strength of prior information and defend it as best as possible. If the nature of the prior information is that all values of ρ are equally likely, then it is satisfactory to use the standard *OLS* t and F tests in the usual way. If one is unwilling to take such a stand, then Sims and Uhlig urged that investigators report both the classical hypothesis test of $\rho = 1$ and the classical confidence region around $\hat{\rho}$ and let the reader interpret the results as he or she sees fit.

APPENDIX 17.A. *Proofs of Chapter 17 Propositions*

■ **Proof of Proposition 17.2.** Observe that

$$\begin{aligned}
\sum_{s=1}^{t} u_s &= \sum_{s=1}^{t} \sum_{j=0}^{\infty} \psi_j \varepsilon_{s-j} \\
&= \{\psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots + \psi_t\varepsilon_0 + \psi_{t+1}\varepsilon_{-1} + \cdots\} \\
&\quad + \{\psi_0\varepsilon_{t-1} + \psi_1\varepsilon_{t-2} + \psi_2\varepsilon_{t-3} + \cdots + \psi_{t-1}\varepsilon_0 + \psi_t\varepsilon_{-1} + \cdots\} \\
&\quad + \{\psi_0\varepsilon_{t-2} + \psi_1\varepsilon_{t-3} + \psi_2\varepsilon_{t-4} + \cdots + \psi_{t-2}\varepsilon_0 + \psi_{t-1}\varepsilon_{-1} + \cdots\} \\
&\quad + \cdots + \{\psi_0\varepsilon_1 + \psi_1\varepsilon_0 + \psi_2\varepsilon_{-1} + \cdots\} \\
&= \psi_0\varepsilon_t + (\psi_0 + \psi_1)\varepsilon_{t-1} + (\psi_0 + \psi_1 + \psi_2)\varepsilon_{t-2} + \cdots \\
&\quad + (\psi_0 + \psi_1 + \psi_2 + \cdots + \psi_{t-1})\varepsilon_1 + (\psi_1 + \psi_2 + \cdots + \psi_t)\varepsilon_0 \\
&\quad + (\psi_2 + \psi_3 + \cdots + \psi_{t+1})\varepsilon_{-1} + \cdots \\
&= (\psi_0 + \psi_1 + \psi_2 + \cdots)\varepsilon_t - (\psi_1 + \psi_2 + \psi_3 + \cdots)\varepsilon_t \\
&\quad + (\psi_0 + \psi_1 + \psi_2 + \cdots)\varepsilon_{t-1} - (\psi_2 + \psi_3 + \cdots)\varepsilon_{t-1} \\
&\quad + (\psi_0 + \psi_1 + \psi_2 + \cdots)\varepsilon_{t-2} - (\psi_3 + \psi_4 + \cdots)\varepsilon_{t-2} + \cdots \\
&\quad + (\psi_0 + \psi_1 + \psi_2 + \cdots)\varepsilon_1 - (\psi_t + \psi_{t+1} + \cdots)\varepsilon_1 \\
&\quad + (\psi_1 + \psi_2 + \psi_3 + \cdots)\varepsilon_0 - (\psi_{t+1} + \psi_{t+2} + \cdots)\varepsilon_0 \\
&\quad + (\psi_2 + \psi_3 + \psi_4 + \cdots)\varepsilon_{-1} - (\psi_{t+2} + \psi_{t+3} + \cdots)\varepsilon_{-1} + \cdots
\end{aligned}$$

or

$$\sum_{s=1}^{t} u_s = \psi(1) \cdot \sum_{s=1}^{t} \varepsilon_s + \eta_t - \eta_0, \qquad [17.A.1]$$

where

$$\begin{aligned}
\eta_t &\equiv -(\psi_1 + \psi_2 + \psi_3 + \cdots)\varepsilon_t - (\psi_2 + \psi_3 + \psi_4 + \cdots)\varepsilon_{t-1} \\
&\quad - (\psi_3 + \psi_4 + \psi_5 + \cdots)\varepsilon_{t-2} - \cdots \\
\eta_0 &= -(\psi_1 + \psi_2 + \psi_3 + \cdots)\varepsilon_0 - (\psi_2 + \psi_3 + \psi_4 + \cdots)\varepsilon_{-1} \\
&\quad - (\psi_3 + \psi_4 + \psi_5 + \cdots)\varepsilon_{-2} - \cdots.
\end{aligned}$$

[15]See the many comments accompanying Phillips (1991a).

Notice that $\eta_t = \Sigma_{j=0}^{\infty}\alpha_j\varepsilon_{t-j}$, where $\alpha_j = -(\psi_{j+1} + \psi_{j+2} + \cdots)$, with $\{\alpha_j\}_{j=0}^{\infty}$ absolutely summable:

$$\begin{aligned}\sum_{j=0}^{\infty}|\alpha_j| &= |\psi_1 + \psi_2 + \psi_3 + \cdots| + |\psi_2 + \psi_3 + \psi_4 + \cdots| + |\psi_3 + \psi_4 + \psi_5 + \cdots| + \cdots \\ &\leq \{|\psi_1| + |\psi_2| + |\psi_3| + \cdots\} + \{|\psi_2| + |\psi_3| + |\psi_4| + \cdots\} \\ &\quad + \{|\psi_3| + |\psi_4| + |\psi_5| + \cdots\} + \cdots \\ &= |\psi_1| + 2|\psi_2| + 3|\psi_3| + \cdots \\ &= \sum_{j=0}^{\infty} j\cdot|\psi_j|,\end{aligned}$$

which is bounded by the assumptions in Proposition 17.2. ■

■ **Proof of Proposition 17.3.**

(a) This was shown in [17.5.9].
(b) This follows from [7.2.17] and the fact that $E(u_t^2) = \gamma_0$.
(c) This is implied by [7.2.14].
(d) Since $\xi_t = \Sigma_{s=1}^{t}u_s$, Proposition 17.2 asserts that

$$\xi_t = \psi(1)\sum_{s=1}^{t}\varepsilon_s + \eta_t - \eta_0. \qquad [17.A.2]$$

Hence,

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\xi_{t-1}\varepsilon_t &= T^{-1}\sum_{t=2}^{T}\left(\psi(1)\sum_{s=1}^{t-1}\varepsilon_s + \eta_{t-1} - \eta_0\right)\varepsilon_t \\ &= \psi(1)\cdot T^{-1}\sum_{t=2}^{T}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{t-1})\varepsilon_t \\ &\quad + T^{-1}\sum_{t=2}^{T}(\eta_{t-1} - \eta_0)\varepsilon_t.\end{aligned} \qquad [17.A.3]$$

But [17.3.26] established that

$$T^{-1}\sum_{t=2}^{T}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{t-1})\varepsilon_t \xrightarrow{L} (1/2)\sigma^2\cdot\{[W(1)]^2 - 1\}. \qquad [17.A.4]$$

Furthermore, Proposition 17.2 ensures that $\{(\eta_{t-1} - \eta_0)\varepsilon_t\}_{t=1}^{\infty}$ is a martingale difference sequence with finite variance, and so, from Example 7.11,

$$T^{-1}\sum_{t=2}^{T}(\eta_{t-1} - \eta_0)\varepsilon_t \xrightarrow{p} 0. \qquad [17.A.5]$$

Substituting [17.A.4] and [17.A.5] into [17.A.3] yields

$$T^{-1}\sum_{t=1}^{T}\xi_{t-1}\varepsilon_t \xrightarrow{L} (1/2)\sigma^2\cdot[\psi(1)]\cdot\{[W(1)]^2 - 1\}, \qquad [17.A.6]$$

as claimed in (d).

(e) For $j = 0$ we have from [17.1.11] that

$$T^{-1}\sum_{t=1}^{T}\xi_{t-1}u_t = (1/2)T^{-1}\xi_T^2 - (1/2)T^{-1}(u_1^2 + u_2^2 + \cdots + u_T^2). \qquad [17.A.7]$$

But

$$T^{-1}\xi_T^2 = [T^{-1/2}(u_1 + u_2 + \cdots + u_T)]^2 \xrightarrow{L} \lambda^2\cdot[W(1)]^2 \qquad [17.A.8]$$

from result (a). Also,

$$T^{-1}(u_1^2 + u_2^2 + \cdots + u_T^2) \xrightarrow{p} \gamma_0$$

from result (c). Thus, [17.A.7] converges to

$$T^{-1}\sum_{t=1}^{T}\xi_{t-1}u_t \xrightarrow{L} (1/2)\{\lambda^2\cdot[W(1)]^2 - \gamma_0\}, \tag{17.A.9}$$

which establishes result (e) for $j = 0$.

For $j > 0$, observe that

$$\xi_{t-1} = \xi_{t-j-1} + u_{t-j} + u_{t-j+1} + \cdots + u_{t-1},$$

implying that

$$\begin{aligned} T^{-1}\sum_{t=j+1}^{T}\xi_{t-1}u_{t-j} &= T^{-1}\sum_{t=j+1}^{T}(\xi_{t-j-1} + u_{t-j} + u_{t-j+1} + \cdots + u_{t-1})u_{t-j} \\ &= T^{-1}\sum_{t=j+1}^{T}\xi_{t-j-1}u_{t-j} \\ &\quad + T^{-1}\sum_{t=j+1}^{T}(u_{t-j} + u_{t-j+1} + \cdots + u_{t-1})u_{t-j}. \end{aligned} \tag{17.A.10}$$

But

$$T^{-1}\sum_{t=j+1}^{T}\xi_{t-j-1}u_{t-j} = [(T-j)/T]\cdot(T-j)^{-1}\sum_{t=1}^{T-j}\xi_{t-1}u_t \xrightarrow{L} (1/2)\{\lambda^2\cdot[W(1)]^2 - \gamma_0\},$$

as in [17.A.9]. Also,

$$T^{-1}\sum_{t=j+1}^{T}(u_{t-j} + u_{t-j+1} + \cdots + u_{t-1})u_{t-j} \xrightarrow{p} \gamma_0 + \gamma_1 + \gamma_2 + \cdots + \gamma_{j-1},$$

from result (c). This, [17.A.10] converges to

$$T^{-1}\sum_{t=j+1}^{T}\xi_{t-1}u_{t-j} \xrightarrow{L} (1/2)\{\lambda^2\cdot[W(1)]^2 - \gamma_0\} + \{\gamma_0 + \gamma_1 + \gamma_2 + \cdots + \gamma_{j-1}\}.$$

Clearly, $T^{-1}\Sigma_{t=1}^{T}\xi_{t-1}u_{t-j}$ has the same asymptotic distribution, since

$$T^{-1}\sum_{t=1}^{j}\xi_{t-1}u_{t-j} \xrightarrow{p} 0.$$

(f) From the definition of ξ_t in [17.5.11] and $X_T(r)$ in [17.5.4], it follows as in [17.3.15] that

$$\int_0^1 \sqrt{T}\cdot X_T(r)\,dr = T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}.$$

Result (f) then follows immediately from [17.5.5].

(g) First notice that

$$T^{-3/2}\sum_{t=1}^{T}tu_{t-j} = T^{-3/2}\sum_{t=1}^{T}(t-j+j)u_{t-j},$$

where $j\cdot T^{-3/2}\Sigma_{t=1}^{T}u_{t-j} \xrightarrow{p} 0$. Hence,

$$T^{-3/2}\sum_{t=1}^{T}tu_{t-j} \xrightarrow{p} T^{-3/2}\sum_{t=1}^{T}(t-j)u_{t-j} \xrightarrow{p} T^{-3/2}\sum_{t=1}^{T}tu_t.$$

But from [17.3.19],

$$T^{-3/2}\sum_{t=1}^{T}tu_t = T^{-1/2}\sum_{t=1}^{T}u_t - T^{-3/2}\sum_{t=1}^{T}\xi_{t-1} \xrightarrow{L} \lambda\cdot W(1) - \lambda\cdot\int_0^1 W(r)\,dr,$$

by virtue of (a) and (f).

(h) Using the same analysis as in [17.3.20] through [17.3.22], for ξ_t defined in [17.5.11] and $X_T(r)$ defined in [17.5.4], we have

$$T^{-1}\{\xi_1^2/T + \xi_2^2/T + \cdots + \xi_{T-1}^2/T\} = \int_0^1 [\sqrt{T}\cdot X_T(r)]^2\,dr \xrightarrow{L} [\sigma\cdot\psi(1)]^2\cdot\int_0^1 [W(r)]^2\,dr,$$

by virtue of [17.5.5].

(i) As in [17.3.23],

$$\begin{aligned}
T^{-5/2}\sum_{t=1}^{T} t\xi_{t-1} &= T^{1/2}\sum_{t=1}^{T}(t/T)\cdot(\xi_{t-1}/T^2)\\
&= T^{1/2}\int_0^1 \{([Tr]^* + 1)/T\}\cdot\{(u_1 + u_2 + \cdots + u_{[Tr]^*})/T\}\,dr\\
&= T^{1/2}\int_0^1 \{([Tr]^* + 1)/T\}\cdot X_T(r)\,dr\\
&\xrightarrow{L} \sigma\cdot\psi(1)\cdot\int_0^1 rW(r)\,dr,
\end{aligned}$$

from [17.5.5] and the continuous mapping theorem.

(j) From the same argument as in (i),

$$\begin{aligned}
T^{-3}\sum_{t=1}^{T} t\xi_{t-1}^2 &= \sum_{t=1}^{T}(t/T)(\xi_{t-1}^2/T^2)\\
&= T\int_0^1 \{([Tr]^* + 1)/T\}\cdot\{(u_1 + u_2 + \cdots + u_{[Tr]^*})/T\}^2\,dr\\
&= T\int_0^1 \{([Tr]^* + 1)/T\}\cdot[X_T(r)]^2\,dr\\
&\xrightarrow{L} [\sigma\cdot\psi(1)]^2\cdot\int_0^1 r[W(r)]^2\,dr.
\end{aligned}$$

(k) This is identical to result (h) from Proposition 17.1, repeated in this proposition for the reader's convenience. ■

Chapter 17 Exercises

17.1. Let $\{u_t\}$ be an i.i.d. sequence with mean zero and variance σ^2, and let $y_t = u_1 + u_2 + \cdots + u_t$ with $y_0 = 0$. Deduce from [17.3.17] and [17.3.18] that

$$\begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-3/2}\Sigma y_{t-1}\end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \sigma^2\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{3}\end{bmatrix}\right),$$

where Σ indicates summation over t from 1 to T. Comparing this result with Proposition 17.1, argue that

$$\begin{bmatrix} W(1) \\ \int W(r)\,dr\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{3}\end{bmatrix}\right),$$

where the integral sign denotes integration over r from 0 to 1.

17.2. *Phillips (1987) generalization of case 1.* Suppose that data are generated from the process $y_t = y_{t-1} + u_t$, where $u_t = \psi(L)\varepsilon_t$, $\Sigma_{j=0}^{\infty} j\cdot|\psi_j| < \infty$, and ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment. Consider OLS estimation of the autoregression $y_t = \rho y_{t-1} + u_t$. Let $\hat\rho_T = (\Sigma y_{t-1}^2)^{-1}(\Sigma y_{t-1}y_t)$ be the OLS estimate of ρ, $s_T^2 = (T-1)^{-1}\times\Sigma\hat u_t^2$ the OLS estimate of the variance of the regression error, $\hat\sigma_{\hat\rho_T}^2 = s_T^2\cdot(\Sigma y_{t-1}^2)^{-1}$ the OLS estimate of the variance of $\hat\rho_T$, and $t_T = (\hat\rho_T - 1)/\hat\sigma_{\hat\rho_T}$ the OLS t test of $\rho = 1$, and define

$\lambda \equiv \sigma\cdot\psi(1)$. Use Proposition 17.3 to show that

(a) $T(\hat{\rho}_T - 1) \xrightarrow{L} \dfrac{\frac{1}{2}\{\lambda^2\cdot[W(1)]^2 - \gamma_0\}}{\lambda^2\cdot\int [W(r)]^2\,dr}$;

(b) $T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} \xrightarrow{L} \dfrac{\gamma_0}{\lambda^2\cdot\int [W(r)]^2\,dr}$;

(c) $t_T \xrightarrow{L} (\lambda^2/\gamma_0)^{1/2}\left\{\dfrac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int [W(r)]^2\,dr\right\}^{1/2}} + \dfrac{\frac{1}{2}(\lambda^2 - \gamma_0)}{\lambda^2\left\{\int [W(r)]^2\,dr\right\}^{1/2}}\right\}$;

(d) $T(\hat{\rho}_T - 1) - \frac{1}{2}(T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2)(\lambda^2 - \gamma_0) \xrightarrow{L} \dfrac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\int [W(r)]^2\,dr}$;

(e) $(\gamma_0/\lambda^2)^{1/2}\cdot t_T - \{\frac{1}{2}(\lambda^2 - \gamma_0)/\lambda\} \times \{T\cdot\hat{\sigma}_{\hat{\rho}_T} \div s_T\} \xrightarrow{L} \dfrac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int [W(r)]^2\,dr\right\}^{1/2}}$.

Suggest estimates of γ_0 and λ^2 that could be used to construct the statistics in (d) and (e), and indicate where one could find critical values for these statistics.

17.3. *Phillips and Perron (1988) generalization of case 4.* Suppose that data are generated from the process $y_t = \alpha + y_{t-1} + u_t$, where $u_t = \psi(L)\varepsilon_t$, $\sum_{j=0}^{\infty} j\cdot|\psi_j| < \infty$, and ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment, and where α can be any value, including zero. Consider *OLS* estimation of

$$y_t = \alpha + \rho y_{t-1} + \delta t + u_t.$$

As in [17.4.49], note that the fitted values and estimate of ρ from this regression are identical to those from an *OLS* regression of y_t on a constant, time trend, and $\xi_{t-1} \equiv y_{t-1} - \alpha(t-1)$:

$$y_t = \alpha^* + \rho^*\xi_{t-1} + \delta^* t + u_t,$$

where, under the assumed data-generating process, ξ_t satisfies the assumptions of Proposition 17.3. Let $(\hat{\alpha}_T^*, \hat{\rho}_T^*, \hat{\delta}_T^*)'$ be the *OLS* estimates given by equation [17.4.50], $s_T^2 = (T-3)^{-1} \times \Sigma\hat{u}_t^2$ the *OLS* estimate of the variance of the regression error, $\hat{\sigma}^2_{\hat{\rho}_T^*}$ the *OLS* estimate of the variance of $\hat{\rho}_T^*$ given in [17.4.54], and $t_T^* = (\hat{\rho}_T^* - 1)/\hat{\sigma}_{\hat{\rho}_T^*}$ the *OLS* t test of $\rho = 1$. Recall further that $\hat{\rho}_T^*$, $\hat{\sigma}^2_{\hat{\rho}_T^*}$, and t_T^* are numerically identical to the analogous magnitudes for the original regression, $\hat{\rho}_T$, $\hat{\sigma}^2_{\hat{\rho}_T}$, and t_T. Finally, define $\lambda \equiv \sigma\cdot\psi(1)$. Use Proposition 17.3 to show that

(a)
$$\begin{bmatrix} 1 & T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma t \\ T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma\xi_{t-1}^2 & T^{-5/2}\Sigma\xi_{t-1}t \\ T^{-2}\Sigma t & T^{-5/2}\Sigma t\xi_{t-1} & T^{-3}\Sigma t^2 \end{bmatrix}$$
$$\xrightarrow{L} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

(b)
$$\begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma\xi_{t-1}u_t \\ T^{-3/2}\Sigma tu_t \end{bmatrix} \xrightarrow{L} \lambda\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - [\gamma_0/\lambda^2]\} \\ W(1) - \int W(r)\,dr \end{bmatrix};$$

(c) $\begin{bmatrix} T^{1/2}\hat{\alpha}_T^* \\ T(\hat{\rho}_T^* - 1) \\ T^{3/2}(\hat{\delta}_T^* - \alpha_0) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}$

$$\times \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - [\gamma_0/\lambda^2]\} \\ \left\{W(1) - \int W(r)\,dr\right\} \end{bmatrix};$$

(d) $T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \xrightarrow{p} (s_T^2/\lambda^2)[0 \quad 1 \quad 0] \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$$\equiv (s_T^2/\lambda^2) \cdot Q;$$

(e) $t_T \xrightarrow{p} (\lambda^2/\gamma_0)^{1/2} \cdot T(\hat{\rho}_T - 1)/\sqrt{Q}$;

(f) $T(\hat{\rho}_T - 1) - \frac{1}{2}(T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2)(\lambda^2 - \gamma_0)$

$$\xrightarrow{L} [0 \quad 1 \quad 0] \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \\ W(1) - \int W(r)\,dr \end{bmatrix}$$

$$\equiv V;$$

(g) $(\gamma_0/\lambda^2)^{1/2} \cdot t_T - \{\frac{1}{2}(\lambda^2 - \gamma_0)/\lambda\} \times \{T \cdot \hat{\sigma}_{\hat{\rho}_T} \div s_T\} \xrightarrow{L} V \div \sqrt{Q}$.

Suggest estimates of γ_0 and λ^2 that could be used to construct the statistics in (f) and (g), and indicate where one could find critical values for these statistics.

17.4. *Generalization of case 1 for autoregression.* Consider *OLS* estimation of

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \rho y_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment and the roots of $(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1} z^{p-1}) = 0$ are outside the unit circle. Define $\lambda \equiv \sigma/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})$ and $\gamma_j = E\{(\Delta y_t)(\Delta y_{t-j})\}$. Let $\hat{\boldsymbol{\zeta}}_T \equiv (\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{p-1,T})'$ be the $(p - 1) \times 1$ vector of estimated *OLS* coefficients on the lagged changes in y, and let $\boldsymbol{\zeta}$ be the corresponding true value. Show that if the true value of ρ is unity, then

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \\ T(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0}' & \lambda^2 \cdot \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_1 \\ \frac{1}{2}\sigma\lambda\{[W(1)]^2 - 1\} \end{bmatrix},$$

where $\mathbf{V}$ is the $[(p - 1) \times (p - 1)]$ matrix defined in [17.7.19] and $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V})$. Deduce from this that

(a) $T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{V}^{-1})$;

(b) $T(\hat{\rho}_T - 1)/(1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \cdots - \hat{\zeta}_{p-1,T}) \xrightarrow{L} \dfrac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\int [W(r)]^2\,dr}$;

(c) $(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T} \xrightarrow{L} \dfrac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int [W(r)]^2\, dr\right\}^{1/2}}.$

Where could you find critical values for the statistics in (b) and (c)?

17.5. *Generalization of case 3 for autoregression.* Consider *OLS* estimation of

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1}\Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment and the roots of $(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1}z^{p-1}) = 0$ are outside the unit circle.

(a) Show that the fitted values for this regression are identical to those for the following transformed specification:

$$y_t = \zeta_1 u_{t-1} + \zeta_2 u_{t-2} + \cdots + \zeta_{p-1}u_{t-p+1} + \mu + \rho y_{t-1} + \varepsilon_t,$$

where $u_t \equiv \Delta y_t - \mu$ and $\mu \equiv \alpha/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})$.

(b) Suppose that the true value of ρ is 1 and the true value of α is nonzero. Show that under these assumptions,

$$\begin{aligned} u_t &= [1/(1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_{p-1}L^{p-1})]\varepsilon_t \\ y_{t-1} &= \mu(t-1) + \xi_{t-1}, \end{aligned}$$

where

$$\xi_{t-1} \equiv y_0 + u_1 + u_2 + \cdots + u_{t-1}.$$

Conclude that for fixed y_0, the variables u_t and ξ_t satisfy the assumptions of Proposition 17.3 and that y_t is dominated asymptotically by a time trend.

(c) Let $\gamma_j \equiv E(u_t u_{t-j})$, and let $\hat{\boldsymbol{\zeta}}_T \equiv (\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{p-1,T})'$ be the $(p-1) \times 1$ vector of estimated *OLS* coefficients on $(u_{t-1}, u_{t-2}, \ldots, u_{t-p+1})$; these, of course, are identical to the coefficients on $(\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1})$ in the original regression. Show that if $p = 1$ and $\alpha \neq 0$,

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \\ T^{1/2}(\hat{\mu}_T - \mu) \\ T^{3/2}(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \mu/2 \\ \mathbf{0}' & \mu/2 & \mu^2/3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_1 \\ h_2 \\ h_3 \end{bmatrix},$$

where

$$\begin{bmatrix} \mathbf{h}_1 \\ h_2 \\ h_3 \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{V} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \mu/2 \\ \mathbf{0}' & \mu/2 & \mu^2/3 \end{bmatrix} \right)$$

and $\mathbf{V}$ is the matrix in [17.7.19]. Conclude as in the analysis of Section 16.3 that any *OLS* t or F test on the original regression can be compared with the standard t and F tables to give an asymptotically valid inference.

17.6. *Generalization of case 4 for autoregression.* Consider *OLS* estimation of

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1}\Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \delta t + \varepsilon_t,$$

where ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment and then roots of $(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1}z^{p-1}) = 0$ are outside the unit circle.

(a) Show that the fitted values of this regression are numerically identical to those of the following specification:

$$y_t = \zeta_1 u_{t-1} + \zeta_2 u_{t-2} + \cdots + \zeta_{p-1}u_{t-p+1} + \mu^* + \rho \xi_{t-1} + \delta^* t + \varepsilon_t,$$

where $u_t \equiv \Delta y_t - \mu$, $\mu \equiv \alpha/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})$, $\mu^* \equiv (1-\rho)\mu$, $\xi_{t-1} \equiv y_{t-1} - \mu(t-1)$, and $\delta^* \equiv \delta + \rho\mu$. Note that the estimated coefficients $\hat{\boldsymbol{\zeta}}_T$ and $\hat{\rho}_T$ and their standard errors will be identical for the two regressions.

(b) Suppose that the true value of ρ is 1 and the true value of δ is 0. Show that under these assumptions,

$$\begin{aligned} u_t &= [1/(1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_{p-1}L^{p-1})]\varepsilon_t \\ \xi_{t-1} &= y_0 + u_1 + u_2 + \cdots + u_{t-1}. \end{aligned}$$

Conclude that for fixed y_0, the variables u_t and ξ_t satisfy the assumptions of Proposition 17.3.

(c) Again let $\rho = 1$ and $\delta = 0$, and define $\gamma_j \equiv E(u_t u_{t-j})$ and

$$\lambda \equiv \sigma/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1}).$$

Show that

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \\ T^{1/2}\hat{\mu}_T^* \\ T(\hat{\rho}_T - 1) \\ T^{3/2}(\hat{\delta}_T^* - \delta^*) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \lambda\cdot\int W(r)\,dr & 1/2 \\ \mathbf{0}' & \lambda\cdot\int W(r)\,dr & \lambda^2\cdot\int [W(r)]^2\,dr & \lambda\cdot\int rW(r)\,dr \\ \mathbf{0}' & 1/2 & \lambda\cdot\int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \mathbf{h}_1 \\ \sigma\cdot W(1) \\ \tfrac{1}{2}\sigma\lambda\{[W(1)]^2 - 1\} \\ \sigma\cdot\left\{W(1) - \int W(r)\,dr\right\} \end{bmatrix}$$

where $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V})$ and $\mathbf{V}$ is as defined in [17.7.19].

(d) Deduce from answer (c) that

$T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{V}^{-1})$;

$T(\hat{\rho}_T - 1)/(1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \cdots - \hat{\zeta}_{p-1,T})$

$$\xrightarrow{L} [0 \;\; 1 \;\; 0]\begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}\begin{bmatrix} W(1) \\ \tfrac{1}{2}\{[W(1)]^2 - 1\} \\ W(1) - \int W(r)\,dr \end{bmatrix}$$

$\equiv V$;

$(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T} \xrightarrow{L} V \div \sqrt{Q}$,

where

$$Q \equiv [0 \;\; 1 \;\; 0]\begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Notice that the distribution of V is the same as the asymptotic distribution of the variable tabulated for case 4 in Table B.5, while the distribution of $V/\sqrt{Q}$ is the same as the asymptotic distribution of the variable tabulated for case 4 in Table B.6.

Chapter 17 References

Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59:817–58.

———. 1993. "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models." *Econometrica* 61:139–65.

Beveridge, Stephen, and Charles R. Nelson. 1981. "A New Approcah to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle.'" *Journal of Monetary Economics* 7:151–74.

Bhargava, Alok. 1986. "On the Theory of Testing for Unit Roots in Observed Time Series." *Review of Economic Studies* 53:369–84.

Billingsley, Patrick. 1968. *Convergence of Probability Measures.* New York: Wiley.

Campbell, John Y., and Pierre Perron. 1991. "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots." *NBER Macroeconomics Annual.* Cambridge, Mass.: MIT Press.

Cecchetti, Stephen G., and Pok-sang Lam. 1991. "What Do We Learn from Variance Ratio Statistics? A Study of Stationary and Nonstationary Models with Breaking Trends." Department of Economics, Ohio State University. Mimeo.

Chan, N. H., and C. Z. Wei. 1987. "Asymptotic Inference for Nearly Nonstationary AR(1) Processes." *Annals of Statistics* 15:1050–63.

——— and ———. 1988. "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes." *Annals of Statistics* 16:367–401.

Cochrane, John H. 1988. "How Big Is the Random Walk in GNP?" *Journal of Political Economy* 96:893–920.

DeJong, David N., and Charles H. Whiteman. 1991. "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series.'" *Journal of Monetary Economics* 28:221–54.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74:427–31.

——— and ———. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49:1057–72.

——— and S. G. Pantula. 1987. "Determining the Order of Differencing in Autoregressive Processes." *Journal of Business and Economic Statistics* 5:455–61.

Evans, G. B. A., and N. E. Savin. 1981. "Testing for Unit Roots: 1." *Econometrica* 49:753–79.

——— and ———. 1984. "Testing for Unit Roots: 2." *Econometrica* 52:1241–69.

Fuller, Wayne A. 1976. *Introduction to Statistical Time Series.* New York: Wiley.

Hall, Alastair. 1989. "Testing for a Unit Root in the Presence of Moving Average Errors." *Biometrika* 76:49–56.

———. 1991. "Testing for a Unit Root in Time Series with Pretest Data Based Model Selection." Department of Economics, North Carolina State University. Mimeo.

Hall, P., and C. C. Heyde. 1980. *Martingale Limit Theory and Its Application.* New York: Academic Press.

Hansen, Bruce E. 1992. "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes." *Econometrica* 60:967–72.

Jeffreys, H. 1946. "An Invariant Form for the Prior Probability in Estimation Problems." *Proceedings of the Royal Society of London* Series A, 186:453–61.

Kim, Kiwhan, and Peter Schmidt. 1990. "Some Evidence on the Accuracy of Phillips-Perron Tests Using Alternative Estimates of Nuisance Parameters." *Economics Letters* 34:345–50.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54:159–78.

Lo, Andrew W., and A. Craig MacKinlay. 1988. "Stock Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." *Review of Financial Studies* 1:41–66.

——— and ———. 1989. "The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation." *Journal of Econometrics* 40:203–38.

Malliaris, A. G., and W. A. Brock. 1982. *Stochastic Methods in Economics and Finance.* Amsterdam: North-Holland.

Pantula, Sastry G., and Alastair Hall. 1991. "Testing for Unit Roots in Autoregressive Moving Average Models: An Instrumental Variable Approach." *Journal of Econometrics* 48:325–53.

Park, Joon Y., and B. Choi. 1988. "A New Approach to Testing for a Unit Root." Cornell University. Mimeo.

Park, Joon Y., and Peter C. B. Phillips. 1988. "Statistical Inference in Regressions with Integrated Processes: Part 1." *Econometric Theory* 4:468–97.