16 Processes with Deterministic Time Trends

The coefficients of regression models involving unit roots or deterministic time trends are typically estimated by ordinary least squares. However, the asymptotic distributions of the coefficient estimates cannot be calculated in the same way as are those for regression models involving stationary variables. Among other difficulties, the estimates of different parameters will in general have different asymptotic rates of convergence. This chapter introduces the idea of different rates of convergence and develops a general approach to obtaining asymptotic distributions suggested by Sims, Stock, and Watson (1990).[1] This chapter deals exclusively with processes involving deterministic time trends but no unit roots. One of the results for such processes will be that the usual *OLS* t and F statistics, calculated in the usual way, have the same asymptotic distributions as they do for stationary regressions. Although the limiting distributions are standard, the techniques used to verify these limiting distributions are different from those used in Chapter 8. These techniques will also be used to develop the asymptotic distributions for processes including unit roots in Chapters 17 and 18.

This chapter begins with the simplest example of i.i.d. innovations around a deterministic time trend. Section 16.1 derives the asymptotic distributions of the coefficient estimates for this model and illustrates a rescaling of variables that is necessary to accommodate different asymptotic rates of convergence. Section 16.2 shows that despite the different asymptotic rates of convergence, the standard *OLS* t and F statistics have the usual limiting distributions for this model. Section 16.3 develops analogous results for a covariance-stationary autoregression around a deterministic time trend. That section also introduces the Sims, Stock, and Watson technique of transforming the regression model into a canonical form for which the asymptotic distribution is simpler to describe.

16.1. *Asymptotic Distribution of* OLS *Estimates of the Simple Time Trend Model*

This section considers *OLS* estimation of the parameters of a simple time trend,

$$y_t = \alpha + \delta t + \varepsilon_t, \qquad [16.1.1]$$

for ε_t a white noise process. If $\varepsilon_t \sim N(0, \sigma^2)$, then the model [16.1.1] satisfies the classical regression assumptions[2] and the standard *OLS* t or F statistics in equations

[1] A simpler version of this theme appeared in the analysis of a univariate process with unit roots by Fuller (1976).

[2] See Assumption 8.1 in Chapter 8.

[8.1.26] and [8.1.32] would have exact small-sample t or F distributions. On the other hand, if ε_t is non-Gaussian, then a slightly different technique for finding the asymptotic distributions of the *OLS* estimates of α and δ would have to be used from that employed for stationary regressions in Chapter 8. This chapter introduces this technique, which will prove useful not only for studying time trends but also for analyzing estimators for a variety of nonstationary processes in Chapters 17 and 18.[3]

Recall the approach used to find asymptotic distributions for regressions with stationary explanatory variables in Chapter 8. Write [16.1.1] in the form of the standard regression model,

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \tag{16.1.2}$$

where

$$\underset{(1\times 2)}{\mathbf{x}_t'} \equiv [1 \quad t] \tag{16.1.3}$$

$$\underset{(2\times 1)}{\boldsymbol{\beta}} \equiv \begin{bmatrix} \alpha \\ \delta \end{bmatrix}. \tag{16.1.4}$$

Let $\mathbf{b}_T$ denote the *OLS* estimate of $\boldsymbol{\beta}$ based on a sample of size T:

$$\mathbf{b}_T \equiv \begin{bmatrix} \hat{\alpha}_T \\ \hat{\delta}_T \end{bmatrix} = \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t y_t\right]. \tag{16.1.5}$$

Recall from equation [8.2.3] that the deviation of the *OLS* estimate from the true value can be expressed as

$$(\mathbf{b}_T - \boldsymbol{\beta}) = \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t\right]. \tag{16.1.6}$$

To find the limiting distribution for a regression with stationary explanatory variables, the approach in Chapter 8 was to multiply [16.1.6] by $\sqrt{T}$, resulting in

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) = \left[(1/T)\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[(1/\sqrt{T})\sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t\right]. \tag{16.1.7}$$

The usual assumption was that $(1/T)\Sigma_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ converged in probability to a nonsingular matrix $\mathbf{Q}$ while $(1/\sqrt{T})\Sigma_{t=1}^{T}\mathbf{x}_t\varepsilon_t$ converged in distribution to a $N(\mathbf{0}, \sigma^2\mathbf{Q})$ random variable, implying that $\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$.

To see why this same argument cannot be used for a deterministic time trend, note that for $\mathbf{x}_t$ and $\boldsymbol{\beta}$ given in equations [16.1.3] and [16.1.4], expression [16.1.6] would be

$$\begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\delta}_T - \delta \end{bmatrix} = \begin{bmatrix} \Sigma 1 & \Sigma t \\ \Sigma t & \Sigma t^2 \end{bmatrix}^{-1}\begin{bmatrix} \Sigma\varepsilon_t \\ \Sigma t\varepsilon_t \end{bmatrix}, \tag{16.1.8}$$

[3]The general approach in these chapters follows Sims, Stock, and Watson (1990).

where Σ denotes summation for $t = 1$ through T. It is straightforward to show by induction that[4]

$$\sum_{t=1}^{T} t = T(T + 1)/2 \qquad [16.1.9]$$

$$\sum_{t=1}^{T} t^2 = T(T + 1)(2T + 1)/6. \qquad [16.1.10]$$

Thus, the leading term in $\Sigma_{t=1}^{T} t$ is $T^2/2$; that is,

$$(1/T^2) \sum_{t=1}^{T} t = (1/T^2)[(T^2/2) + (T/2)] = 1/2 + 1/(2T) \to 1/2. \qquad [16.1.11]$$

Similarly, the leading term in $\Sigma_{t=1}^{T} t^2$ is $T^3/3$:

$$\begin{aligned}(1/T^3) \sum_{t=1}^{T} t^2 &= (1/T^3)[(2T^3/6) + (3T^2/6) + T/6] \\ &= 1/3 + 1/(2T) + 1/(6T^2) \\ &\to 1/3.\end{aligned} \qquad [16.1.12]$$

For future reference, we note here the general pattern—the leading term in $\Sigma_{t=1}^{T} t^v$ is $T^{v+1}/(v + 1)$:

$$(1/T^{v+1}) \sum_{t=1}^{T} t^v \to 1/(v + 1). \qquad [16.1.13]$$

To verify [16.1.13], note that

$$(1/T^{v+1}) \sum_{t=1}^{T} t^v = (1/T) \sum_{t=1}^{T} (t/T)^v. \qquad [16.1.14]$$

The right side of [16.1.14] can be viewed as an approximation to the area under the curve

$$f(r) = r^v$$

for r between zero and unity. To see this, notice that $(1/T) \cdot (t/T)^v$ represents the area of a rectangle with width $(1/T)$ and height r^v evaluated at $r = t/T$ (see Figure 16.1). Thus, [16.1.14] is the sum of the area of these rectangles evaluated

[4] Clearly, [16.1.9] and [16.1.10] hold for $T = 1$. Given that [16.1.9] holds for T,

$$\sum_{t=1}^{T+1} t = \sum_{t=1}^{T} t + (T + 1) = T(T + 1)/2 + (T + 1) = (T + 1)[(T/2) + 1] = (T + 1)(T + 2)/2,$$

establishing that [16.1.9] holds for $T + 1$. Similarly, given that [16.1.10] holds for T,

$$\begin{aligned}\sum_{t=1}^{T+1} t^2 &= T(T + 1)(2T + 1)/6 + (T + 1)^2 \\ &= (T + 1)\{[T(2T + 1)/6] + (T + 1)\} \\ &= (T + 1)(2T^2 + 7T + 6)/6 \\ &= (T + 1)(T + 2)[2(T + 1) + 1]/6,\end{aligned}$$

establishing that [16.1.10] holds for $T + 1$.

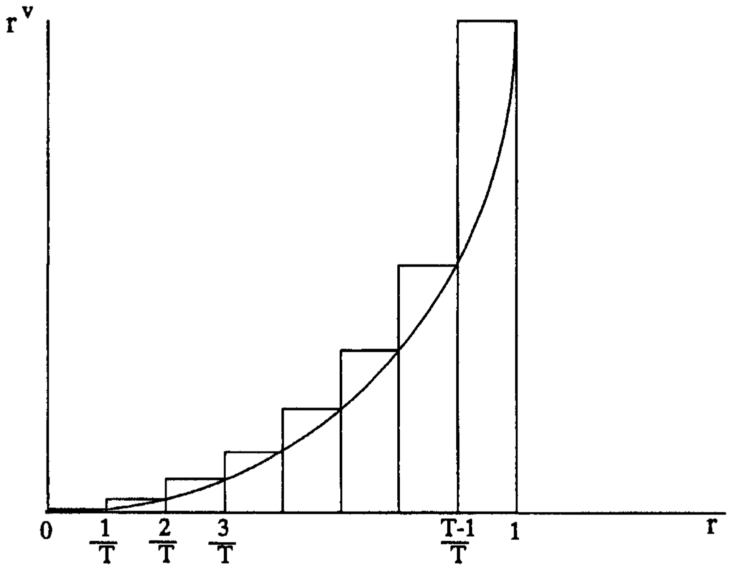

FIGURE 16.1 Demonstration that $(1/T)\sum_{t=1}^{T}(t/T)^{\nu} \to \int_0^1 r^{\nu}\,dr = 1/(\nu + 1)$.

at $r = 1/T, 2/T, \ldots, 1$. As $T \to \infty$, this sum converges to the area under the curve $f(r)$:

$$(1/T)\sum_{t=1}^{T}(t/T)^{\nu} \to \int_0^1 r^{\nu}\,dr = r^{\nu+1}/(\nu + 1)\big|_{r=0}^{1} = 1/(\nu + 1). \qquad [16.1.15]$$

For $\mathbf{x}_t$ given in [16.1.3], results [16.1.9] and [16.1.10] imply that

$$\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' = \begin{bmatrix} \Sigma 1 & \Sigma t \\ \Sigma t & \Sigma t^2 \end{bmatrix} = \begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix}. \qquad [16.1.16]$$

In contrast to the usual result for stationary regressions, for the matrix in [16.1.16], $(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ diverges. To obtain a convergent matrix, [16.1.16] would have to be divided by T^3 rather than T:

$$T^{-3}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' \to \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}.$$

Unfortunately, this limiting matrix cannot be inverted, as $(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ can be in the usual case. Hence, a different approach from that in the stationary case will be needed to calculate the asymptotic distribution of $\mathbf{b}_T$.

It turns out that the *OLS* estimates $\hat{\alpha}_T$ and $\hat{\delta}_T$ have different asymptotic rates of convergence. To arrive at nondegenerate limiting distributions, $\hat{\alpha}_T$ is multiplied by $\sqrt{T}$, whereas $\hat{\delta}_T$ must be multiplied by $T^{3/2}$! We can think of this adjustment as premultiplying [16.1.6] or [16.1.8] by the matrix

$$\mathbf{Y}_T \equiv \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T^{3/2} \end{bmatrix}, \qquad [16.1.17]$$

resulting in

$$\begin{bmatrix}\sqrt{T}(\hat{\alpha}_T - \alpha)\\ T^{3/2}(\hat{\delta}_T - \delta)\end{bmatrix} = \mathbf{Y}_T\left[\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t\right]$$

$$= \mathbf{Y}_T\left[\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}\mathbf{Y}_T\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t\right] \qquad [16.1.18]$$

$$= \left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{Y}_T^{-1}\right\}^{-1}\left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t\right]\right\}.$$

Consider the first term in the last expression of [16.1.18]. Substituting from [16.1.17] and [16.1.16],

$$\left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{Y}_T^{-1}\right\} = \left\{\begin{bmatrix}T^{-1/2} & 0\\ 0 & T^{-3/2}\end{bmatrix}\begin{bmatrix}\Sigma 1 & \Sigma t\\ \Sigma t & \Sigma t^2\end{bmatrix}\begin{bmatrix}T^{-1/2} & 0\\ 0 & T^{-3/2}\end{bmatrix}\right\}$$

$$= \begin{bmatrix}T^{-1}\Sigma 1 & T^{-2}\Sigma t\\ T^{-2}\Sigma t & T^{-3}\Sigma t^2\end{bmatrix}.$$

Thus, it follows from [16.1.11] and [16.1.12] that

$$\left\{\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{Y}_T^{-1}\right\} \to \mathbf{Q}, \qquad [16.1.19]$$

where

$$\mathbf{Q} \equiv \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{3}\end{bmatrix}. \qquad [16.1.20]$$

Turning next to the second term in [16.1.18],

$$\mathbf{Y}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t\right] = \begin{bmatrix}T^{-1/2} & 0\\ 0 & T^{-3/2}\end{bmatrix}\begin{bmatrix}\Sigma\varepsilon_t\\ \Sigma t\varepsilon_t\end{bmatrix} = \begin{bmatrix}(1/\sqrt{T})\Sigma\varepsilon_t\\ (1/\sqrt{T})\Sigma(t/T)\varepsilon_t\end{bmatrix}. \qquad [16.1.21]$$

Under standard assumptions about ε_t, this vector will be asymptotically Gaussian. For example, suppose that ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment. Then the first element of the vector in [16.1.21] satisfies

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

by the central limit theorem.

For the second element of the vector in [16.1.21], observe that $\{(t/T)\varepsilon_t\}$ is a martingale difference sequence that satisfies the conditions of Proposition 7.8. Specifically, its variance is

$$\sigma_t^2 = E[(t/T)\varepsilon_t]^2 = \sigma^2\cdot(t^2/T^2),$$

where

$$(1/T)\sum_{t=1}^{T}\sigma_t^2 = \sigma^2(1/T^3)\sum_{t=1}^{T}t^2 \to \sigma^2/3.$$

Furthermore, $(1/T)\sum_{t=1}^{T}[(t/T)\varepsilon_t]^2 \xrightarrow{p} \sigma^2/3$. To verify this last claim, notice that

$$\begin{aligned}
&E\left((1/T)\sum_{t=1}^{T}[(t/T)\varepsilon_t]^2 - (1/T)\sum_{t=1}^{T}\sigma_t^2\right)^2 \\
&= E\left((1/T)\sum_{t=1}^{T}[(t/T)\varepsilon_t]^2 - (1/T)\sum_{t=1}^{T}(t/T)^2\sigma^2\right)^2 \\
&= E\left((1/T)\sum_{t=1}^{T}(t/T)^2(\varepsilon_t^2 - \sigma^2)\right)^2 \\
&= (1/T)^2\sum_{t=1}^{T}(t/T)^4E(\varepsilon_t^2 - \sigma^2)^2.
\end{aligned} \qquad [16.1.22]$$

But from [16.1.13], T times the magnitude in [16.1.22] converges to

$$(1/T)\sum_{t=1}^{T}(t/T)^4E(\varepsilon_t^2 - \sigma^2)^2 \to (1/5)\cdot E(\varepsilon_t^2 - \sigma^2)^2,$$

meaning that [16.1.22] itself converges to zero:

$$(1/T)\sum_{t=1}^{T}[(t/T)\varepsilon_t]^2 - (1/T)\sum_{t=1}^{T}\sigma_t^2 \xrightarrow{m.s.} 0.$$

But this implies that

$$(1/T)\sum_{t=1}^{T}[(t/T)\varepsilon_t]^2 \xrightarrow{p} \sigma^2/3,$$

as claimed. Hence, from Proposition 7.8, $(1/\sqrt{T})\sum_{t=1}^{T}(t/T)\varepsilon_t$ satisfies the central limit theorem:

$$(1/\sqrt{T})\sum_{t=1}^{T}(t/T)\varepsilon_t \xrightarrow{L} N(0, \sigma^2/3).$$

Finally, consider the joint distribution of the two elements in the (2×1) vector described by [16.1.21]. Any linear combination of these elements takes the form

$$(1/\sqrt{T})\sum_{t=1}^{T}[\lambda_1 + \lambda_2(t/T)]\varepsilon_t.$$

Then $[\lambda_1 + \lambda_2(t/T)]\varepsilon_t$ is also a martingale difference sequence with positive variance[5] given by $\sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2]$ satisfying

$$\begin{aligned}
(1/T)\sum_{t=1}^{T}\sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2] &\to \sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(\tfrac{1}{2}) + \lambda_2^2(\tfrac{1}{3})] \\
&= \sigma^2\boldsymbol{\lambda}'\mathbf{Q}\boldsymbol{\lambda}
\end{aligned}$$

for $\boldsymbol{\lambda} \equiv (\lambda_1, \lambda_2)'$ and $\mathbf{Q}$ the matrix in [16.1.20]. Furthermore,

$$(1/T)\sum_{t=1}^{T}[\lambda_1 + \lambda_2(t/T)]^2\varepsilon_t^2 \xrightarrow{p} \sigma^2\boldsymbol{\lambda}'\mathbf{Q}\boldsymbol{\lambda}; \qquad [16.1.23]$$

see Exercise 16.1. Thus any linear combination of the two elements in the vector in [16.1.21] is asymptotically Gaussian, implying a limiting bivariate Gaussian dis-

[5]More accurately, a given nonzero λ_1 and λ_2 will produce a zero variance for $[\lambda_1 + \lambda_2(t/T)]\varepsilon_t$ for at most a single value of t, which does not affect the validity of the asymptotic claim.

tribution:

$$\begin{bmatrix} (1/\sqrt{T})\Sigma\varepsilon_t \\ (1/\sqrt{T})\Sigma(t/T)\varepsilon_t \end{bmatrix} \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{Q}). \qquad [16.1.24]$$

From [16.1.19] and [16.1.24], the asymptotic distribution of [16.1.18] can be calculated as in Example 7.5 of Chapter 7:

$$\begin{bmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{bmatrix} \xrightarrow{L} N(\mathbf{0}, [\mathbf{Q}^{-1}\cdot\sigma^2\mathbf{Q}\cdot\mathbf{Q}^{-1}]) = N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1}). \qquad [16.1.25]$$

These results can be summarized as follows.

Proposition 16.1: Let y_t be generated according to the simple deterministic time trend [16.1.1] where ε_t is i.i.d. with $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t^4) < \infty$. Then

$$\begin{bmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1}\right). \qquad [16.1.26]$$

Note that the resulting estimate of the coefficient on the time trend ($\hat{\delta}_T$) is *superconsistent*—not only does $\hat{\delta}_T \xrightarrow{p} \delta$, but even when multiplied by T, we still have

$$T(\hat{\delta}_T - \delta) \xrightarrow{p} 0; \qquad [16.1.27]$$

see Exercise 16.2.

Different rates of convergence are sometimes described in terms of *order in probability*. A sequence of random variables $\{X_T\}_{T=1}^{\infty}$ is said to be $O_p(T^{-1/2})$ if for every $\varepsilon > 0$, there exists an $M > 0$ such that

$$P\{|X_T| > M/\sqrt{T}\} < \varepsilon \qquad [16.1.28]$$

for all T; in other words, the random variable $\sqrt{T}\cdot X_T$ is almost certain to fall within $\pm M$ for any T. Most of the estimators encountered for stationary time series are $O_p(T^{-1/2})$. For example, suppose that X_T represents the mean of a sample of size T,

$$X_T = (1/T)\sum_{t=1}^{T} y_t,$$

where $\{y_t\}$ is i.i.d. with mean zero and variance σ^2. Then the variance of X_T is σ^2/T. But Chebyshev's inequality implies that

$$P\{|X_T| > M/\sqrt{T}\} \le \frac{\sigma^2/T}{M^2/T} = (\sigma/M)^2$$

for any M. By choosing M so that $(\sigma/M)^2 < \varepsilon$, condition [16.1.28] is guaranteed. Since the standard deviation of the estimator is $\sigma/\sqrt{T}$, by choosing M to be a suitable multiple of σ, the band $X_T \pm M/\sqrt{T}$ can include as much of the density as desired.

As another example, the estimator $\hat{\alpha}_T$ in [16.1.26] would also be said to be $O_p(T^{-1/2})$. Since $\sqrt{T}$ times $(\hat{\alpha}_T - \alpha)$ is asymptotically Gaussian, there exists a band $\pm M/\sqrt{T}$ around $\hat{\alpha}_T$ that contains as much of the probability distribution as desired.

In general, a sequence of random variables $\{X_T\}_{T=1}^{\infty}$ is said to be $O_p(T^{-k})$ if for every $\varepsilon > 0$ there exists an $M > 0$ such that

$$P\{|X_T| > M/(T^k)\} < \varepsilon. \qquad [16.1.29]$$

Thus, for example, the estimator $\hat{\delta}_T$ in [16.1.26] is $O_p(T^{-3/2})$, since there exists a band $\pm M$ around $T^{3/2}(\hat{\delta}_T - \delta)$ that contains as much of the probability distribution as desired.

16.2. *Hypothesis Testing for the Simple Time Trend Model*

If the innovations ε_t for the simple time trend [16.1.1] are Gaussian, then the *OLS* estimates $\hat{\alpha}_T$ and $\hat{\delta}_T$ are Gaussian and the usual *OLS* t and F tests have exact small-sample t and F distributions for all sample sizes T. Thus, despite the fact that $\hat{\alpha}_T$ and $\hat{\delta}_T$ have different asymptotic rates of convergence, the standard errors $\hat{\sigma}_{\hat{\alpha}_T}$ and $\hat{\sigma}_{\hat{\delta}_T}$ evidently have offsetting asymptotic behavior so that the statistics such as $(\hat{\delta}_T - \delta_0)/\hat{\sigma}_{\hat{\delta}_T}$ are asymptotically $N(0, 1)$ when the innovations are Gaussian. We might thus conjecture that the usual t and F tests are asymptotically valid for non-Gaussian innovations as well. This conjecture is indeed correct, as we now verify.

First consider the *OLS* t test of the null hypothesis $\alpha = \alpha_0$, which can be written as

$$t_T = \frac{\hat{\alpha}_T - \alpha_0}{\left\{s_T^2[1 \quad 0](\mathbf{X}_T'\mathbf{X}_T)^{-1}\begin{bmatrix}1\\0\end{bmatrix}\right\}^{1/2}}. \qquad [16.2.1]$$

Here s_T^2 denotes the *OLS* estimate of σ^2:

$$s_T^2 = [1/(T-2)]\sum_{t=1}^{T}(y_t - \hat{\alpha}_T - \hat{\delta}_T t)^2; \qquad [16.2.2]$$

and $(\mathbf{X}_T'\mathbf{X}_T) = \Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t'$ denotes the matrix in equation [16.1.16]. The numerator and denominator of [16.2.1] can further be multiplied by $\sqrt{T}$, resulting in

$$t_T = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\left\{s_T^2[\sqrt{T} \quad 0](\mathbf{X}_T'\mathbf{X}_T)^{-1}\begin{bmatrix}\sqrt{T}\\0\end{bmatrix}\right\}^{1/2}}. \qquad [16.2.3]$$

Note further from [16.1.17] that

$$[\sqrt{T} \quad 0] = [1 \quad 0]\mathbf{Y}_T. \qquad [16.2.4]$$

Substituting [16.2.4] into [16.2.3],

$$t_T = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\left\{s_T^2[1 \quad 0]\mathbf{Y}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{Y}_T\begin{bmatrix}1\\0\end{bmatrix}\right\}^{1/2}}. \qquad [16.2.5]$$

But recall from [16.1.19] that

$$\mathbf{Y}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{Y}_T = [\mathbf{Y}_T^{-1}(\mathbf{X}_T'\mathbf{X}_T)\mathbf{Y}_T^{-1}]^{-1} \to \mathbf{Q}^{-1}. \qquad [16.2.6]$$

It is straightforward to show that $s_T^2 \xrightarrow{p} \sigma^2$. Recall further that $\sqrt{T}(\hat{\alpha}_T - \alpha_0) \xrightarrow{L} N(0, \sigma^2 q^{11})$ for q^{11} the (1, 1) element of $\mathbf{Q}^{-1}$. Hence, from [16.2.5],

$$t_T \xrightarrow{p} \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\left\{\sigma^2[1 \quad 0]\mathbf{Q}^{-1}\begin{bmatrix}1\\0\end{bmatrix}\right\}^{1/2}} = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\sigma\sqrt{q^{11}}}. \qquad [16.2.7]$$

But this is an asymptotically Gaussian variable divided by the square root of its variance, and so asymptotically it has a $N(0, 1)$ distribution. Thus, the usual *OLS* t test of $\alpha = \alpha_0$ will give an asymptotically valid inference.

Similarly, consider the usual *OLS* t test of $\delta = \delta_0$:

$$t_T = \frac{\hat{\delta}_T - \delta_0}{\left\{s_T^2[0 \quad 1](\mathbf{X}_T'\mathbf{X}_T)^{-1}\begin{bmatrix}0\\1\end{bmatrix}\right\}^{1/2}}.$$

Multiplying numerator and denominator by $T^{3/2}$,

$$t_T = \frac{T^{3/2}(\hat{\delta}_T - \delta_0)}{\left\{s_T^2[0 \quad T^{3/2}](\mathbf{X}_T'\mathbf{X}_T)^{-1}\begin{bmatrix}0\\T^{3/2}\end{bmatrix}\right\}^{1/2}}$$
$$= \frac{T^{3/2}(\hat{\delta}_T - \delta_0)}{\left\{s_T^2[0 \quad 1]\mathbf{Y}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{Y}_T\begin{bmatrix}0\\1\end{bmatrix}\right\}^{1/2}}$$
$$\xrightarrow{p} \frac{T^{3/2}(\hat{\delta}_T - \delta_0)}{\sigma\sqrt{q^{22}}},$$

which again is asymptotically a $N(0, 1)$ variable. Thus, although $\hat{\alpha}_T$ and $\hat{\delta}_T$ converge at different rates, the corresponding standard errors $\hat{\sigma}_{\hat{\alpha}_T}$ and $\hat{\sigma}_{\hat{\delta}_T}$ also incorporate different orders of T, with the result that the usual *OLS* t tests are asymptotically valid.

It is interesting also to consider a test of a single hypothesis involving both α and δ,

$$H_0\colon r_1\alpha + r_2\delta = r,$$

where r_1, r_2, and r are parameters that describe the hypothesis. A t test of H_0 can be obtained from the square root of the *OLS* F test (expression [8.1.32]):[6]

$$t_T = \frac{(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\left\{s_T^2[r_1 \quad r_2](\mathbf{X}_T'\mathbf{X}_T)^{-1}\begin{bmatrix}r_1\\r_2\end{bmatrix}\right\}^{1/2}}.$$

In this case we multiply numerator and denominator by $\sqrt{T}$, the slower rate of convergence among the two estimators $\hat{\alpha}_T$ and $\hat{\delta}_T$:

$$t_T = \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\left\{s_T^2\sqrt{T}[r_1 \quad r_2](\mathbf{X}_T'\mathbf{X}_T)^{-1}\begin{bmatrix}r_1\\r_2\end{bmatrix}\sqrt{T}\right\}^{1/2}}$$
$$= \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\left\{s_T^2\sqrt{T}[r_1 \quad r_2]\mathbf{Y}_T^{-1}\mathbf{Y}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{Y}_T\mathbf{Y}_T^{-1}\begin{bmatrix}r_1\\r_2\end{bmatrix}\sqrt{T}\right\}^{1/2}}$$
$$= \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\{s_T^2\mathbf{r}_T'[\mathbf{Y}_T(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{Y}_T]\mathbf{r}_T\}^{1/2}},$$

where

$$\mathbf{r}_T \equiv \mathbf{Y}_T^{-1}\begin{bmatrix}r_1\\r_2\end{bmatrix}\sqrt{T} = \begin{bmatrix}r_1\\r_2/T\end{bmatrix} \rightarrow \begin{bmatrix}r_1\\0\end{bmatrix}. \qquad [16.2.8]$$

Similarly, recall from [16.1.27] that $\hat{\delta}_T$ is superconsistent, implying that

$$\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r) \xrightarrow{p} \sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r), \qquad [16.2.9]$$

[6]With a single linear restriction as here, $m = 1$ and expression [8.1.32] describes an $F(1, T - k)$ variable when the innovations are Gaussian. But an $F(1, T - k)$ variable is the square of a $t(T - k)$ variable. The test is described here in terms of a t test rather than an F test in order to facilitate comparison with the earlier results in this section.

where δ is the true population value for the time trend parameter. Again applying [16.2.6], it follows that

$$t_T \xrightarrow{p} \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r)}{\left\{\sigma^2[r_1 \quad 0]\mathbf{Q}^{-1}\begin{bmatrix} r_1 \\ 0 \end{bmatrix}\right\}^{1/2}} = \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r)}{\{r_1^2\sigma^2 q^{11}\}^{1/2}}. \qquad [16.2.10]$$

But notice that

$$\begin{aligned} \sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r) &= \sqrt{T}[r_1(\hat{\alpha}_T - \alpha) + r_1\alpha + r_2\delta - r] \\ &= \sqrt{T}[r_1(\hat{\alpha}_T - \alpha)] \end{aligned}$$

under the null hypothesis. Hence, under the null,

$$t_T \xrightarrow{p} \frac{\sqrt{T}[r_1(\hat{\alpha}_T - \alpha)]}{\{r_1^2\sigma^2 q^{11}\}^{1/2}} = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha)}{\{\sigma^2 q^{11}\}^{1/2}},$$

which asymptotically has a $N(0, 1)$ distribution. Thus, again, the usual *OLS* t test of H_0 is valid asymptotically.

This last example illustrates the following general principle: A test involving a single restriction across parameters with different rates of convergence is dominated asymptotically by the parameters with the slowest rates of convergence. This means that a test involving both α and δ that employs the estimated value of δ would have the same asymptotic properties under the null as a test that employs the true value of δ.

Finally, consider a joint test of separate hypotheses about α and δ,

$$H_0: \begin{bmatrix} \alpha \\ \delta \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \delta_0 \end{bmatrix},$$

or, in vector form,

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0.$$

The Wald form of the *OLS* χ^2 test of H_0 is found from [8.2.23] by taking $\mathbf{R} = \mathbf{I}_2$:

$$\begin{aligned} \chi_T^2 &= (\mathbf{b}_T - \boldsymbol{\beta}_0)'[s_T^2(\mathbf{X}_T'\mathbf{X}_T)^{-1}]^{-1}(\mathbf{b}_T - \boldsymbol{\beta}_0) \\ &= (\mathbf{b}_T - \boldsymbol{\beta}_0)'\mathbf{Y}_T[\mathbf{Y}_T s_T^2(\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{Y}_T]^{-1}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}_0) \\ &\xrightarrow{p} [\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}_0)]'[\sigma^2\mathbf{Q}^{-1}]^{-1}[\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}_0)]. \end{aligned}$$

Recalling [16.1.25], this is a quadratic form in a two-dimensional Gaussian vector of the sort considered in Proposition 8.1, from which

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

Thus, again, the usual *OLS* test is asymptotically valid.

16.3. *Asymptotic Inference for an Autoregressive Process Around a Deterministic Time Trend*

The same principles can be used to study a general autoregressive process around a deterministic time trend:

$$y_t = \alpha + \delta t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t. \qquad [16.3.1]$$

It is assumed throughout this section that ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment, and that roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

lie outside the unit circle. Consider a sample of $T + p$ observations on y, $\{y_{-p+1}, y_{-p+2}, \ldots, y_T\}$, and let $\hat{\alpha}_T$, $\hat{\delta}_T$, $\hat{\phi}_{1,T}, \ldots, \hat{\phi}_{p,T}$ denote coefficient estimates based on ordinary least squares estimation of [16.3.1] for $t = 1, 2, \ldots, T$.

A Useful Transformation of the Regressors

By adding and subtracting $\phi_j[\alpha + \delta(t - j)]$ for $j = 1, 2, \ldots, p$ on the right side, the regression model [16.3.1] can equivalently be written as

$$\begin{aligned} y_t = {} & \alpha(1 + \phi_1 + \phi_2 + \cdots + \phi_p) + \delta(1 + \phi_1 + \phi_2 + \cdots + \phi_p)t \\ & - \delta(\phi_1 + 2\phi_2 + \cdots + p\phi_p) + \phi_1[y_{t-1} - \alpha - \delta(t - 1)] \\ & + \phi_2[y_{t-2} - \alpha - \delta(t - 2)] + \cdots \\ & + \phi_p[y_{t-p} - \alpha - \delta(t - p)] + \varepsilon_t \end{aligned} \qquad [16.3.2]$$

or

$$y_t = \alpha^* + \delta^* t + \phi_1^* y_{t-1}^* + \phi_2^* y_{t-2}^* + \cdots + \phi_p^* y_{t-p}^* + \varepsilon_t, \qquad [16.3.3]$$

where

$$\begin{aligned} \alpha^* &\equiv [\alpha(1 + \phi_1 + \phi_2 + \cdots + \phi_p) - \delta(\phi_1 + 2\phi_2 + \cdots + p\phi_p)] \\ \delta^* &\equiv \delta(1 + \phi_1 + \phi_2 + \cdots + \phi_p) \\ \phi_j^* &\equiv \phi_j \qquad \text{for } j = 1, 2, \ldots, p \end{aligned}$$

and

$$y_{t-j}^* \equiv y_{t-j} - \alpha - \delta(t - j) \qquad \text{for } j = 1, 2, \ldots, p. \qquad [16.3.4]$$

The idea of transforming the regression into a form such as [16.3.3] is due to Sims, Stock, and Watson (1990).[7] The objective is to rewrite the regressors of [16.3.1] in terms of zero-mean covariance-stationary random variables (the terms y_{t-j}^* for $j = 1, 2, \ldots, p$), a constant term, and a time trend. Transforming the regressors in this way isolates components of the *OLS* coefficient vector with different rates of convergence and provides a general technique for finding the asymptotic distribution of regressions involving nonstationary variables. A general result is that, if such a transformed equation were estimated by *OLS*, the coefficients on zero-mean covariance-stationary random variables (in this case, $\hat{\phi}_{1,T}^*, \hat{\phi}_{2,T}^*, \ldots, \hat{\phi}_{p,T}^*$) would converge at rate $\sqrt{T}$ to a Gaussian distribution. The coefficients $\hat{\alpha}_T^*$ and $\hat{\delta}_T^*$ from *OLS* estimation of [16.3.3] turn out to behave asymptotically exactly like $\hat{\alpha}_T$ and $\hat{\delta}_T$ for the simple time trend model analyzed in Section 16.1 and are asymptotically independent of the $\hat{\phi}^*$'s.

It is helpful to describe this transformation in more general notation that will also apply to more complicated models in the chapters that follow. The original regression model [16.3.1] can be written

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \qquad [16.3.5]$$

[7]A simpler version of this theme appeared in the analysis of a univariate process with unit roots by Fuller (1976).

where

$$\underset{(p+2)\times 1}{\mathbf{x}_t} \equiv \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \\ 1 \\ t \end{bmatrix} \qquad \underset{(p+2)\times 1}{\boldsymbol{\beta}} \equiv \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \\ \alpha \\ \delta \end{bmatrix} \qquad [16.3.6]$$

The algebraic transformation in arriving at [16.3.3] could then be described as rewriting [16.3.5] in the form

$$y_t = \mathbf{x}_t'\mathbf{G}'[\mathbf{G}']^{-1}\boldsymbol{\beta} + \varepsilon_t = [\mathbf{x}_t^*]'\boldsymbol{\beta}^* + \varepsilon_t, \qquad [16.3.7]$$

where

$$\underset{(p+2)\times(p+2)}{\mathbf{G}'} \equiv \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ -\alpha+\delta & -\alpha+2\delta & \cdots & -\alpha+p\delta & 1 & 0 \\ -\delta & -\delta & \cdots & -\delta & 0 & 1 \end{bmatrix} \qquad [16.3.8]$$

$$\underset{(p+2)\times(p+2)}{[\mathbf{G}']^{-1}} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ \alpha-\delta & \alpha-2\delta & \cdots & \alpha-p\delta & 1 & 0 \\ \delta & \delta & \cdots & \delta & 0 & 1 \end{bmatrix}$$

$$\mathbf{x}_t^* \equiv \mathbf{G}\mathbf{x}_t = \begin{bmatrix} y_{t-1}^* \\ y_{t-2}^* \\ \vdots \\ y_{t-p}^* \\ 1 \\ t \end{bmatrix} \qquad [16.3.9]$$

$$\boldsymbol{\beta}^* \equiv [\mathbf{G}']^{-1}\boldsymbol{\beta} = \begin{bmatrix} \phi_1^* \\ \phi_2^* \\ \vdots \\ \phi_p^* \\ \alpha^* \\ \delta^* \end{bmatrix} \qquad [16.3.10]$$

The system of [16.3.7] is just an algebraically equivalent representation of the regression model [16.3.5]. Notice that the estimate of $\boldsymbol{\beta}^*$ based on an *OLS* regression of y_t on $\mathbf{x}_t^*$ is given by

$$\begin{aligned} \mathbf{b}^* &= \left[\sum_{t=1}^{T} \mathbf{x}_t^*[\mathbf{x}_t^*]'\right]^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t^* y_t\right] \\ &= \left[\mathbf{G}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)\mathbf{G}'\right]^{-1}\mathbf{G}\left(\sum_{t=1}^{T} \mathbf{x}_t y_t\right) \\ &= [\mathbf{G}']^{-1}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{G}^{-1}\mathbf{G}\left(\sum_{t=1}^{T} \mathbf{x}_t y_t\right) \\ &= [\mathbf{G}']^{-1}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\sum_{t=1}^{T} \mathbf{x}_t y_t\right) \\ &= [\mathbf{G}']^{-1}\mathbf{b}, \end{aligned} \qquad [16.3.11]$$

where $\mathbf{b}$ denotes the estimated coefficient vector from an *OLS* regression of y_t on $\mathbf{x}_t$. Thus, the coefficient estimate for the transformed regression ($\mathbf{b}^*$) is a simple linear transformation of the coefficient estimate for the original system ($\mathbf{b}$). The fitted value for date t associated with the transformed regression is

$$[\mathbf{x}_t^*]'\mathbf{b}^* = [\mathbf{G}\mathbf{x}_t]'[\mathbf{G}']^{-1}\mathbf{b} = \mathbf{x}_t'\mathbf{b}.$$

Thus, the fitted values for the transformed regression are numerically identical to the fitted values from the original regression.

Of course, given data only on $\{y_t\}$, we could not actually estimate the transformed regression by *OLS*, because construction of $\mathbf{x}_t^*$ from $\mathbf{x}_t$ requires knowledge of the true values of the parameters α and δ. It is nevertheless helpful to summarize the properties of hypothetical *OLS* estimation of [16.3.7], because [16.3.7] is easier to analyze than [16.3.5]. Moreover, once we find the asymptotic distribution of $\mathbf{b}^*$, the asymptotic distribution of $\mathbf{b}$ can be inferred by inverting [16.3.11]:

$$\mathbf{b} = \mathbf{G}'\mathbf{b}^*. \qquad [16.3.12]$$

The Asymptotic Distribution of OLS *Estimates for the Transformed Regression*

Appendix 16.A to this chapter demonstrates that

$$\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*) \xrightarrow{L} N(0, \sigma^2[\mathbf{Q}^*]^{-1}), \qquad [16.3.13]$$

where

$$\underset{(p+2)\times(p+2)}{\mathbf{Y}_T} = \begin{bmatrix} \sqrt{T} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \sqrt{T} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{T} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sqrt{T} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & T^{3/2} \end{bmatrix} \qquad [16.3.14]$$

$$\underset{(p+2)\times(p+2)}{\mathbf{Q}^*} = \begin{bmatrix} \gamma_0^* & \gamma_1^* & \gamma_2^* & \cdots & \gamma_{p-1}^* & 0 & 0 \\ \gamma_1^* & \gamma_0^* & \gamma_1^* & \cdots & \gamma_{p-2}^* & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \gamma_{p-1}^* & \gamma_{p-2}^* & \gamma_{p-3}^* & \cdots & \gamma_0^* & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} & \frac{1}{3} \end{bmatrix} \qquad [16.3.15]$$

for $\gamma_j^* \equiv E(y_t^* y_{t-j}^*)$. In other words, the *OLS* estimate $\mathbf{b}^*$ is asymptotically Gaussian, with the coefficient on the time trend ($\hat{\delta}^*$) converging at rate $T^{3/2}$ and all other coefficients converging at rate $\sqrt{T}$. The earlier result [16.1.26] is a special case of [16.3.13] with $p = 0$.

The Asymptotic Distribution of OLS *Estimates for the Original Regression*

What does this result imply about the asymptotic distribution of $\mathbf{b}$, the estimated coefficient vector for the *OLS* regression that is actually estimated? Writing out [16.3.12] explicitly using [16.3.8], we have

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \\ \hat{\alpha} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ -\alpha + \delta & -\alpha + 2\delta & \cdots & -\alpha + p\delta & 1 & 0 \\ -\delta & -\delta & \cdots & -\delta & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_1^* \\ \hat{\phi}_2^* \\ \vdots \\ \hat{\phi}_p^* \\ \hat{\alpha}^* \\ \hat{\delta}^* \end{bmatrix}. \qquad [16.3.16]$$

The *OLS* estimates $\hat{\phi}_j$ of the untransformed regression are identical to the corresponding coefficients of the transformed regression $\hat{\phi}_j^*$, so the asymptotic distribution of $\hat{\phi}_j$ is given immediately by [16.3.13]. The estimate $\hat{\alpha}_T$ is a linear combination of variables that converge to a Gaussian distribution at rate $\sqrt{T}$, and so $\hat{\alpha}_T$ behaves the same way. Specifically, $\hat{\alpha}_T = \mathbf{g}_\alpha' \mathbf{b}_T^*$, where

$$\mathbf{g}_\alpha' \equiv [-\alpha + \delta \quad -\alpha + 2\delta \quad \cdots \quad -\alpha + p\delta \quad 1 \quad 0],$$

and so, from [16.3.13],

$$\sqrt{T}(\hat{\alpha}_T - \alpha) \xrightarrow{L} N(0, \sigma^2 \mathbf{g}_\alpha' [\mathbf{Q}^*]^{-1} \mathbf{g}_\alpha). \qquad [16.3.17]$$

Finally, the estimate $\hat{\delta}_T$ is a linear combination of variables converging at different rates:

$$\hat{\delta}_T = \mathbf{g}_\delta' \mathbf{b}_T^* + \hat{\delta}_T^*,$$

where

$$\mathbf{g}_\delta' \equiv [-\delta \quad -\delta \quad \cdots \quad -\delta \quad 0 \quad 0].$$

Its asymptotic distribution is governed by the variables with the slowest rate of convergence:

$$\begin{aligned} \sqrt{T}(\hat{\delta}_T - \delta) &= \sqrt{T}(\hat{\delta}_T^* + \mathbf{g}_\delta' \mathbf{b}_T^* - \delta^* - \mathbf{g}_\delta' \boldsymbol{\beta}^*) \\ &\xrightarrow{p} \sqrt{T}(\delta^* + \mathbf{g}_\delta' \mathbf{b}_T^* - \delta^* - \mathbf{g}_\delta' \boldsymbol{\beta}^*) \\ &= \mathbf{g}_\delta' \sqrt{T}(\mathbf{b}_T^* - \boldsymbol{\beta}^*) \\ &\xrightarrow{L} N(0, \sigma^2 \mathbf{g}_\delta' [\mathbf{Q}^*]^{-1} \mathbf{g}_\delta). \end{aligned}$$

Thus, each of the elements of $\mathbf{b}_T$ individually is asymptotically Gaussian and $O_p(T^{-1/2})$. The asymptotic distribution of the full vector $\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta})$ is multivariate Gaussian, though with a singular variance-covariance matrix. Specifically, the particular linear combination of the elements of $\mathbf{b}_T$ that recovers $\hat{\delta}_T^*$, the time trend coefficient of the hypothetical regression,

$$\hat{\delta}_T^* = -\mathbf{g}_\delta' \mathbf{b}_T^* + \hat{\delta}_T = \delta \hat{\phi}_{1,T} + \delta \hat{\phi}_{2,T} + \cdots + \delta \hat{\phi}_{p,T} + \hat{\delta}_T,$$

converges to a point mass around δ^* even when scaled by $\sqrt{T}$:

$$\sqrt{T}(\hat{\delta}_T^* - \delta^*) \xrightarrow{p} 0.$$

However, [16.3.13] establishes that

$$T^{3/2}(\hat{\delta}_T^* - \delta^*) \xrightarrow{L} N\left(0, \sigma^2 (q^*)^{p+2,p+2}\right)$$

for $(q^*)^{p+2,p+2}$ the bottom right element of $[\mathbf{Q}^*]^{-1}$.

Hypothesis Tests

The preceding analysis described the asymptotic distribution of $\mathbf{b}$ in terms of the properties of the transformed regression estimates $\mathbf{b}^*$. This might seem to imply that knowledge of the transformation matrix $\mathbf{G}$ in [16.3.8] is necessary in order to conduct hypothesis tests. Fortunately, this is not the case. The results of Section 16.2 turn out to apply equally well to the general model [16.3.1]—the usual t and F tests about $\boldsymbol{\beta}$ calculated in the usual way on the untransformed system are all asymptotically valid.

Consider the following null hypothesis about the parameters of the untransformed system:

$$H_0\colon \mathbf{R}\boldsymbol{\beta} = \mathbf{r}. \tag{16.3.18}$$

Here $\mathbf{R}$ is a known $[m \times (p + 2)]$ matrix, $\mathbf{r}$ is a known $(m \times 1)$ vector, and m is the number of restrictions. The Wald form of the OLS χ^2 test of H_0 (expression [8.2.23]) is

$$\chi_T^2 = (\mathbf{R}\mathbf{b}_T - \mathbf{r})'\left[s_T^2\mathbf{R}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{R}'\right]^{-1}(\mathbf{R}\mathbf{b}_T - \mathbf{r}). \tag{16.3.19}$$

Here $\mathbf{b}_T$ is the OLS estimate of $\boldsymbol{\beta}$ based on observation of $\{y_{-p+1}, y_{-p+2}, \ldots, y_0, y_1, \ldots, y_T\}$ and $s_T^2 = [1/(T - p - 2)]\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\mathbf{b}_T)^2$.

Under the null hypothesis [16.3.18], expression [16.3.19] can be rewritten

$$\begin{aligned}
\chi_T^2 &= [\mathbf{R}(\mathbf{b}_T - \boldsymbol{\beta})]'\left[s_T^2\mathbf{R}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{R}'\right]^{-1}[\mathbf{R}(\mathbf{b}_T - \boldsymbol{\beta})] \\
&= [\mathbf{R}\mathbf{G}'(\mathbf{G}')^{-1}(\mathbf{b}_T - \boldsymbol{\beta})]' \\
&\quad\times\left[s_T^2\mathbf{R}\mathbf{G}'(\mathbf{G}')^{-1}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}(\mathbf{G})^{-1}\mathbf{G}\mathbf{R}'\right]^{-1}[\mathbf{R}\mathbf{G}'(\mathbf{G}')^{-1}(\mathbf{b}_T - \boldsymbol{\beta})].
\end{aligned} \tag{16.3.20}$$

Notice that

$$(\mathbf{G}')^{-1}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}(\mathbf{G})^{-1} = \left[\mathbf{G}\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)\mathbf{G}'\right]^{-1} = \left(\sum_{t=1}^{T} \mathbf{x}_t^*[\mathbf{x}_t^*]'\right)^{-1}$$

for $\mathbf{x}_t^*$ given by [16.3.9]. Similarly, from [16.3.10] and [16.3.11],

$$(\mathbf{b}_T^* - \boldsymbol{\beta}^*) = (\mathbf{G}')^{-1}(\mathbf{b}_T - \boldsymbol{\beta}).$$

Defining

$$\mathbf{R}^* \equiv \mathbf{R}\mathbf{G}',$$

expression [16.3.20] can be written

$$\begin{aligned}
\chi_T^2 &= [\mathbf{R}^*(\mathbf{b}_T^* - \boldsymbol{\beta}^*)]'\left[s_T^2\mathbf{R}^*\left(\sum_{t=1}^{T} \mathbf{x}_t^*[\mathbf{x}_t^*]'\right)^{-1}[\mathbf{R}^*]'\right]^{-1} \\
&\quad\times[\mathbf{R}^*(\mathbf{b}_T^* - \boldsymbol{\beta}^*)].
\end{aligned} \tag{16.3.21}$$

Expression [16.3.21] will be recognized as the χ^2 test that would be calculated if we had estimated the transformed system and wanted to test the hypothesis that $\mathbf{R}^*\boldsymbol{\beta}^* = \mathbf{r}$ (recall that the fitted values for the transformed and untransformed regressions are identical, so that s_T^2 will be the same value for either representation). Observe that the transformed regression does not actually have to be estimated in order to calculate this statistic, since [16.3.21] is numerically identical to the χ^2 statistic [16.3.20] that is calculated from the untransformed system in the usual way. Nevertheless, expression [16.3.21] gives us another way of thinking about the distribution of the statistic as actually calculated in [16.3.20].

Expression [16.3.21] can be further rewritten as

$$\begin{aligned}\chi_T^2 &= [\mathbf{R}^*\mathbf{Y}_T^{-1}\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)]' \\ &\times \left[s_T^2\mathbf{R}^*\mathbf{Y}_T^{-1}\mathbf{Y}_T\left(\sum_{t=1}^{T}\mathbf{x}_t^*[\mathbf{x}_t^*]'\right)^{-1}\mathbf{Y}_T\mathbf{Y}_T^{-1}[\mathbf{R}^*]'\right]^{-1} \\ &\times [\mathbf{R}^*\mathbf{Y}_T^{-1}\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)]\end{aligned} \qquad [16.3.22]$$

for $\mathbf{Y}_T$ the matrix in [16.3.14]. Recall the insight from Section 16.2 that hypothesis tests involving coefficients with different rates of convergence will be dominated by the variables with the slowest rate of convergence. This means that some of the elements of $\mathbf{R}^*$ may be irrelevant asymptotically, so that [16.3.22] has the same asymptotic distribution as a simpler expression. To describe this expression, consider two possibilities.

Case 1. Each of the m *Hypotheses Represented by* $\mathbf{R}^*\boldsymbol{\beta}^* = \mathbf{r}$ *Involves a Parameter that Converges at Rate* $\sqrt{T}$

Of course, we could trivially rewrite any system of restrictions so as to involve $O_p(T^{-1/2})$ parameters in every equation. For example, the null hypothesis

$$H_0\colon \phi_2^* = 0, \qquad \delta^* = 0 \qquad [16.3.23]$$

could be rewritten as

$$H_0\colon \phi_2^* = 0, \qquad \delta^* = \phi_2^*, \qquad [16.3.24]$$

which seems to include ϕ_2^* in each restriction. For purposes of *implementing* a test of H_0, it does not matter which representation of H_0 is used, since either will produce the identical value for the test statistic.[8] For purposes of *analyzing the properties* of the test, we distinguish a hypothesis such as [16.3.23] from a hypothesis involving only ϕ_2^* and ϕ_3^*. For this distinction to be meaningful, we will assume that H_0 would be written in the form of [16.3.23] rather than [16.3.24].

[8]More generally, let $\mathbf{H}$ be any nonsingular $(m \times m)$ matrix. Then the null hypothesis $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ can equivalently be written as $\dot{\mathbf{R}}\boldsymbol{\beta} = \dot{\mathbf{r}}$, where $\dot{\mathbf{R}} \equiv \mathbf{HR}$ and $\dot{\mathbf{r}} \equiv \mathbf{Hr}$. The χ^2 statistic constructed from the second parameterization is

$$\begin{aligned}\chi^2 &= (\dot{\mathbf{R}}\mathbf{b} - \dot{\mathbf{r}})'\left[s_T^2\dot{\mathbf{R}}\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\dot{\mathbf{R}}'\right]^{-1}(\dot{\mathbf{R}}\mathbf{b} - \dot{\mathbf{r}}) \\ &= (\mathbf{Rb} - \mathbf{r})'\mathbf{H}'[\mathbf{H}']^{-1}\left[s_T^2\mathbf{R}\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{R}'\right]^{-1}\mathbf{H}^{-1}\mathbf{H}(\mathbf{Rb} - \mathbf{r}) \\ &= (\mathbf{Rb} - \mathbf{r})'\left[s_T^2\mathbf{R}\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{R}'\right]^{-1}(\mathbf{Rb} - \mathbf{r}),\end{aligned}$$

which is identical to the χ^2 statistic constructed from the first parameterization. The representation [16.3.24] is an example of such a transformation of [16.3.23], with

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

In general terms, this means that $\mathbf{R}^*$ is "upper triangular."[9] "Case 1" describes the situation in which the first $p + 1$ elements of the last row of $\mathbf{R}^*$ are not all zero.

For case 1, even though some of the hypotheses may involve $\hat{\delta}_T^*$, a test of the null hypothesis will be asymptotically equivalent to a test that treated δ^* as if known with certainty. This is a consequence of $\hat{\delta}_T^*$ being superconsistent. To develop this result rigorously, notice that

$$\mathbf{R}^*\mathbf{Y}_T^{-1} = \begin{bmatrix} r_{11}^*/\sqrt{T} & r_{12}^*/\sqrt{T} & \cdots & r_{1,p+1}^*/\sqrt{T} & r_{1,p+2}^*/T^{3/2} \\ r_{21}^*/\sqrt{T} & r_{22}^*/\sqrt{T} & \cdots & r_{2,p+1}^*/\sqrt{T} & r_{2,p+2}^*/T^{3/2} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{m1}^*/\sqrt{T} & r_{m2}^*/\sqrt{T} & \cdots & r_{m,p+1}^*/\sqrt{T} & r_{m,p+2}^*/T^{3/2} \end{bmatrix},$$

and define

$$\underset{(m\times m)}{\tilde{\mathbf{Y}}_T} \equiv \sqrt{T}\mathbf{I}_m$$

$$\tilde{\mathbf{R}}_T^* \equiv \begin{bmatrix} r_{11}^* & r_{12}^* & \cdots & r_{1,p+1}^* & r_{1,p+2}^*/T \\ r_{21}^* & r_{22}^* & \cdots & r_{2,p+1}^* & r_{2,p+2}^*/T \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{m1}^* & r_{m2}^* & \cdots & r_{m,p+1}^* & r_{m,p+2}^*/T \end{bmatrix}.$$

These matrices were chosen so that

$$\mathbf{R}^*\mathbf{Y}_T^{-1} = \tilde{\mathbf{Y}}_T^{-1}\tilde{\mathbf{R}}_T^*. \qquad [16.3.25]$$

The matrix $\tilde{\mathbf{R}}_T^*$ has the further property that

$$\tilde{\mathbf{R}}_T^* \to \tilde{\mathbf{R}}^*, \qquad [16.3.26]$$

where $\tilde{\mathbf{R}}^*$ involves only those restrictions that affect the asymptotic distribution:

$$\tilde{\mathbf{R}}^* = \begin{bmatrix} r_{11}^* & r_{12}^* & \cdots & r_{1,p+1}^* & 0 \\ r_{21}^* & r_{22}^* & \cdots & r_{2,p+1}^* & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{m1}^* & r_{m2}^* & \cdots & r_{m,p+1}^* & 0 \end{bmatrix}.$$

[9] "Upper triangular" means that if the set of restrictions in H_0 involves parameters $\beta_{i_1}^*, \beta_{i_2}^*, \ldots, \beta_{i_n}^*$ with $i_1 < i_2 < \cdots < i_n$, then elements of $\mathbf{R}^*$ in rows 2 through m and columns 1 through i_1 are all zero. This is simply a normalization—any hypothesis $\mathbf{R}^*\boldsymbol{\beta}^* = \mathbf{r}$ can be written in such a form by selecting a restriction involving $\beta_{i_1}^*$ to be the first row of $\mathbf{R}^*$ and then multiplying the first row of this system of equations by a suitable constant and subtracting it from each of the following rows. If the system of restrictions represented by rows 2 through m of the resulting matrix involves parameters $\beta_{j_1}^*, \beta_{j_2}^*, \ldots, \beta_{j_l}^*$ with $j_1 < j_2 < \cdots < j_l$, then it is assumed that the elements in rows 3 through m and columns 1 through j_1 are all zero. An example of an upper triangular system is

$$\mathbf{R}^* = \begin{bmatrix} 0 & r_{1,i_1}^* & r_{1,i_2}^* & 0 & \cdots & 0 & r_{1,i_n}^* \\ 0 & 0 & 0 & r_{2,j_1}^* & \cdots & r_{2,j_l}^* & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & r_{m,k_z-1}^* & r_{m,k_z}^* \end{bmatrix}.$$

Substituting [16.3.25] into [16.3.22],

$$\begin{aligned}
\chi_T^2 &= [\tilde{\mathbf{Y}}_T^{-1}\tilde{\mathbf{R}}_T^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)]' \\
&\quad \times \left[s_T^2\tilde{\mathbf{Y}}_T^{-1}\tilde{\mathbf{R}}_T^*\mathbf{Y}_T\left(\sum_{t=1}^T \mathbf{x}_t^*[\mathbf{x}_t^*]'\right)^{-1}\mathbf{Y}_T[\tilde{\mathbf{Y}}_T^{-1}\tilde{\mathbf{R}}_T^*]'\right]^{-1}[\tilde{\mathbf{Y}}_T^{-1}\tilde{\mathbf{R}}_T^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)] \\
&= [\tilde{\mathbf{R}}_T^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)]'\tilde{\mathbf{Y}}_T^{-1} \\
&\quad \times \tilde{\mathbf{Y}}_T\left[s_T^2\tilde{\mathbf{R}}_T^*\mathbf{Y}_T\left(\sum_{t=1}^T \mathbf{x}_t^*[\mathbf{x}_t^*]'\right)^{-1}\mathbf{Y}_T[\tilde{\mathbf{R}}_T^*]'\right]^{-1}\tilde{\mathbf{Y}}_T\tilde{\mathbf{Y}}_T^{-1}[\tilde{\mathbf{R}}_T^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)] \\
&= [\tilde{\mathbf{R}}_T^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)]' \\
&\quad \times \left[s_T^2\tilde{\mathbf{R}}_T^*\mathbf{Y}_T\left(\sum_{t=1}^T \mathbf{x}_t^*[\mathbf{x}_t^*]'\right)^{-1}\mathbf{Y}_T[\tilde{\mathbf{R}}_T^*]'\right]^{-1}[\tilde{\mathbf{R}}_T^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)] \\
&\xrightarrow{p} [\tilde{\mathbf{R}}^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)]'\left[\sigma^2\tilde{\mathbf{R}}^*[\mathbf{Q}^*]^{-1}[\tilde{\mathbf{R}}^*]'\right]^{-1}[\tilde{\mathbf{R}}^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*)] \qquad [16.3.27]
\end{aligned}$$

by virtue of [16.3.26] and [16.A.4].

Now [16.3.13] implies that

$$\tilde{\mathbf{R}}^*\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*) \xrightarrow{L} N(\mathbf{0}, \tilde{\mathbf{R}}^*\sigma^2[\mathbf{Q}^*]^{-1}[\tilde{\mathbf{R}}^*]'),$$

and so [16.3.27] is a quadratic form in an asymptotically Gaussian variable of the kind covered in Proposition 8.1. It is therefore asymptotically $\chi^2(m)$. Since [16.3.27] is numerically identical to [16.3.19], the Wald form of the *OLS* χ^2 test, calculated in the usual way from the untransformed regression [16.3.1], has the usual $\chi^2(m)$ distribution.

Case 2. One of the Hypotheses Involves Only the Time Trend Parameter δ^*

Again assuming for purposes of discussion that $\mathbf{R}^*$ is upper triangular, for case 2 the hypothesis about δ^* will be the sole entry in the mth row of $\mathbf{R}^*$:

$$\mathbf{R}^* = \begin{bmatrix} r_{11}^* & r_{12}^* & \cdots & r_{1,p+1}^* & r_{1,p+2}^* \\ r_{21}^* & r_{22}^* & \cdots & r_{2,p+1}^* & r_{2,p+2}^* \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{m-1,1}^* & r_{m-1,2}^* & \cdots & r_{m-1,p+1}^* & r_{m-1,p+2}^* \\ 0 & 0 & \cdots & 0 & r_{m,p+2}^* \end{bmatrix}.$$

For this case, define

$$\underset{(m\times m)}{\tilde{\mathbf{Y}}_T} \equiv \begin{bmatrix} \sqrt{T} & 0 & \cdots & 0 & 0 \\ 0 & \sqrt{T} & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{T} & 0 \\ 0 & 0 & \cdots & 0 & T^{3/2} \end{bmatrix}$$

and

$$\tilde{\mathbf{R}}_T^* \equiv \begin{bmatrix} r_{11}^* & r_{12}^* & \cdots & r_{1,p+1}^* & r_{1,p+2}^*/T \\ r_{21}^* & r_{22}^* & \cdots & r_{2,p+1}^* & r_{2,p+2}^*/T \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{m-1,1}^* & r_{m-1,2}^* & \cdots & r_{m-1,p+1}^* & r_{m-1,p+2}^*/T \\ 0 & 0 & \cdots & 0 & r_{m,p+2}^* \end{bmatrix}.$$

Notice that these matrices again satisfy [16.3.25] and [16.3.26] with

$$\tilde{\mathbf{R}}^* = \begin{bmatrix} r_{11}^* & r_{12}^* & \cdots & r_{1,p+1}^* & 0 \\ r_{21}^* & r_{22}^* & \cdots & r_{2,p+1}^* & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ r_{m-1,1}^* & r_{m-1,2}^* & \cdots & r_{m-1,p+1}^* & 0 \\ 0 & 0 & \cdots & 0 & r_{m,p+2}^* \end{bmatrix}.$$

The analysis of [16.3.27] thus goes through for this case as well with no change.

Summary

Any standard *OLS* χ^2 test of the null hypothesis $\mathbf{R\beta} = \mathbf{r}$ for the regression model [16.3.1] can be calculated and interpreted in the usual way. The test is asymptotically valid for any hypothesis about any subset of the parameters in $\boldsymbol{\beta}$. The elements of $\mathbf{R}$ do not have to be ordered or expressed in any particular form for this to be true.

APPENDIX 16.A. *Derivation of Selected Equations for Chapter 16*

■ **Derivation of [16.3.13].** As in [16.1.6],

$$\mathbf{b}_T^* - \boldsymbol{\beta}^* = \left[\sum_{t=1}^{T} \mathbf{x}_t^*[\mathbf{x}_t^*]'\right]^{-1}\left[\sum_{t=1}^{T} \mathbf{x}_t^*\varepsilon_t\right], \qquad [16.A.1]$$

since the population residuals ε_t are identical for the transformed and untransformed representations. As in [16.1.18], premultiply by $\mathbf{Y}_T$ to write

$$\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*) = \left\{\mathbf{Y}_T^{-1}\sum_{t=1}^{T} \mathbf{x}_t^*[\mathbf{x}_t^*]'\mathbf{Y}_T^{-1}\right\}^{-1}\left\{\mathbf{Y}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t^*\varepsilon_t\right\}. \qquad [16.A.2]$$

From [16.3.9],

$$\sum_{t=1}^{T} \mathbf{x}_t^*[\mathbf{x}_t^*]' = \begin{bmatrix} \Sigma(y_{t-1}^*)^2 & \Sigma y_{t-1}^* y_{t-2}^* & \cdots & \Sigma y_{t-1}^* y_{t-p}^* & \Sigma y_{t-1}^* & \Sigma t y_{t-1}^* \\ \Sigma y_{t-2}^* y_{t-1}^* & \Sigma(y_{t-2}^*)^2 & \cdots & \Sigma y_{t-2}^* y_{t-p}^* & \Sigma y_{t-2}^* & \Sigma t y_{t-2}^* \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \Sigma y_{t-p}^* y_{t-1}^* & \Sigma y_{t-p}^* y_{t-2}^* & \cdots & \Sigma(y_{t-p}^*)^2 & \Sigma y_{t-p}^* & \Sigma t y_{t-p}^* \\ \Sigma y_{t-1}^* & \Sigma y_{t-2}^* & \cdots & \Sigma y_{t-p}^* & \Sigma 1 & \Sigma t \\ \Sigma t y_{t-1}^* & \Sigma t y_{t-2}^* & \cdots & \Sigma t y_{t-p}^* & \Sigma t & \Sigma t^2 \end{bmatrix}$$

and

$$\mathbf{Y}_T^{-1}\sum_{t=1}^{T} \mathbf{x}_t^*[\mathbf{x}_t^*]'\mathbf{Y}_T^{-1}$$

$$= \begin{bmatrix} T^{-1}\Sigma(y_{t-1}^*)^2 & T^{-1}\Sigma y_{t-1}^* y_{t-2}^* & \cdots & T^{-1}\Sigma y_{t-1}^* y_{t-p}^* & T^{-1}\Sigma y_{t-1}^* & T^{-2}\Sigma t y_{t-1}^* \\ T^{-1}\Sigma y_{t-2}^* y_{t-1}^* & T^{-1}\Sigma(y_{t-2}^*)^2 & \cdots & T^{-1}\Sigma y_{t-2}^* y_{t-p}^* & T^{-1}\Sigma y_{t-2}^* & T^{-2}\Sigma t y_{t-2}^* \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ T^{-1}\Sigma y_{t-p}^* y_{t-1}^* & T^{-1}\Sigma y_{t-p}^* y_{t-2}^* & \cdots & T^{-1}\Sigma(y_{t-p}^*)^2 & T^{-1}\Sigma y_{t-p}^* & T^{-2}\Sigma t y_{t-p}^* \\ T^{-1}\Sigma y_{t-1}^* & T^{-1}\Sigma y_{t-2}^* & \cdots & T^{-1}\Sigma y_{t-p}^* & T^{-1}\cdot T & T^{-2}\cdot\Sigma t \\ T^{-2}\Sigma t y_{t-1}^* & T^{-2}\Sigma t y_{t-2}^* & \cdots & T^{-2}\Sigma t y_{t-p}^* & T^{-2}\cdot\Sigma t & T^{-3}\cdot\Sigma t^2 \end{bmatrix}.$$

[16.A.3]

For the first p rows and columns, the row i, column j element of this matrix is

$$T^{-1}\sum_{t=1}^{T} y_{t-i}^{*}y_{t-j}^{*}.$$

But y_t^* follows a zero-mean stationary $AR(p)$ process satisfying the conditions of Exercise 7.7. Thus, these terms converge in probability to $\gamma_{|i-j|}^{*}$. The first p elements of row $p + 1$ (or the first p elements of column $p + 1$) are of the form

$$T^{-1}\sum_{t=1}^{T} y_{t-i}^{*},$$

which converge in probability to zero. The first p elements of row $p + 2$ (or the first p elements of column $p + 2$) are of the form

$$T^{-1}\sum_{t=1}^{T} (t/T)y_{t-i}^{*},$$

which can be shown to converge in probability to zero with a ready adaptation of the techniques in Chapter 7 (see Exercise 16.3). Finally, the (2×2) matrix in the bottom right corner of [16.A.3] converges to

$$\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

Thus

$$\mathbf{Y}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t^{*}[\mathbf{x}_t^{*}]'\mathbf{Y}_T^{-1} \xrightarrow{p} \mathbf{Q}^{*} \qquad [16.A.4]$$

for $\mathbf{Q}^*$ the matrix in [16.3.15].

Turning next to the second term in [16.A.2],

$$\mathbf{Y}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t^{*}\varepsilon_t = \begin{bmatrix} T^{-1/2}\Sigma y_{t-1}^{*}\varepsilon_t \\ T^{-1/2}\Sigma y_{t-2}^{*}\varepsilon_t \\ \vdots \\ T^{-1/2}\Sigma y_{t-p}^{*}\varepsilon_t \\ T^{-1/2}\Sigma \varepsilon_t \\ T^{-1/2}\Sigma (t/T)\varepsilon_t \end{bmatrix} = T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\xi}_t, \qquad [16.A.5]$$

where

$$\boldsymbol{\xi}_t \equiv \begin{bmatrix} y_{t-1}^{*}\varepsilon_t \\ y_{t-2}^{*}\varepsilon_t \\ \vdots \\ y_{t-p}^{*}\varepsilon_t \\ \varepsilon_t \\ (t/T)\varepsilon_t \end{bmatrix}.$$

But $\boldsymbol{\xi}_t$ is a martingale difference sequence with variance

$$E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \sigma^2\mathbf{Q}_t^{*},$$

where

$$\mathbf{Q}_t^{*} = \begin{bmatrix} \gamma_0^{*} & \gamma_1^{*} & \gamma_2^{*} & \cdots & \gamma_{p-1}^{*} & 0 & 0 \\ \gamma_1^{*} & \gamma_0^{*} & \gamma_1^{*} & \cdots & \gamma_{p-2}^{*} & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \gamma_{p-1}^{*} & \gamma_{p-2}^{*} & \gamma_{p-3}^{*} & \cdots & \gamma_0^{*} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & t/T \\ 0 & 0 & 0 & \cdots & 0 & t/T & t^2/T^2 \end{bmatrix}$$

and

$$(1/T) \sum_{t=1}^{T} \mathbf{Q}_t^* \rightarrow \mathbf{Q}^*.$$

Applying the arguments used in Exercise 8.3 and in [16.1.24], it can be shown that

$$\mathbf{Y}_T^{-1} \sum_{t=1}^{T} \mathbf{x}_t^* \varepsilon_t \xrightarrow{L} N(\mathbf{0}, \sigma^2 \mathbf{Q}^*). \qquad [16.A.6]$$

It follows from [16.A.4], [16.A.6], and [16.A.2] that

$$\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*) \xrightarrow{L} N(\mathbf{0}, [\mathbf{Q}^*]^{-1}\sigma^2\mathbf{Q}^*[\mathbf{Q}^*]^{-1}) = N(\mathbf{0}, \sigma^2[\mathbf{Q}^*]^{-1}),$$

as claimed in [16.3.13]. ∎

Chapter 16 Exercises

16.1. Verify result [16.1.23].

16.2. Verify expression [16.1.27].

16.3. Let y_t be covariance-stationary with mean zero and absolutely summable autocovariances:

$$\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$$

for $\gamma_j = E(y_t y_{t-j})$. Adapting the argument in expression [7.2.6], show that

$$T^{-1} \sum_{t=1}^{T} (t/T) y_t \xrightarrow{m.s.} 0.$$

Chapter 16 References

Fuller, Wayne A. 1976. *Introduction to Statistical Time Series.* New York: Wiley.

Sims, Christopher A., James H. Stock, and Mark W. Watson. 1990. "Inference in Linear Time Series Models with Some Unit Roots." *Econometrica* 58:113–44.

17

Univariate Processes with Unit Roots

This chapter discusses statistical inference for univariate processes containing a unit root. Section 17.1 gives a brief explanation of why the asymptotic distributions and rates of convergence for the estimated coefficients of unit root processes differ from those for stationary processes. The asymptotic distributions for unit root processes can be described in terms of functionals on Brownian motion. The basic idea behind Brownian motion is introduced in Section 17.2. The technical tools used to establish that the asymptotic distributions of certain statistics involving unit root processes can be represented in terms of such functionals are developed in Section 17.3, though it is not necessary to master these tools in order to read Sections 17.4 through 17.9. Section 17.4 derives the asymptotic distribution of the estimated coefficient for a first-order autoregression when the true process is a random walk. This distribution turns out to depend on whether a constant or time trend is included in the estimated regression and whether the true random walk is characterized by nonzero drift.

Section 17.5 extends the results of Section 17.3 to cover unit root processes whose differences exhibit general serial correlation. These results can be used to develop two different classes of tests for unit roots. One approach, due to Phillips and Perron (1988), adjusts the statistics calculated from a simple first-order autoregression to account for serial correlation of the differenced data. The second approach, due to Dickey and Fuller (1979), adds lags to the autoregression. These approaches are reviewed in Sections 17.6 and 17.7, respectively. Section 17.7 further derives the properties of all of the estimated coefficients for a pth-order autoregression when one of the roots is unity.

Readers interested solely in how these results are applied in practice may want to begin with the summaries in Table 17.2 or Table 17.3 and with the empirical applications described in Examples 17.6 through 17.9.

17.1. *Introduction*

Consider OLS estimation of a Gaussian $AR(1)$ process,

$$y_t = \rho y_{t-1} + u_t, \qquad [17.1.1]$$

where $u_t \sim$ i.i.d. $N(0, \sigma^2)$, and $y_0 = 0$. The OLS estimate of ρ is given by

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2}. \qquad [17.1.2]$$