——— and ———. 1989. "Statistical Inference in Regressions with Integrated Processes: Part 2." *Econometric Theory* 5:95–131.

Phillips, P. C. B. 1986. "Understanding Spurious Regressions in Econometrics." *Journal of Econometrics* 33:311–40.

———. 1987. "Time Series Regression with a Unit Root." *Econometrica* 55:277–301.

———. 1988. "Regression Theory for Near-Integrated Time Series." *Econometrica* 56:1021–43.

———. 1991a. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends." *Journal of Applied Econometrics* 6:333–64.

———. 1991b. "Bayesian Routes and Unit Roots: De Rebus Prioribus Semper Est Disputandum." *Journal of Applied Econometrics* 6:435–73.

——— and Pierre Perron. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrika* 75:335–46.

——— and Victor Solo. 1992. "Asymptotics for Linear Processes." *Annals of Statistics* 20:971–1001.

Said, Said E. 1991. "Unit-Root Tests for Time-Series Data with a Linear Time Trend." *Journal of Econometrics* 47:285–303.

——— and David A. Dickey. 1984. "Testing for Unit Roots in Autoregressive–Moving Average Models of Unknown Order." *Biometrika* 71:599–607.

——— and ———. 1985. "Hypothesis Testing in ARIMA($p$, 1, $q$) Models." *Journal of the American Statistical Association* 80:369–74.

Sargan, J. D., and Alok Bhargava. 1983. "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk." *Econometrica* 51:153–74.

Schmidt, Peter, and Peter C. B. Phillips. 1992. "LM Tests for a Unit Root in the Presence of Deterministic Trends." *Oxford Bulletin of Economics and Statistics* 54:257–87.

Schwert, G. William. 1989. "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business and Economic Statistics* 7:147–59.

Sims, Christopher A. 1988. "Bayesian Skepticism on Unit Root Econometrics." *Journal of Economic Dynamics and Control* 12:463–74.

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

——— and Harald Uhlig. 1991. "Understanding Unit Rooters: A Helicopter Tour." *Econometrica* 59:1591–99.

Solo, V. 1984. "The Order of Differencing in ARIMA Models." *Journal of the American Statistical Association* 79:916–21.

Sowell, Fallaw. 1990. "The Fractional Unit Root Distribution." *Econometrica* 58:495–505.

Stinchcombe, Maxwell, and Halbert White. 1993. "An Approach to Consistent Specification Testing Using Duality and Banach Limit Theory." University of California, San Diego. Mimeo.

Stock, James H. 1991. "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series." *Journal of Monetary Economics* 28:435–59.

———. 1993. "Unit Roots and Trend Breaks," in Robert Engle and Daniel McFadden, eds., *Handbook of Econometrics*, Vol. 4. Amsterdam: North-Holland.

White, J. S. 1958. "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case." *Annals of Mathematical Statistics* 29:1188–97.

# 18 Unit Roots in Multivariate Time Series

The previous chapter investigated statistical inference for univariate processes containing unit roots. This chapter develops comparable results for vector processes. The first section develops a vector version of the functional central limit theorem. Section 18.2 uses these results to generalize the analysis of Section 17.7 to vector autoregressions. Section 18.3 discusses an important problem, known as *spurious regression*, that can arise if the error term in a regression is $I(1)$. One should be concerned about the possibility of a spurious regression whenever all the variables in a regression are $I(1)$ and no lags of the dependent variable are included in the regression.

## 18.1. *Asymptotic Results for Nonstationary Vector Processes*

Section 17.2 described univariate standard Brownian motion $W(r)$ as a scalar continuous-time process ($W$: $r \in [0, 1] \rightarrow \mathbb{R}^1$). The variable $W(r)$ has a $N(0, r)$ distribution across realizations, and for any given realization, $W(r)$ is a continuous function of the date $r$ with independent increments. If a set of $n$ such independent processes, denoted $W_1(r)$, $W_2(r)$, . . . , $W_n(r)$, are collected in an $(n \times 1)$ vector $\mathbf{W}(r)$, the result is *n-dimensional standard Brownian motion.*

***Definition:*** *n-dimensional standard Brownian motion* $\mathbf{W}(\cdot)$ *is a continuous-time process associating each date* $r \in [0, 1]$ *with the* $(n \times 1)$ *vector* $\mathbf{W}(r)$ *satisfying the following:*

(*a*) $\mathbf{W}(0) = \mathbf{0}$;

(*b*) *For any dates* $0 \leq r_1 < r_2 < \cdots < r_k \leq 1$, *the changes* $[\mathbf{W}(r_2) - \mathbf{W}(r_1)]$, $[\mathbf{W}(r_3) - \mathbf{W}(r_2)], \ldots, [\mathbf{W}(r_k) - \mathbf{W}(r_{k-1})]$ *are independent multivariate Gaussian with* $[\mathbf{W}(s) - \mathbf{W}(r)] \sim N(\mathbf{0}, (s - r)\cdot\mathbf{I}_n)$;

(*c*) *For any given realization,* $\mathbf{W}(r)$ *is continuous in r with probability* 1.

Suppose that $\{v_t\}_{t=1}^{\infty}$ is a univariate i.i.d. discrete-time process with mean zero and unit variance, and let

$$\bar{X}_T^*(r) \equiv T^{-1}(v_1 + v_2 + \cdots + v_{[Tr]^*}),$$

where $[Tr]^*$ denotes the largest integer that is less than or equal to $Tr$. The functional central limit theorem states that as $T \to \infty$,

$$\sqrt{T} \cdot \bar{X}_T^*(\cdot) \xrightarrow{L} W(\cdot).$$

This readily generalizes. Suppose that $\{\mathbf{v}_t\}_{t=1}^{\infty}$ is an $n$-dimensional i.i.d. vector process with $E(\mathbf{v}_t) = \mathbf{0}$ and $E(\mathbf{v}_t \mathbf{v}_t') = \mathbf{I}_n$, and let

$$\bar{\mathbf{X}}_T^*(r) \equiv T^{-1}(\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_{[Tr]^*}).$$

Then

$$\sqrt{T} \cdot \bar{\mathbf{X}}_T^*(\cdot) \xrightarrow{L} \mathbf{W}(\cdot). \qquad [18.1.1]$$

Next, consider an i.i.d. $n$-dimensional process $\{\boldsymbol{\varepsilon}_t\}_{t=1}^{\infty}$ with mean zero and variance-covariance matrix given by $\boldsymbol{\Omega}$. Let $\mathbf{P}$ be any matrix such that

$$\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'; \qquad [18.1.2]$$

for example, $\mathbf{P}$ might be the Cholesky factor of $\boldsymbol{\Omega}$. We could think of $\boldsymbol{\varepsilon}_t$ as having been generated from

$$\boldsymbol{\varepsilon}_t = \mathbf{P}\mathbf{v}_t, \qquad [18.1.3]$$

for $\mathbf{v}_t$ i.i.d. with mean zero and variance $\mathbf{I}_n$. To see why, notice that [18.1.3] implies that $\boldsymbol{\varepsilon}_t$ is i.i.d. with mean zero and variance given by

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \mathbf{P} \cdot E(\mathbf{v}_t \mathbf{v}_t') \cdot \mathbf{P}' = \mathbf{P} \cdot \mathbf{I}_n \cdot \mathbf{P}' = \boldsymbol{\Omega}.$$

Let

$$\begin{aligned} \mathbf{X}_T^*(r) &\equiv T^{-1}(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_{[Tr]^*}) \\ &= \mathbf{P} \cdot T^{-1}(\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_{[Tr]^*}) \\ &= \mathbf{P} \cdot \bar{\mathbf{X}}_T^*(r). \end{aligned}$$

It then follows from [18.1.1] and the continuous mapping theorem that

$$\sqrt{T} \cdot \mathbf{X}_T^*(\cdot) \xrightarrow{L} \mathbf{P} \cdot \mathbf{W}(\cdot). \qquad [18.1.4]$$

For given $r$, the variable $\mathbf{P} \cdot \mathbf{W}(r)$ represents $\mathbf{P}$ times a $N(\mathbf{0}, r \cdot \mathbf{I}_n)$ vector and so has a $N(\mathbf{0}, r \cdot \mathbf{P}\mathbf{P}') = N(\mathbf{0}, r \cdot \boldsymbol{\Omega})$ distribution. The process $\mathbf{P} \cdot \mathbf{W}(r)$ is described as $n$-dimensional Brownian motion with variance matrix $\boldsymbol{\Omega}$.

The functional central limit theorem can also be applied to serially dependent vector processes using a generalization of Proposition 17.2.[1] Suppose that

$$\mathbf{u}_t = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\varepsilon}_{t-s}, \qquad [18.1.5]$$

where if $\psi_{ij}^{(s)}$ denotes the row $i$, column $j$ element of $\boldsymbol{\Psi}_s$,

$$\sum_{s=0}^{\infty} s \cdot |\psi_{ij}^{(s)}| < \infty$$

for each $i, j = 1, 2, \ldots, n$. Then algebra virtually identical to that in Proposition 17.2 can be used to show that

$$\sum_{s=1}^{t} \mathbf{u}_s = \boldsymbol{\Psi}(1) \cdot \sum_{s=1}^{t} \boldsymbol{\varepsilon}_s + \boldsymbol{\eta}_t - \boldsymbol{\eta}_0, \qquad [18.1.6]$$

where $\boldsymbol{\Psi}(1) \equiv (\boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2 + \cdots)$ and $\boldsymbol{\eta}_t = \Sigma_{s=0}^{\infty} \boldsymbol{\alpha}_s \boldsymbol{\varepsilon}_{t-s}$ for $\boldsymbol{\alpha}_s =$

[1]This is the approach used by Phillips and Solo (1992).

$-(\mathbf{\Psi}_{s+1} + \mathbf{\Psi}_{s+2} + \mathbf{\Psi}_{s+3} + \cdots)$, and $\{\boldsymbol{\alpha}_s\}_{s=0}^{\infty}$ is absolutely summable. Expression [18.1.6] provides a multivariate generalization of the Beveridge-Nelson decomposition.

If $\mathbf{u}_t$ satisfies [18.1.5] where $\boldsymbol{\varepsilon}_t$ is i.i.d. with mean zero, variance given by $\mathbf{\Omega} = \mathbf{PP}'$, and finite fourth moments, then it is straightforward to generalize to vector process the statements in Proposition 17.3 about univariate processes. For example, if we define

$$\mathbf{X}_T(r) = (1/T) \sum_{s=1}^{[Tr]^*} \mathbf{u}_s, \qquad [18.1.7]$$

then it follows from [18.1.6] that

$$\sqrt{T} \cdot \mathbf{X}_T(r) = T^{-1/2}\left(\mathbf{\Psi}(1) \sum_{s=1}^{[Tr]^*} \boldsymbol{\varepsilon}_s + \boldsymbol{\eta}_{[Tr]^*} - \boldsymbol{\eta}_0\right).$$

As in Example 17.2, one can show that

$$\sup_{\substack{r \in [0,1] \\ i=1,2,\ldots,n}} T^{-1/2} |\eta_{i,[Tr]^*} - \eta_{i,0}| \xrightarrow{p} 0.$$

It then follows from [18.1.4] that

$$\sqrt{T} \cdot \mathbf{X}_T(\cdot) \xrightarrow{p} \mathbf{\Psi}(1) \cdot \mathbf{P} \cdot \sqrt{T} \cdot \tilde{\mathbf{X}}_T^*(\cdot) \xrightarrow{L} \mathbf{\Psi}(1) \cdot \mathbf{P} \cdot \mathbf{W}(\cdot), \qquad [18.1.8]$$

where $\mathbf{\Psi}(1) \cdot \mathbf{P} \cdot \mathbf{W}(r)$ is distributed $N(\mathbf{0}, r[\mathbf{\Psi}(1)] \cdot \mathbf{\Omega} \cdot [\mathbf{\Psi}(1)]')$ across realizations. Furthermore, for $\boldsymbol{\xi}_t \equiv \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_t$, we have as in [17.3.15] that

$$T^{-3/2} \sum_{t=1}^{T} \boldsymbol{\xi}_{t-1} = \int_0^1 \sqrt{T} \cdot \mathbf{X}_T(r)\, dr \xrightarrow{L} \mathbf{\Psi}(1) \cdot \mathbf{P} \cdot \int_0^1 \mathbf{W}(r)\, dr, \qquad [18.1.9]$$

which generalizes result (f) of Proposition 17.3.

Generalizing result (e) of Proposition 17.3 requires a little more care. Consider for illustration the simplest case, where $\mathbf{v}_t$ is an i.i.d. $(n \times 1)$ vector with mean zero and $E(\mathbf{v}_t \mathbf{v}_t') = \mathbf{I}_n$. Define

$$\boldsymbol{\xi}_t^* \equiv \begin{cases} \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_t & \text{for } t = 1, 2, \ldots, T \\ \mathbf{0} & \text{for } t = 0; \end{cases}$$

we use the symbols $\mathbf{v}_t$ and $\boldsymbol{\xi}_t^*$ here in place of $\mathbf{u}_t$ and $\boldsymbol{\xi}_t$ to emphasize that $\mathbf{v}_t$ is i.i.d. with variance matrix given by $\mathbf{I}_n$. For the scalar i.i.d. unit variance case ($n = 1$, $\lambda = \gamma_0 = 1$), result (e) of Proposition 17.3 stated that

$$T^{-1} \sum_{t=1}^{T} \xi_{t-1}^* v_t \xrightarrow{L} \tfrac{1}{2}\{[W(1)]^2 - 1\}. \qquad [18.1.10]$$

The corresponding generalization for the i.i.d. unit variance vector case ($n > 1$) turns out to be

$$T^{-1} \sum_{t=1}^{T} \{\boldsymbol{\xi}_{t-1}^* \mathbf{v}_t' + \mathbf{v}_t \boldsymbol{\xi}_{t-1}^{*\prime}\} \xrightarrow{L} [\mathbf{W}(1)] \cdot [\mathbf{W}(1)]' - \mathbf{I}_n; \qquad [18.1.11]$$

see result (d) of Proposition 18.1, to follow. Expression [18.1.11] generalizes the scalar result [18.1.10] to an $(n \times n)$ matrix. The row $i$, column $i$ diagonal element of this matrix expression states that

$$T^{-1} \sum_{t=1}^{T} \{\xi_{i,t-1}^* v_{it} + v_{it} \xi_{i,t-1}^*\} \xrightarrow{L} [W_i(1)]^2 - 1, \qquad [18.1.12]$$

where $\xi_{it}^*$, $v_{it}$, and $W_i(r)$ denote the $i$th elements of the vectors $\boldsymbol{\xi}_t^*$, $\mathbf{v}_t$, and $\mathbf{W}(r)$, respectively. The row $i$, column $j$ off-diagonal element of [18.1.11] asserts that

$$T^{-1} \sum_{t=1}^{T} \{\xi^*_{i,t-1} v_{jt} + v_{it}\xi^*_{j,t-1}\} \xrightarrow{L} [W_i(1)]\cdot[W_j(1)] \qquad \text{for } i \neq j. \quad [18.1.13]$$

Thus, the sum of the random variables $T^{-1}\Sigma_{t=1}^{T}\xi^*_{i,t-1}v_{jt}$ and $T^{-1}\Sigma_{t=1}^{T}v_{it}\xi^*_{j,t-1}$ converges in distribution to the product of two independent standard Normal variables.

It is sometimes convenient to describe the asymptotic distribution of $T^{-1}\Sigma_{t=1}^{T}\xi^*_{i,t-1}v_{jt}$ alone. It turns out that

$$T^{-1} \sum_{t=1}^{T} \xi^*_{i,t-1} v_{jt} \xrightarrow{L} \int_0^1 W_i(r)\, dW_j(r). \qquad [18.1.14]$$

This expression makes use of the differential of Brownian motion, denoted $dW_j(r)$. A formal definition of the differential $dW_j(r)$ and derivation of [18.1.14] are somewhat involved—see Phillips (1988) for details. For our purposes, we will simply regard the right side of [18.1.14] as a compact notation for indicating the limiting distribution of the sequence represented by the left side. In practice, this distribution is constructed by Monte Carlo generation of the statistic on the left side of [18.1.14] for suitably large $T$.

It is evident from [18.1.13] and [18.1.14] that

$$\int_0^1 W_i(r)\, dW_j(r) + \int_0^1 W_j(r)\, dW_i(r) = W_i(1)\cdot W_j(1) \qquad \text{for } i \neq j,$$

whereas comparing [18.1.14] with [18.1.12] reveals that

$$\int_0^1 W_i(r)\, dW_i(r) = \tfrac{1}{2}\{[W_i(1)]^2 - 1\}. \qquad [18.1.15]$$

The expressions in [18.1.14] can be collected for $i, j = 1, 2, \ldots, n$ in an $(n \times n)$ matrix:

$$T^{-1} \sum_{t=1}^{T} \boldsymbol{\xi}^*_{t-1}\mathbf{v}'_t \xrightarrow{L} \int_0^1 [\mathbf{W}(r)]\, [d\mathbf{W}(r)]'. \qquad [18.1.16]$$

The following proposition summarizes the multivariate convergence results that will be used in this chapter.[2]

***Proposition 18.1:*** *Let $\mathbf{u}_t$ be an $(n \times 1)$ vector with*

$$\mathbf{u}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\varepsilon}_{t-s},$$

*where $\{s\cdot\boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable, that is, $\Sigma_{s=0}^{\infty} s\cdot|\psi_{ij}^{(s)}| < \infty$ for each $i, j = 1, 2, \ldots, n$ for $\psi_{ij}^{(s)}$ the row $i$, column $j$ element of $\boldsymbol{\Psi}_s$. Suppose that $\{\boldsymbol{\varepsilon}_t\}$ is an i.i.d. sequence with mean zero, finite fourth moments, and $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega}$ a positive definite matrix. Let $\boldsymbol{\Omega} = \mathbf{PP}'$ denote the Cholesky factorization of $\boldsymbol{\Omega}$, and define*

$$\sigma_{ij} \equiv E(\varepsilon_{it}\varepsilon_{jt}) = \text{row } i, \text{ column } j \text{ element of } \boldsymbol{\Omega}$$

$$\underset{(n\times n)}{\boldsymbol{\Gamma}_s} \equiv E(\mathbf{u}_t\mathbf{u}'_{t-s}) = \sum_{v=0}^{\infty} \boldsymbol{\Psi}_{s+v}\boldsymbol{\Omega}\boldsymbol{\Psi}'_v \qquad \text{for } s = 0, 1, 2, \ldots$$

$$\underset{(nv\times 1)}{\mathbf{z}_t} \equiv \begin{bmatrix} \mathbf{u}_{t-1} \\ \mathbf{u}_{t-2} \\ \vdots \\ \mathbf{u}_{t-v} \end{bmatrix} \qquad \text{for arbitrary } v \geq 1 \qquad [18.1.17]$$

[2]These or similar results were derived by Phillips and Durlauf (1986), Park and Phillips (1988, 1989), Sims, Stock, and Watson (1990), and Phillips and Solo (1992).

$$\underset{(n\nu\times n\nu)}{\mathbf{V}} \equiv E(\mathbf{z}_t\mathbf{z}_t') = \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{\nu-1} \\ \boldsymbol{\Gamma}_{-1} & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{\nu-2} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{\Gamma}_{-\nu+1} & \boldsymbol{\Gamma}_{-\nu+2} & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix}$$

$$\underset{(n\times n)}{\boldsymbol{\Lambda}} \equiv \boldsymbol{\Psi}(1)\cdot\mathbf{P} = (\boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2 + \cdots)\cdot\mathbf{P} \qquad [18.1.18]$$

$$\underset{(n\times 1)}{\boldsymbol{\xi}_t} \equiv \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_t \qquad \textit{for } t = 1, 2, \ldots, T \qquad [18.1.19]$$

*with* $\boldsymbol{\xi}_0 \equiv \mathbf{0}$. *Then*

(*a*) $T^{-1/2}\sum_{t=1}^{T}\mathbf{u}_t \xrightarrow{L} \boldsymbol{\Lambda}\cdot\mathbf{W}(1);$

(*b*) $T^{-1/2}\sum_{t=1}^{T}\mathbf{z}_t\varepsilon_{it} \xrightarrow{L} N(\mathbf{0}, \sigma_{ii}\cdot\mathbf{V}) \qquad \textit{for } i = 1, 2, \ldots, n;$

(*c*) $T^{-1}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_{t-s}' \xrightarrow{p} \boldsymbol{\Gamma}_s \qquad \textit{for } s = 0, 1, 2, \ldots;$

(*d*) $T^{-1}\sum_{t=1}^{T}(\boldsymbol{\xi}_{t-1}\mathbf{u}_{t-s}' + \mathbf{u}_{t-s}\boldsymbol{\xi}_{t-1}')$

$$\xrightarrow{L}\begin{cases} \boldsymbol{\Lambda}\cdot[\mathbf{W}(1)]\cdot[\mathbf{W}(1)]'\cdot\boldsymbol{\Lambda}' - \boldsymbol{\Gamma}_0 & \textit{for } s = 0 \\ \boldsymbol{\Lambda}\cdot[\mathbf{W}(1)]\cdot[\mathbf{W}(1)]'\cdot\boldsymbol{\Lambda}' + \sum_{v=-s+1}^{s-1}\boldsymbol{\Gamma}_v & \textit{for } s = 1, 2, \ldots; \end{cases}$$

(*e*) $T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{t-1}\mathbf{u}_t' \xrightarrow{L} \boldsymbol{\Lambda}\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\boldsymbol{\Lambda}' + \sum_{v=1}^{\infty}\boldsymbol{\Gamma}_v';$

(*f*) $T^{-1}\sum_{t=1}^{T}\boldsymbol{\xi}_{t-1}\boldsymbol{\varepsilon}_t' \xrightarrow{L} \boldsymbol{\Lambda}\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\mathbf{P}';$

(*g*) $T^{-3/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{t-1} \xrightarrow{L} \boldsymbol{\Lambda}\cdot\int_0^1 \mathbf{W}(r)\,dr;$

(*h*) $T^{-3/2}\sum_{t=1}^{T}t\mathbf{u}_{t-s} \xrightarrow{L} \boldsymbol{\Lambda}\cdot\left\{\mathbf{W}(1) - \int_0^1 \mathbf{W}(r)\,dr\right\} \qquad \textit{for } s = 0, 1, 2, \ldots;$

(*i*) $T^{-2}\sum_{t=1}^{T}\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}' \xrightarrow{L} \boldsymbol{\Lambda}\cdot\left\{\int_0^1 [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\boldsymbol{\Lambda}';$

(*j*) $T^{-5/2}\sum_{t=1}^{T}t\boldsymbol{\xi}_{t-1} \xrightarrow{L} \boldsymbol{\Lambda}\cdot\int_0^1 r\mathbf{W}(r)\,dr;$

(*k*) $T^{-3}\sum_{t=1}^{T}t\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}' \xrightarrow{L} \boldsymbol{\Lambda}\cdot\left\{\int_0^1 r[\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\boldsymbol{\Lambda}';$

(*l*) $T^{-(v+1)}\sum_{t=1}^{T}t^v \rightarrow 1/(v + 1) \qquad \textit{for } v = 0, 1, 2, \ldots.$

## 18.2. *Vector Autoregressions Containing Unit Roots*

Suppose that a vector $\mathbf{y}_t$ could be described by a vector autoregression in the differences $\Delta\mathbf{y}_t$. This section presents results developed by Park and Phillips (1988, 1989) and Sims, Stock, and Watson (1990) for the consequences of estimating the *VAR* in levels. We begin by generalizing the Dickey-Fuller variable transformation that was used in analyzing a univariate autoregression.

### *An Alternative Representation of a* VAR(p) *Process*

Let $\mathbf{y}_t$ be an $(n \times 1)$ vector satisfying

$$(\mathbf{I}_n - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \cdots - \mathbf{\Phi}_p L^p)\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t, \qquad [18.2.1]$$

where $\mathbf{\Phi}_s$ denotes an $(n \times n)$ matrix for $s = 1, 2, \ldots, p$ and $\boldsymbol{\alpha}$ and $\boldsymbol{\varepsilon}_t$ are $(n \times 1)$ vectors. The scalar algebra in [17.7.4] works perfectly well for matrices, establishing that for any values of $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \ldots, \mathbf{\Phi}_p$, the following polynomials are equivalent:

$$\begin{aligned}(\mathbf{I}_n &- \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \cdots - \mathbf{\Phi}_p L^p) \\ &= (\mathbf{I}_n - \boldsymbol{\rho} L) - (\boldsymbol{\zeta}_1 L + \boldsymbol{\zeta}_2 L^2 + \cdots + \boldsymbol{\zeta}_{p-1} L^{p-1})(1 - L),\end{aligned} \qquad [18.2.2]$$

where

$$\boldsymbol{\rho} \equiv \mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \cdots + \mathbf{\Phi}_p \qquad [18.2.3]$$

$$\boldsymbol{\zeta}_s \equiv -[\mathbf{\Phi}_{s+1} + \mathbf{\Phi}_{s+2} + \cdots + \mathbf{\Phi}_p] \quad \text{for } s = 1, 2, \ldots, p - 1. \qquad [18.2.4]$$

It follows that any *VAR*$(p)$ process [18.2.1] can always be written in the form

$$(\mathbf{I}_n - \boldsymbol{\rho} L)\mathbf{y}_t - (\boldsymbol{\zeta}_1 L + \boldsymbol{\zeta}_2 L^2 + \cdots + \boldsymbol{\zeta}_{p-1} L^{p-1})(1 - L)\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t$$

or

$$\mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\rho}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \qquad [18.2.5]$$

The null hypothesis considered throughout this section is that the first difference of $\mathbf{y}$ follows a *VAR*$(p - 1)$ process;

$$\Delta\mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t, \qquad [18.2.6]$$

requiring from [18.2.5] that

$$\boldsymbol{\rho} = \mathbf{I}_n \qquad [18.2.7]$$

or, from [18.2.3],

$$\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \cdots + \mathbf{\Phi}_p = \mathbf{I}_n. \qquad [18.2.8]$$

Recalling Proposition 10.1, the vector autoregression [18.2.1] will be said to contain at least one unit root if the following determinant is zero:

$$|\mathbf{I}_n - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \cdots - \mathbf{\Phi}_p| = 0. \qquad [18.2.9]$$

Note that [18.2.8] implies [18.2.9] but [18.2.9] does not imply [18.2.8]. Thus, this section is considering only a subset of the class of vector autoregressions containing a unit root, namely, the class described by [18.2.8]. Vector autoregressions for which [18.2.9] holds but [18.2.8] does not will be considered in Chapter 19.

This section begins with a vector generalization of case 2 from Chapter 17.

## A Vector Autoregression with No Drift in Any of the Variables

Here we assume that the $VAR$ [18.2.1] satisfies [18.2.8] along with $\boldsymbol{\alpha} = \mathbf{0}$ and consider the consequences of estimating each equation in levels by $OLS$ using observations $t = 1, 2, \ldots, T$ and conditioning on $\mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}$. A constant term is assumed to be included in each regression. Under the maintained hypothesis [18.2.8], the data-generating process can be described as

$$(\mathbf{I}_n - \boldsymbol{\zeta}_1 L - \boldsymbol{\zeta}_2 L^2 - \cdots - \boldsymbol{\zeta}_{p-1} L^{p-1})\Delta \mathbf{y}_t = \boldsymbol{\varepsilon}_t. \qquad [18.2.10]$$

Assuming that all values of $z$ satisfying

$$|\mathbf{I}_n - \boldsymbol{\zeta}_1 z - \boldsymbol{\zeta}_2 z^2 - \cdots - \boldsymbol{\zeta}_{p-1} z^{p-1}| = 0$$

lie outside the unit circle, [18.2.10] implies that

$$\Delta \mathbf{y}_t = \mathbf{u}_t, \qquad [18.2.11]$$

where

$$\mathbf{u}_t = (\mathbf{I}_n - \boldsymbol{\zeta}_1 L - \boldsymbol{\zeta}_2 L^2 - \cdots - \boldsymbol{\zeta}_{p-1} L^{p-1})^{-1}\boldsymbol{\varepsilon}_t.$$

If $\boldsymbol{\varepsilon}_t$ is i.i.d. with mean zero, positive definite variance-covariance matrix $\boldsymbol{\Omega} = \mathbf{PP}'$, and finite fourth moments, then $\mathbf{u}_t$ satisfies the conditions of Proposition 18.1 with

$$\boldsymbol{\Psi}(L) = (\mathbf{I}_n - \boldsymbol{\zeta}_1 L - \boldsymbol{\zeta}_2 L^2 - \cdots - \boldsymbol{\zeta}_{p-1} L^{p-1})^{-1}. \qquad [18.2.12]$$

Also from [18.2.11], we have

$$\mathbf{y}_t = \mathbf{y}_0 + \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_t,$$

so that $\mathbf{y}_t$ will have the same asymptotic behavior as $\boldsymbol{\xi}_t$ in Proposition 18.1.

Recall that the fitted values of a $VAR$ estimated in levels [18.2.1] are identical to the fitted values for a $VAR$ estimated in the form of [18.2.5]. Consider the $i$th equation in [18.2.5], which we write as

$$y_{it} = \boldsymbol{\zeta}_{i1}'\mathbf{u}_{t-1} + \boldsymbol{\zeta}_{i2}'\mathbf{u}_{t-2} + \cdots + \boldsymbol{\zeta}_{i,p-1}'\mathbf{u}_{t-p+1} + \alpha_i + \boldsymbol{\rho}_i'\mathbf{y}_{t-1} + \varepsilon_{it}, \qquad [18.2.13]$$

where $\mathbf{u}_t = \Delta\mathbf{y}_t$ and $\boldsymbol{\zeta}_{is}'$ denotes the $i$th row of $\boldsymbol{\zeta}_s$ for $s = 1, 2, \ldots, p-1$. Similarly, $\boldsymbol{\rho}_i'$ denotes the $i$th row of $\boldsymbol{\rho}$. Under the null hypothesis [18.2.7], $\boldsymbol{\rho}_i' = \mathbf{e}_i'$, where $\mathbf{e}_i'$ is the $i$th row of the $(n \times n)$ identity matrix. Recall the usual expression [8.2.3] for the deviation of the $OLS$ estimate $\mathbf{b}_T$ from its hypothesized true value:

$$\mathbf{b}_T - \boldsymbol{\beta} = (\Sigma \mathbf{x}_t\mathbf{x}_t')^{-1}(\Sigma \mathbf{x}_t\varepsilon_t), \qquad [18.2.14]$$

where $\Sigma$ denotes summation over $t = 1$ through $T$. In the case of $OLS$ estimation of [18.2.13],

$$\mathbf{b}_T - \boldsymbol{\beta} = \begin{bmatrix} \hat{\boldsymbol{\zeta}}_{i1} - \boldsymbol{\zeta}_{i1} \\ \hat{\boldsymbol{\zeta}}_{i2} - \boldsymbol{\zeta}_{i2} \\ \vdots \\ \hat{\boldsymbol{\zeta}}_{i,p-1} - \boldsymbol{\zeta}_{i,p-1} \\ \hat{\alpha}_i \\ \hat{\boldsymbol{\rho}}_i - \mathbf{e}_i \end{bmatrix} \qquad [18.2.15]$$

$$\Sigma \mathbf{x}_t \mathbf{x}_t'$$

$$= \begin{bmatrix} \Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{u}_{t-1} & \Sigma \mathbf{u}_{t-1}\mathbf{y}_{t-1}' \\ \Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{u}_{t-2} & \Sigma \mathbf{u}_{t-2}\mathbf{y}_{t-1}' \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{u}_{t-p+1} & \Sigma \mathbf{u}_{t-p+1}\mathbf{y}_{t-1}' \\ \Sigma \mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-p+1}' & T & \Sigma \mathbf{y}_{t-1}' \\ \Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-1}' & \Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{y}_{t-1} & \Sigma \mathbf{y}_{t-1}\mathbf{y}_{t-1}' \end{bmatrix}$$

[18.2.16]

$$\Sigma \mathbf{x}_t \varepsilon_t = \begin{bmatrix} \Sigma \mathbf{u}_{t-1}\varepsilon_{it} \\ \Sigma \mathbf{u}_{t-2}\varepsilon_{it} \\ \vdots \\ \Sigma \mathbf{u}_{t-p+1}\varepsilon_{it} \\ \Sigma \varepsilon_{it} \\ \Sigma \mathbf{y}_{t-1}\varepsilon_{it} \end{bmatrix}. \qquad [18.2.17]$$

Our earlier convention would append a subscript $T$ to the estimated coefficients $\hat{\zeta}_{is}$ in [18.2.15]. For this discussion, the subscript $T$ will be suppressed to avoid excessively cumbersome notation.

Define $\mathbf{Y}_T$ to be the following matrix:

$$\underset{(np+1)\times(np+1)}{\mathbf{Y}_T} \equiv \begin{bmatrix} T^{1/2}\cdot \mathbf{I}_{n(p-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & T^{1/2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & T\cdot \mathbf{I}_n \end{bmatrix}. \qquad [18.2.18]$$

Premultiplying [18.2.14] by $\mathbf{Y}_T$ and rearranging as in [17.4.20] results in

$$\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}) = (\mathbf{Y}_T^{-1}\Sigma \mathbf{x}_t\mathbf{x}_t'\mathbf{Y}_T^{-1})^{-1}(\mathbf{Y}_T^{-1}\Sigma \mathbf{x}_t\varepsilon_t). \qquad [18.2.19]$$

Using results (a), (c), (d), (g), and (i) of Proposition 18.1, we find

$$(\mathbf{Y}_T^{-1}\Sigma \mathbf{x}_t\mathbf{x}_t'\mathbf{Y}_T^{-1}) = \begin{bmatrix} T^{-1}\Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-1}' & T^{-1}\Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-2}' & \cdots & T^{-1}\Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-p+1}' & T^{-1}\Sigma \mathbf{u}_{t-1} & T^{-3/2}\Sigma \mathbf{u}_{t-1}\mathbf{y}_{t-1}' \\ T^{-1}\Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-1}' & T^{-1}\Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-2}' & \cdots & T^{-1}\Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-p+1}' & T^{-1}\Sigma \mathbf{u}_{t-2} & T^{-3/2}\Sigma \mathbf{u}_{t-2}\mathbf{y}_{t-1}' \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ T^{-1}\Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-1}' & T^{-1}\Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-2}' & \cdots & T^{-1}\Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-p+1}' & T^{-1}\Sigma \mathbf{u}_{t-p+1} & T^{-3/2}\Sigma \mathbf{u}_{t-p+1}\mathbf{y}_{t-1}' \\ T^{-1}\Sigma \mathbf{u}_{t-1}' & T^{-1}\Sigma \mathbf{u}_{t-2}' & \cdots & T^{-1}\Sigma \mathbf{u}_{t-p+1}' & 1 & T^{-3/2}\Sigma \mathbf{y}_{t-1}' \\ T^{-3/2}\Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-1}' & T^{-3/2}\Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-2}' & \cdots & T^{-3/2}\Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-p+1}' & T^{-3/2}\Sigma \mathbf{y}_{t-1} & T^{-2}\Sigma \mathbf{y}_{t-1}\mathbf{y}_{t-1}' \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}, \qquad [18.2.20]$$

where

$$\underset{[n(p-1)\times n(p-1)]}{\mathbf{V}} \equiv \begin{bmatrix} \mathbf{\Gamma}_0 & \mathbf{\Gamma}_1 & \cdots & \mathbf{\Gamma}_{p-2} \\ \mathbf{\Gamma}_{-1} & \mathbf{\Gamma}_0 & \cdots & \mathbf{\Gamma}_{p-3} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{\Gamma}_{-p+2} & \mathbf{\Gamma}_{-p+3} & \cdots & \mathbf{\Gamma}_0 \end{bmatrix} \qquad [18.2.21]$$

$$\mathbf{\Gamma}_s \equiv E(\Delta\mathbf{y}_t)(\Delta\mathbf{y}_{t-s})'$$

$$\underset{(n+1)\times(n+1)}{\mathbf{Q}} \equiv \begin{bmatrix} 1 & \left[\int \mathbf{W}(r)\,dr\right]'\cdot\mathbf{\Lambda}' \\ \mathbf{\Lambda}\cdot\int \mathbf{W}(r)\,dr & \mathbf{\Lambda}\cdot\left\{\int[\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\mathbf{\Lambda}' \end{bmatrix}. \qquad [18.2.22]$$

Also, the integral sign denotes integration over $r$ from 0 to 1, and

$$\mathbf{\Lambda} \equiv (\mathbf{I}_n - \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2 - \cdots - \boldsymbol{\zeta}_{p-1})^{-1}\mathbf{P} \qquad [18.2.23]$$

with $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{PP}'$. Similarly, applying results (a), (b), and (f) from Proposition 18.1 to the second term in [18.2.19] reveals

$$(\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t\varepsilon_t) = \begin{bmatrix} T^{-1/2}\Sigma\mathbf{u}_{t-1}\varepsilon_{it} \\ T^{-1/2}\Sigma\mathbf{u}_{t-2}\varepsilon_{it} \\ \vdots \\ T^{-1/2}\Sigma\mathbf{u}_{t-p+1}\varepsilon_{it} \\ T^{-1/2}\Sigma\varepsilon_{it} \\ T^{-1}\Sigma\mathbf{y}_{t-1}\varepsilon_{it} \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix}, \qquad [18.2.24]$$

where

$$\underset{[n(p-1)\times 1]}{\mathbf{h}_1} \sim N(\mathbf{0}, \sigma_{ii}\mathbf{V})$$

$$\sigma_{ii} = E(\varepsilon_{it}^2)$$

$$\underset{[(n+1)\times 1]}{\mathbf{h}_2} = \begin{bmatrix} \mathbf{e}_i'\mathbf{PW}(1) \\ \mathbf{\Lambda}\cdot\left\{\int[\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\mathbf{P}'\mathbf{e}_i \end{bmatrix}$$

for $\mathbf{e}_i$ the $i$th column of $\mathbf{I}_n$. Results [18.2.19], [18.2.20], and [18.2.24] establish that

$$\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{L} \begin{bmatrix} \mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2 \end{bmatrix}. \qquad [18.2.25]$$

The first $n(p-1)$ elements of [18.2.25] imply that the coefficients on $\Delta\mathbf{y}_{t-1}$, $\Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$ converge at rate $\sqrt{T}$ to Gaussian variables:

$$\sqrt{T}\begin{bmatrix} \hat{\boldsymbol{\zeta}}_{i1} - \boldsymbol{\zeta}_{i1} \\ \hat{\boldsymbol{\zeta}}_{i2} - \boldsymbol{\zeta}_{i2} \\ \vdots \\ \hat{\boldsymbol{\zeta}}_{i,p-1} - \boldsymbol{\zeta}_{i,p-1} \end{bmatrix} \xrightarrow{L} \mathbf{V}^{-1}\mathbf{h}_1 \sim N(\mathbf{0}, \sigma_{ii}\cdot\mathbf{V}^{-1}). \qquad [18.2.26]$$

This means that the Wald form of the *OLS* $\chi^2$ test of any linear hypothesis that involves only the coefficients on $\Delta\mathbf{y}_{t-s}$ has the usual asymptotic $\chi^2$ distribution, as the reader is invited to confirm in Exercise 18.1.

Notice that [18.2.26] is identical to the asymptotic distribution that would characterize the estimates if the *VAR* were estimated in differences:

$$\Delta y_{it} = \alpha_i + \boldsymbol{\zeta}_{i1}' \Delta \mathbf{y}_{t-1} + \boldsymbol{\zeta}_{i2}' \Delta \mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{i,p-1}' \Delta \mathbf{y}_{t-p+1} + \varepsilon_{it}. \quad [18.2.27]$$

Thus, as in the case of a univariate autoregression, if the goal is to estimate the parameters $\boldsymbol{\zeta}_{i1}, \boldsymbol{\zeta}_{i2}, \ldots, \boldsymbol{\zeta}_{i,p-1}$ or test hypotheses about these coefficients, there is no need based on the asymptotic distributions for estimating the *VAR* in the difference form [18.2.27] rather than in the levels form,

$$\begin{aligned} y_{it} = {} & \boldsymbol{\zeta}_{i1}'\Delta \mathbf{y}_{t-1} + \boldsymbol{\zeta}_{i2}'\Delta \mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{i,p-1}'\Delta \mathbf{y}_{t-p+1} \\ & + \alpha_i + \boldsymbol{\rho}_i'\mathbf{y}_{t-1} + \varepsilon_{it}. \end{aligned} \quad [18.2.28]$$

Nevertheless, the small-sample distributions may well be improved by estimating the *VAR* in differences, assuming that the restriction [18.2.8] is valid.

Although the asymptotic distribution of the coefficient on $\mathbf{y}_{t-1}$ is non-Gaussian, the fact that this estimate converges at rate $T$ means that a hypothesis test involving a single linear combination of $\boldsymbol{\rho}_i$ and $\boldsymbol{\zeta}_{i1}, \boldsymbol{\zeta}_{i2}, \ldots, \boldsymbol{\zeta}_{i,p-1}$ will be dominated asymptotically by the coefficients with the slower rate of convergence, namely, $\boldsymbol{\zeta}_{i1}, \boldsymbol{\zeta}_{i2}, \ldots, \boldsymbol{\zeta}_{i,p-1}$, and indeed will have the same asymptotic distribution as if the true value of $\boldsymbol{\rho} = \mathbf{I}_n$ were used. For example, if the *VAR* is estimated in levels form [18.2.1], the individual coefficient matrices $\boldsymbol{\Phi}_s$ are related to the coefficients for the transformed *VAR* [18.2.5] by

$$\hat{\boldsymbol{\Phi}}_p = -\hat{\boldsymbol{\zeta}}_{p-1} \quad [18.2.29]$$

$$\hat{\boldsymbol{\Phi}}_s = \hat{\boldsymbol{\zeta}}_s - \hat{\boldsymbol{\zeta}}_{s-1} \qquad \text{for } s = 2, 3, \ldots, p-1 \quad [18.2.30]$$

$$\hat{\boldsymbol{\Phi}}_1 = \hat{\boldsymbol{\rho}} + \hat{\boldsymbol{\zeta}}_1. \quad [18.2.31]$$

Since $\sqrt{T}(\hat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}_s)$ is asymptotically Gaussian and since $\hat{\boldsymbol{\rho}}$ is $O_p(T^{-1})$, it follows that $\sqrt{T}(\hat{\boldsymbol{\Phi}}_s - \boldsymbol{\Phi}_s)$ is asymptotically Gaussian for $s = 1, 2, \ldots, p$ assuming that $p \geq 2$. This means that if the *VAR* is estimated in levels in the standard way, any individual autoregressive coefficient converges at rate $\sqrt{T}$ to a Gaussian variable and the usual $t$ test of a hypothesis involving that coefficient is asymptotically valid. Moreover, an $F$ test involving a linear combination other than $\boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 + \cdots + \boldsymbol{\Phi}_p$ has the usual asymptotic distribution.

Another important example is testing the null hypothesis that the data follow a $VAR(p_0)$ with $p_0 \geq 1$ against the alternative of a $VAR(p)$ with $p > p_0$. Consider *OLS* estimation of the $i$th equation of the *VAR* as represented in levels,

$$y_{it} = \alpha_i + \boldsymbol{\Phi}_{i1}'\mathbf{y}_{t-1} + \boldsymbol{\Phi}_{i2}'\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_{ip}'\mathbf{y}_{t-p} + \varepsilon_{it}, \quad [18.2.32]$$

where $\boldsymbol{\Phi}_{is}'$ denotes the $i$th row of $\boldsymbol{\Phi}_s$. Consider the null hypothesis

$$H_0\text{: } \boldsymbol{\Phi}_{i,p_0+1} = \boldsymbol{\Phi}_{i,p_0+2} = \cdots = \boldsymbol{\Phi}_{ip} = \mathbf{0}. \quad [18.2.33]$$

The Wald form of the *OLS* $\chi^2$ test of this hypothesis will be numerically identical to the test of

$$H_0\text{: } \boldsymbol{\zeta}_{i,p_0} = \boldsymbol{\zeta}_{i,p_0+1} = \cdots = \boldsymbol{\zeta}_{i,p-1} = \mathbf{0} \quad [18.2.34]$$

for *OLS* estimation of

$$\begin{aligned} y_{it} = {} & \boldsymbol{\zeta}_{i1}'\Delta \mathbf{y}_{t-1} + \boldsymbol{\zeta}_{i2}'\Delta \mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{i,p-1}'\Delta \mathbf{y}_{t-p+1} \\ & + \alpha_i + \boldsymbol{\rho}_i'\mathbf{y}_{t-1} + \varepsilon_{it}. \end{aligned} \quad [18.2.35]$$

Since we have seen that the usual $F$ test of [18.2.34] is asymptotically valid and since a test of [18.2.33] is based on the identical test statistic, it follows that the usual Wald test for assessing the number of lags to include in the regression is perfectly appropriate when the regression is estimated in levels form as in [18.2.32].

Of course, some hypothesis tests based on a $VAR$ estimated in levels will not have the usual asymptotic distribution. An important example is a Granger-causality test of the null hypothesis that some of the variables in $\mathbf{y}_t$ do not appear in the regression explaining $y_{it}$. Partition $\mathbf{y}_t = (\mathbf{y}_{1t}', \mathbf{y}_{2t}')'$, where $\mathbf{y}_{2t}$ denotes the subset of variables that do not affect $y_{it}$ under the null hypothesis. Write the regression in levels as

$$\begin{aligned} y_{it} = {} & \boldsymbol{\omega}_1'\mathbf{y}_{1,t-1} + \boldsymbol{\lambda}_1'\mathbf{y}_{2,t-1} + \boldsymbol{\omega}_2'\mathbf{y}_{1,t-2} + \boldsymbol{\lambda}_2'\mathbf{y}_{2,t-2} + \cdots \\ & + \boldsymbol{\omega}_p'\mathbf{y}_{1,t-p} + \boldsymbol{\lambda}_p'\mathbf{y}_{2,t-p} + \alpha_i + \varepsilon_{it} \end{aligned} \qquad [18.2.36]$$

and the transformed regression as

$$\begin{aligned} y_{it} = {} & \boldsymbol{\beta}_1'\Delta\mathbf{y}_{1,t-1} + \boldsymbol{\gamma}_1'\Delta\mathbf{y}_{2,t-1} + \boldsymbol{\beta}_2'\Delta\mathbf{y}_{1,t-2} + \boldsymbol{\gamma}_2'\Delta\mathbf{y}_{2,t-2} + \cdots \\ & + \boldsymbol{\beta}_{p-1}'\Delta\mathbf{y}_{1,t-p+1} + \boldsymbol{\gamma}_{p-1}'\Delta\mathbf{y}_{2,t-p+1} + \alpha_i + \boldsymbol{\eta}'\mathbf{y}_{1,t-1} \\ & + \boldsymbol{\delta}'\mathbf{y}_{2,t-1} + \varepsilon_{it}. \end{aligned} \qquad [18.2.37]$$

The $F$ test of the null hypothesis $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = \cdots = \boldsymbol{\lambda}_p = \mathbf{0}$ based on $OLS$ estimation of [18.2.36] is numerically identical to the $F$ test of the null hypothesis $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \cdots = \boldsymbol{\gamma}_{p-1} = \boldsymbol{\delta} = \mathbf{0}$ based on $OLS$ estimation of [18.2.37]. Since $\hat{\boldsymbol{\delta}}$ has a nonstandard limiting distribution, a test for Granger-causality based on a $VAR$ estimated in levels typically does not have the usual limiting $\chi^2$ distribution (see Exercise 18.2 and Toda and Phillips, 1993b, for further discussion). Monte Carlo simulations by Ohanian (1988), for example, found that if an independent random walk is added to a vector autoregression, the random walk might spuriously appear to Granger-cause the other variables in 20% of the samples if the 5% critical value for a $\chi^2$ variable is mistakenly used to interpret the test statistic. Toda and Phillips (1993a) have an analytical treatment of this issue.

### *A Vector Autoregression with Drift in Some of the Variables*

Here we again consider estimation of a $VAR$ written in the form

$$\mathbf{y}_t = \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\rho}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \qquad [18.2.38]$$

As before, it is assumed that roots of

$$|\mathbf{I}_n - \boldsymbol{\zeta}_1 z - \boldsymbol{\zeta}_2 z^2 - \cdots - \boldsymbol{\zeta}_{p-1}z^{p-1}| = 0$$

are outside the unit circle, that $\boldsymbol{\varepsilon}_t$ is i.i.d. with mean zero, positive definite variance $\boldsymbol{\Omega}$, and finite fourth moments, and that the true value of $\boldsymbol{\rho}$ is the $(n \times n)$ identity matrix. These assumptions imply that

$$\Delta\mathbf{y}_t = \boldsymbol{\delta} + \mathbf{u}_t \qquad [18.2.39]$$

where

$$\boldsymbol{\delta} \equiv (\mathbf{I}_n - \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2 - \cdots - \boldsymbol{\zeta}_{p-1})^{-1}\boldsymbol{\alpha} \qquad [18.2.40]$$

$$\mathbf{u}_t \equiv \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t \qquad [18.2.41]$$

$$\boldsymbol{\Psi}(L) \equiv (\mathbf{I}_n - \boldsymbol{\zeta}_1 L - \boldsymbol{\zeta}_2 L^2 - \cdots - \boldsymbol{\zeta}_{p-1}L^{p-1})^{-1}.$$

In contrast to the previous case, in which it was assumed that $\boldsymbol{\delta} = \mathbf{0}$, here we suppose that at least one and possibly all of the elements of $\boldsymbol{\delta}$ are nonzero.

Since this is a vector generalization of case 3 for the univariate autoregression considered in Chapter 17, one's first thought might be that, because of the nonzero drift in the $I(1)$ regressors, if all of the elements of $\boldsymbol{\delta}$ are nonzero, then all the coefficients will have the usual Gaussian limiting distribution. However, this turns out not to be the case. Any individual element $y_{jt}$ of the vector $\mathbf{y}_t$ is dominated by a deterministic time trend, and if $y_{jt}$ appeared alone in the regression, the asymptotic results would be the same as if $y_{jt}$ were replaced by the time trend $t$. Indeed, as noted by West (1988), in a regression in which there is a single $I(1)$ regressor with nonzero drift and in which all other regressors are $I(0)$, all of the coefficients would be asymptotically Gaussian and $F$ tests would have their usual limiting distribution. This can be shown using essentially the same algebra as in the univariate autoregression analyzed in case 3 in Chapter 17. However, as noted by Sims, Stock, and Watson (1990), in [18.2.38] there are $n$ different $I(1)$ regressors (the $n$ elements of $\mathbf{y}_{t-1}$), and if each of these were replaced by $\delta_i(t-1)$, the resulting regressors would be perfectly collinear. *OLS* will fit $n$ separate linear combinations of $\mathbf{y}_t$ so as to try to minimize the sum of squared residuals, and while one of these will indeed pick up the deterministic time trend $t$, the other linear combinations correspond to $I(1)$ driftless variables.

To develop the correct asymptotic distribution, it is convenient to work with a transformation of [18.2.38] that isolates these different linear combinations. Note that the difference equation [18.2.39] implies that

$$\mathbf{y}_t = \mathbf{y}_0 + \boldsymbol{\delta}\cdot t + \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_t. \qquad [18.2.42]$$

Suppose for illustration that the $n$th variable in the system exhibits nonzero drift ($\delta_n \neq 0$); whether in addition $\delta_i \neq 0$ for $i = 1, 2, \ldots, n-1$ then turns out to be irrelevant, assuming that [18.2.8] holds. Define

$$\begin{aligned}
y_{1t}^* &\equiv y_{1t} - (\delta_1/\delta_n)y_{nt} \\
y_{2t}^* &\equiv y_{2t} - (\delta_2/\delta_n)y_{nt} \\
&\vdots \\
y_{n-1,t}^* &\equiv y_{n-1,t} - (\delta_{n-1}/\delta_n)y_{nt} \\
y_{nt}^* &\equiv y_{nt}.
\end{aligned}$$

Thus, for $i = 1, 2, \ldots, n-1$,

$$\begin{aligned}
y_{it}^* &= [y_{i0} + \delta_i t + u_{i1} + u_{i2} + \cdots + u_{it}] \\
&\quad - (\delta_i/\delta_n)[y_{n0} + \delta_n t + u_{n1} + u_{n2} + \cdots + u_{nt}] \\
&\equiv y_{i0}^* + \xi_{it}^*,
\end{aligned}$$

where we have defined

$$\begin{aligned}
y_{i0}^* &\equiv [y_{i0} - (\delta_i/\delta_n)y_{n0}] \\
\xi_{it}^* &\equiv u_{i1}^* + u_{i2}^* + \cdots + u_{it}^* \\
u_{it}^* &\equiv u_{it} - (\delta_i/\delta_n)u_{nt}.
\end{aligned}$$

Collecting $u_{1t}^*, u_{2t}^*, \ldots, u_{n-1,t}^*$ in an $[(n-1) \times 1]$ vector $\mathbf{u}_t^*$, it follows from [18.2.41] that

$$\mathbf{u}_t^* = \boldsymbol{\Psi}^*(L)\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\Psi}^*(L)$ denotes the following $[(n-1) \times n]$ matrix polynomial:

$$\boldsymbol{\Psi}^*(L) \equiv \mathbf{H}\cdot\boldsymbol{\Psi}(L)$$

for

$$\underset{[(n-1)\times n]}{\mathbf{H}} \equiv \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & -(\delta_1/\delta_n) \\ 0 & 1 & 0 & \cdots & 0 & -(\delta_2/\delta_n) \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -(\delta_{n-1}/\delta_n) \end{bmatrix}.$$

Since $\{s\cdot\Psi_s\}_{s=0}^{\infty}$ is absolutely summable, so is $\{s\cdot\Psi_s^*\}_{s=0}^{\infty}$. Hence, the $[(n-1)\times 1]$ vector $\mathbf{y}_t^* = (y_{1t}^*, y_{2t}^*, \ldots, y_{n-1,t}^*)'$ has the same asymptotic properties as the vector $\boldsymbol{\xi}_t$ in Proposition 18.1 with the matrix $\Psi(1)$ in Proposition 18.1 replaced by $\Psi^*(1)$.

If we had direct observations on $\mathbf{y}_t^*$ and $\mathbf{u}_t$, the fitted values of the *VAR* as estimated from [18.2.38] would clearly be identical to those from estimation of

$$\begin{aligned}\mathbf{y}_t = \boldsymbol{\zeta}_1\mathbf{u}_{t-1} + \boldsymbol{\zeta}_2\mathbf{u}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\mathbf{u}_{t-p+1} + \boldsymbol{\alpha}^* \\ + \boldsymbol{\rho}^*\mathbf{y}_{t-1}^* + \boldsymbol{\gamma}\cdot y_{n,t-1} + \boldsymbol{\varepsilon}_t,\end{aligned} \qquad [18.2.43]$$

where $\boldsymbol{\rho}^*$ denotes an $[n\times(n-1)]$ matrix of coefficients while $\boldsymbol{\gamma}$ is an $(n\times 1)$ vector of coefficients. This representation separates the zero-mean stationary regressors ($\mathbf{u}_{t-s} = \Delta\mathbf{y}_{t-s} - \boldsymbol{\delta}$), the constant term ($\boldsymbol{\alpha}^*$), the driftless $I(0)$ regressors ($\mathbf{y}_{t-1}^*$), and a term dominated asymptotically by a time trend ($y_{n,t-1}$). As in Section 16.3, once the hypothetical *VAR* [18.2.43] is analyzed, we can infer the properties of the *VAR* as actually estimated ([18.2.38] or [18.2.1]) from the relation between the fitted values for the different representations.

Consider the $i$th equation in [18.2.43],

$$\begin{aligned}y_{it} = \boldsymbol{\zeta}_{i1}'\mathbf{u}_{t-1} + \boldsymbol{\zeta}_{i2}'\mathbf{u}_{t-2} + \cdots + \boldsymbol{\zeta}_{i,p-1}'\mathbf{u}_{t-p+1} + \alpha_i^* \\ + \boldsymbol{\rho}_i^{*\prime}\mathbf{y}_{t-1}^* + \gamma_i y_{n,t-1} + \varepsilon_{it},\end{aligned} \qquad [18.2.44]$$

where $\boldsymbol{\zeta}_{is}'$ denotes the $i$th row of $\boldsymbol{\zeta}_s$ and $\boldsymbol{\rho}_i^{*\prime}$ is the $i$th row of $\boldsymbol{\rho}^*$. Define

$$\begin{aligned}
\underset{[(np+1)\times 1]}{\mathbf{x}_t^*} &\equiv (\mathbf{u}_{t-1}', \mathbf{u}_{t-2}', \ldots, \mathbf{u}_{t-p+1}', 1, \mathbf{y}_{t-1}^{*\prime}, y_{n,t-1})' \\
\underset{[(np+1)\times(np+1)]}{\mathbf{Y}_T} &\equiv \begin{bmatrix} T^{1/2}\cdot\mathbf{I}_{n(p-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & T^{1/2} & \mathbf{0}' & 0 \\ \mathbf{0} & \mathbf{0} & T\cdot\mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{0}' & 0 & \mathbf{0}' & T^{3/2}\end{bmatrix} \\
\underset{[(n-1)\times n]}{\boldsymbol{\Lambda}^*} &\equiv \Psi^*(1)\cdot\mathbf{P},
\end{aligned} \qquad [18.2.45]$$

where $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{PP}'$. Then, from Proposition 18.1,

$$\left(\mathbf{Y}_T^{-1}\sum_{t=1}^{T}(\mathbf{x}_t^*)(\mathbf{x}_t^{*\prime})\mathbf{Y}_T^{-1}\right) \qquad [18.2.46]$$

$$\xrightarrow{L}\begin{bmatrix}
\mathbf{V} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0}' & 1 & \left[\int \mathbf{W}(r)\,dr\right]'\cdot\boldsymbol{\Lambda}^{*\prime} & \delta_n/2 \\
\mathbf{0} & \boldsymbol{\Lambda}^*\cdot\int \mathbf{W}(r)\,dr & \boldsymbol{\Lambda}^*\cdot\left\{\int [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\boldsymbol{\Lambda}^{*\prime} & \delta_n\cdot\boldsymbol{\Lambda}^*\cdot\int r\mathbf{W}(r)\,dr \\
\mathbf{0}' & \delta_n/2 & \delta_n\cdot\left[\int r\mathbf{W}(r)\,dr\right]'\cdot\boldsymbol{\Lambda}^{*\prime} & \delta_n^2/3
\end{bmatrix},$$

where

$$\underset{[n(p-1)\times n(p-1)]}{\mathbf{V}} \equiv \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{p-2} \\ \boldsymbol{\Gamma}_{-1} & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{p-3} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{\Gamma}_{-p+2} & \boldsymbol{\Gamma}_{-p+3} & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix} \qquad [18.2.47]$$

and $\mathbf{W}(r)$ denotes $n$-dimensional standard Brownian motion while the integral sign indicates integration over $r$ from 0 to 1. Similarly,

$$\mathbf{Y}_T^{-1} \sum_{t=1}^{T} \mathbf{x}_t^* \varepsilon_{it} \xrightarrow{L} \begin{bmatrix} \mathbf{h}_1 \\ h_2 \\ \mathbf{h}_3 \\ h_4 \end{bmatrix}, \qquad [18.2.48]$$

where $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma_{ii}\mathbf{V})$. The variables $h_2$ and $h_4$ are also Gaussian, though $\mathbf{h}_3$ is non-Gaussian. If we define $\boldsymbol{\omega}$ to be the vector of coefficients on lagged $\Delta\mathbf{y}$,

$$\boldsymbol{\omega} \equiv (\boldsymbol{\zeta}_{i1}', \boldsymbol{\zeta}_{i2}', \ldots, \boldsymbol{\zeta}_{i,n-1}')',$$

then the preceding results imply that

$$\mathbf{Y}_T(\mathbf{b}_T^* - \boldsymbol{\beta}^*) = \begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\omega}}_T - \boldsymbol{\omega}) \\ T^{1/2}(\hat{\alpha}_{i,T}^* - \alpha_i^*) \\ T(\hat{\boldsymbol{\rho}}_{i,T}^* - \boldsymbol{\rho}_i^*) \\ T^{3/2}(\hat{\gamma}_{i,T} - \gamma_i) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\boldsymbol{\eta} \end{bmatrix}, \qquad [18.2.49]$$

where $\boldsymbol{\eta} \equiv (h_2, \mathbf{h}_3', h_4)'$ and $\mathbf{Q}$ is the $[(n+1) \times (n+1)]$ lower right block of the matrix in [18.2.46]. Thus, as usual, the coefficients on $\mathbf{u}_{t-s}$ in [18.2.43] are asymptotically Gaussian:

$$\sqrt{T}(\hat{\boldsymbol{\omega}}_{i,T} - \boldsymbol{\omega}_i) \xrightarrow{L} N(\mathbf{0}, \sigma_{ii}\mathbf{V}^{-1}).$$

These coefficients are, of course, numerically identical to the coefficients on $\Delta\mathbf{y}_{t-s}$ in [18.2.38]. Any $F$ tests involving just these coefficients are also identical for the two parameterizations. Hence, an $F$ test about $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1}$ in [18.2.38] has the usual limiting $\chi^2$ distribution. This is the same asymptotic distribution as if [18.2.38] were estimated with $\boldsymbol{\rho} = \mathbf{I}_n$ imposed; that is, it is the same asymptotic distribution whether the regression is estimated in levels or in differences.

Since $\hat{\boldsymbol{\rho}}_T^*$ and $\hat{\boldsymbol{\gamma}}_T$ converge at a faster rate than $\hat{\boldsymbol{\omega}}_T$, the asymptotic distribution of a linear combination of $\hat{\boldsymbol{\omega}}_T$, $\hat{\boldsymbol{\rho}}_T^*$, and $\hat{\boldsymbol{\gamma}}_T$ that puts nonzero weight on $\hat{\boldsymbol{\omega}}_T$ has the same asymptotic distribution as a linear combination that uses the true values for $\boldsymbol{\rho}$ and $\boldsymbol{\gamma}$. This means, for example, that the original coefficients $\hat{\boldsymbol{\Phi}}_s$ of the *VAR* estimated in levels as in [18.2.1] are all individually Gaussian and can be interpreted using the usual $t$ tests. A Wald test of the null hypothesis of $p_0 \geq 1$ lag against the alternative of $p > p_0$ lags again has the usual $\chi^2$ distribution. However, Granger-causality tests typically have nonstandard distributions.

## 18.3. *Spurious Regressions*

Consider a regression of the form

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t,$$

for which elements of $y_t$ and $\mathbf{x}_t$ might be nonstationary. If there does not exist some population value for $\boldsymbol{\beta}$ for which the residual $u_t = y_t - \mathbf{x}_t'\boldsymbol{\beta}$ is $I(0)$, then *OLS* is quite likely to produce spurious results. This phenomenon was first discovered in Monte Carlo experimentation by Granger and Newbold (1974) and later explained theoretically by Phillips (1986).

A general statement of the spurious regression problem can be made as follows. Let $\mathbf{y}_t$ be an $(n \times 1)$ vector of $I(1)$ variables. Define $g \equiv (n - 1)$, and partition $\mathbf{y}_t$ as

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ \mathbf{y}_{2t} \end{bmatrix},$$

where $\mathbf{y}_{2t}$ denotes a $(g \times 1)$ vector. Consider the consequences of an *OLS* regression of the first variable on the others and a constant,

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + u_t. \qquad [18.3.1]$$

The *OLS* coefficient estimates for a sample of size $T$ are given by

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\boldsymbol{\gamma}}_T \end{bmatrix} = \begin{bmatrix} T & \Sigma\mathbf{y}'_{2t} \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}'_{2t} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_{1t} \\ \Sigma\mathbf{y}_{2t}y_{1t} \end{bmatrix}, \qquad [18.3.2]$$

where $\Sigma$ indicates summation over $t$ from 1 to $T$. It turns out that even if $y_{1t}$ is completely unrelated to $\mathbf{y}_{2t}$, the estimated value of $\boldsymbol{\gamma}$ is likely to appear to be statistically significantly different from zero. Indeed, consider any null hypothesis of the form $H_0$: $\mathbf{R}\boldsymbol{\gamma} = \mathbf{r}$ where $\mathbf{R}$ is a known $(m \times g)$ matrix representing $m$ separate hypotheses involving $\boldsymbol{\gamma}$ and $\mathbf{r}$ is a known $(m \times 1)$ vector. The *OLS F* test of this null hypothesis is

$$F_T = \{\mathbf{R}\hat{\boldsymbol{\gamma}}_T - \mathbf{r}\}'\left\{s_T^2\cdot[\mathbf{0} \quad \mathbf{R}]\begin{bmatrix} T & \Sigma\mathbf{y}'_{2t} \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}'_{2t} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}' \end{bmatrix}\right\}^{-1} \times \{\mathbf{R}\hat{\boldsymbol{\gamma}}_T - \mathbf{r}\} \div m, \qquad [18.3.3]$$

where

$$s_T^2 \equiv (T - n)^{-1} \sum_{t=1}^{T} \hat{u}_t^2. \qquad [18.3.4]$$

Unless there is some value for $\boldsymbol{\gamma}$ such that $y_{1t} - \boldsymbol{\gamma}'\mathbf{y}_{2t}$ is stationary, the *OLS* estimate $\hat{\boldsymbol{\gamma}}_T$ will appear to be spuriously precise in the sense that the $F$ test is virtually certain to reject any null hypothesis if the sample size is sufficiently large, even though $\hat{\boldsymbol{\gamma}}_T$ does not provide a consistent estimate of any well-defined population constant!

The following proposition, adapted from Phillips (1986), provides the formal basis for these statements.

***Proposition 18.2:*** *Consider an* $(n \times 1)$ *vector* $\mathbf{y}_t$ *whose first difference is described by*

$$\Delta\mathbf{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s\boldsymbol{\varepsilon}_{t-s}$$

*for* $\boldsymbol{\varepsilon}_t$ *an i.i.d.* $(n \times 1)$ *vector with mean zero, variance* $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{PP}'$, *and finite fourth moments and where* $\{s\cdot\boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ *is absolutely summable. Let* $g \equiv (n - 1)$ *and* $\boldsymbol{\Lambda} \equiv \boldsymbol{\Psi}(1)\cdot\mathbf{P}$. *Partition* $\mathbf{y}_t$ *as* $\mathbf{y}_t = (y_{1t}, \mathbf{y}'_{2t})'$, *and partition* $\boldsymbol{\Lambda\Lambda}'$ *as*

$$\underset{(n\times n)}{\boldsymbol{\Lambda\Lambda}'} = \begin{bmatrix} \underset{(1\times 1)}{\Sigma_{11}} & \underset{(1\times g)}{\boldsymbol{\Sigma}'_{21}} \\ \underset{(g\times 1)}{\boldsymbol{\Sigma}_{21}} & \underset{(g\times g)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}. \qquad [18.3.5]$$

*Suppose that* $\boldsymbol{\Lambda\Lambda}'$ *is nonsingular, and define*

$$(\sigma_1^*)^2 \equiv (\Sigma_{11} - \boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}). \qquad [18.3.6]$$

*Let* $\mathbf{L}_{22}$ *denote the Cholesky factor of* $\boldsymbol{\Sigma}_{22}^{-1}$; *that is,* $\mathbf{L}_{22}$ *is the lower triangular matrix* *satisfying*

$$\Sigma_{22}^{-1} = \mathbf{L}_{22}\mathbf{L}_{22}'. \qquad [18.3.7]$$

*Then the following hold.*

(*a*) *The OLS estimates* $\hat{\alpha}_T$ *and* $\hat{\boldsymbol{\gamma}}_T$ *in* [18.3.2] *are characterized by*

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T \\ \hat{\boldsymbol{\gamma}}_T - \Sigma_{22}^{-1}\Sigma_{21} \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma_1^* h_1 \\ \sigma_1^* \mathbf{L}_{22}\mathbf{h}_2 \end{bmatrix}, \qquad [18.3.8]$$

*where*

$$\begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & \int [\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int [\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1} \times \begin{bmatrix} \int W_1^*(r)\,dr \\ \int \mathbf{W}_2^*(r)\cdot W_1^*(r)\,dr \end{bmatrix} \qquad [18.3.9]$$

*and the integral sign indicates integration over r from* 0 *to* 1, $W_1^*(r)$ *denotes scalar standard Brownian motion, and* $\mathbf{W}_2^*(r)$ *denotes g-dimensional standard Brownian motion with* $\mathbf{W}_2^*(r)$ *independent of* $W_1^*(r)$.

(*b*) *The sum of squared residuals* $RSS_T$ *from OLS estimation of* [18.3.1] *satisfies*

$$T^{-2}\cdot RSS_T \xrightarrow{L} (\sigma_1^*)^2\cdot H, \qquad [18.3.10]$$

*where*

$$H \equiv \int [W_1^*(r)]^2\,dr - \left\{ \begin{bmatrix} \int W_1^*(r)\,dr & \int [W_1^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix} \times \begin{bmatrix} 1 & \int [\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int [\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1} \begin{bmatrix} \int W_1^*(r)\,dr \\ \int [\mathbf{W}_2^*(r)]\cdot[W_1^*(r)]\,dr \end{bmatrix} \right\}. \qquad [18.3.11]$$

(*c*) *The OLS F test* [18.3.3] *satisfies*

$$T^{-1}\cdot F_T \xrightarrow{L} \{\sigma_1^*\cdot\mathbf{R}^*\mathbf{h}_2 - \mathbf{r}^*\}' \times \left\{ (\sigma_1^*)^2\cdot H[\mathbf{0} \quad \mathbf{R}^*] \times \begin{bmatrix} 1 & \int [\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int [\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}' \\ \mathbf{R}^{*\prime} \end{bmatrix} \right\}^{-1} \times \{\sigma_1^*\cdot\mathbf{R}^*\mathbf{h}_2 - \mathbf{r}^*\} \div m, \qquad [18.3.12]$$

*where*

$$\mathbf{R}^* \equiv \mathbf{R}\cdot\mathbf{L}_{22}$$
$$\mathbf{r}^* \equiv \mathbf{r} - \mathbf{R}\Sigma_{22}^{-1}\Sigma_{21}.$$

The simplest illustration of Proposition 18.2 is provided when $y_{1t}$ and $y_{2t}$ are scalars following totally unrelated random walks:

$$y_{1t} = y_{1,t-1} + \varepsilon_{1t} \qquad [18.3.13]$$

$$y_{2t} = y_{2,t-1} + \varepsilon_{2t}, \qquad [18.3.14]$$

where $\varepsilon_{1t}$ is i.i.d. with mean zero and variance $\sigma_1^2$, $\varepsilon_{2t}$ is i.i.d. with mean zero and variance $\sigma_2^2$, and $\varepsilon_{1t}$ is independent of $\varepsilon_{2\tau}$ for all $t$ and $\tau$. For $\mathbf{y}_t = (y_{1t}, y_{2t})'$, this specification implies

$$\mathbf{P} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

$$\mathbf{\Psi}(1) = \mathbf{I}_2$$

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \mathbf{\Psi}(1)\cdot\mathbf{P}\cdot\mathbf{P}'\cdot[\mathbf{\Psi}(1)]' = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$\sigma_1^* = \sigma_1$$

$$L_{22} = 1/\sigma_2.$$

Result (a) then claims that an *OLS* regression of $y_{1t}$ on $y_{2t}$ and a constant,

$$y_{1t} = \alpha + \gamma y_{2t} + u_t, \qquad [18.3.15]$$

produces estimates $\hat{\alpha}_T$ and $\hat{\gamma}_T$ characterized by

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T \\ \hat{\gamma}_T \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma_1\cdot h_1 \\ (\sigma_1/\sigma_2)\cdot h_2 \end{bmatrix}.$$

Note the contrast between this result and any previous asymptotic distribution analyzed. Usually, the *OLS* estimates are consistent with $\mathbf{b}_T \xrightarrow{p} \mathbf{0}$ and must be multiplied by some increasing function of $T$ in order to obtain a nondegenerate asymptotic distribution. Here, however, neither estimate is consistent—different arbitrarily large samples will have randomly differing estimates $\hat{\gamma}_T$. Indeed, the estimate of the constant term $\hat{\alpha}_T$ actually *diverges*, and must be *divided* by $T^{1/2}$ to obtain a random variable with a well-specified distribution—the estimate $\hat{\alpha}_T$ itself is likely to get farther and farther from the true value of zero as the sample size $T$ increases.

Result (b) implies that the usual *OLS* estimate of the variance of $u_t$,

$$s_T^2 = (T - n)^{-1}\cdot RSS_T,$$

again diverges as $T \to \infty$. To obtain an estimate that does not grow with the sample size, the residual sum of squares has to be divided by $T^2$ rather than $T$. In this respect, the residuals $\hat{u}_t$ from a spurious regression behave like a unit root process; if $\xi_t$ is a scalar $I(1)$ series, then $T^{-1}\Sigma\xi_t^2$ diverges and $T^{-2}\Sigma\xi_t^2$ converges. To see why $\hat{u}_t$ behaves like an $I(1)$ series, notice that the *OLS* residual is given by

$$\hat{u}_t = y_{1t} - \hat{\alpha}_T - \hat{\boldsymbol{\gamma}}_T'\mathbf{y}_{2t},$$

from which

$$\Delta\hat{u}_t = \Delta y_{1t} - \hat{\boldsymbol{\gamma}}_T'\cdot\Delta\mathbf{y}_{2t} = [1 \quad -\hat{\boldsymbol{\gamma}}_T']\begin{bmatrix} \Delta y_{1t} \\ \Delta\mathbf{y}_{2t} \end{bmatrix} \xrightarrow{L} [1 \quad -\mathbf{h}_2^{*\prime}]\Delta\mathbf{y}_t, \qquad [18.3.16]$$

where $\mathbf{h}_2^* \equiv \Sigma_{22}^{-1}\Sigma_{21} + \sigma_1^*\mathbf{L}_{22}\mathbf{h}_2$. This is a random vector $[1 \quad -\mathbf{h}_2^{*\prime}]$ times the $I(0)$ vector $\Delta\mathbf{y}_t$.

Result (c) means that any $OLS$ $t$ or $F$ test based on the spurious regression [18.3.1] also diverges; the $OLS$ $F$ statistic [18.3.3] must be divided by $T$ to obtain a variable that does not grow with the sample size. Since an $F$ test of a single restriction is the square of the corresponding $t$ test, any $t$ statistic would have to be divided by $T^{1/2}$ to obtain a convergent variable. Thus, as the sample size $T$ becomes larger, it becomes increasingly likely that the absolute value of an $OLS$ $t$ test will exceed any arbitrary finite value (such as the usual critical value of $t = 2$). For example, in the regression of [18.3.15], it will appear that $y_{1t}$ and $y_{2t}$ are significantly related whereas in reality they are completely independent.

In more general regressions of the form of [18.3.1], $\Delta y_{1t}$ and $\Delta \mathbf{y}_{2t}$ may be dynamically related through nonzero off-diagonal elements of $\mathbf{P}$ and $\mathbf{\Psi}(L)$. While such correlations will influence the values of the nuisance parameters $\sigma_1^*$, $\Sigma_{21}$, and $\Sigma_{22}$, provided that the conditions of Proposition 18.2 are satisfied, these correlations do not affect the overall nature of the results or rates of convergence for any of the statistics. Note that since $W_1^*(r)$ and $\mathbf{W}_2^*(r)$ are standard Brownian motion, the distributions of $h_1$, $\mathbf{h}_2$, and $H$ in Proposition 18.2 depend only on the number of variables in the regression and not on their dynamic relations.

The condition in Proposition 18.2 that $\Lambda \cdot \Lambda'$ is nonsingular might appear innocuous but is actually quite important. In the case of a single variable ($\mathbf{y}_t = y_{1t}$ with $\Delta y_{1t} = \psi(L)\varepsilon_{1t}$), the matrix $\Lambda \cdot \Lambda'$ would just be the scalar $[\psi(1)\cdot\sigma_1]^2$ and the condition that $\Lambda \cdot \Lambda'$ is nonsingular would come down to the requirement that $\psi(1)$ be nonzero. To understand what this means, suppose that $y_{1t}$ were actually stationary with Wold representation:

$$y_{1t} = \varepsilon_{1t} + C_1\varepsilon_{1,t-1} + C_2\varepsilon_{1,t-2} + \cdots = C(L)\varepsilon_{1t}.$$

Then the first difference $\Delta y_{1t}$ would be described by

$$\Delta y_{1t} = (1 - L)C(L)\varepsilon_{1t} \equiv \psi(L)\varepsilon_{1t},$$

where $\psi(L) \equiv (1 - L)C(L)$, meaning $\psi(1) = (1 - 1)\cdot C(1) = 0$. Thus, if $y_{1t}$ were actually $I(0)$ rather than $I(1)$, the condition that $\Lambda \cdot \Lambda'$ is nonsingular would not be satisfied.

For the more general case in which $\mathbf{y}_t$ is an $(n \times 1)$ vector, the condition that $\Lambda \cdot \Lambda'$ is nonsingular will not be satisfied if some explanatory variable $y_{it}$ is $I(0)$ or if some linear combination of the elements of $\mathbf{y}_t$ is $I(0)$. If $\mathbf{y}_t$ is an $I(1)$ vector but some linear combination of $\mathbf{y}_t$ is $I(0)$, then the elements of $\mathbf{y}_t$ are said to be *cointegrated*. Thus, Proposition 18.2 describes the consequences of $OLS$ estimation of [18.3.1] only when all of the elements of $\mathbf{y}_t$ are $I(1)$ with zero drift and when the vector $\mathbf{y}_t$ is not cointegrated. A regression is spurious only when the residual $u_t$ is nonstationary for all possible values of the coefficient vector.

### Cures for Spurious Regressions

There are three ways in which the problems associated with spurious regressions can be avoided. The first approach is to include lagged values of both the dependent and independent variable in the regression. For example, consider the following model as an alternative to [18.3.15]:

$$y_{1t} = \alpha + \phi y_{1,t-1} + \gamma y_{2t} + \delta y_{2,t-1} + u_t. \qquad [18.3.17]$$

This regression does not satisfy the conditions of Proposition 18.1, because there exist values for the coefficients, specifically $\phi = 1$ and $\gamma = \delta = 0$, for which the error term $u_t$ is $I(0)$. It can be shown that $OLS$ estimation of [18.3.17] yields consistent estimates of all of the parameters. The coefficients $\hat{\gamma}_T$ and $\hat{\delta}_T$ each individually converge at rate $\sqrt{T}$ to a Gaussian distribution, and the $t$ test of the hypothesis that $\gamma = 0$ is asymptotically $N(0, 1)$, as is the $t$ test of the hypothesis that $\delta = 0$. However, an $F$ test of the joint null hypothesis that $\gamma$ and $\delta$ are both zero has a nonstandard limiting distribution; see Exercise 18.3. Hence, including lagged values in the regression is sufficient to solve many of the problems associated with spurious regressions, although tests of some hypotheses will still involve nonstandard distributions.

A second approach is to difference the data before estimating the relation, as in

$$\Delta y_{1t} = \alpha + \gamma \Delta y_{2t} + u_t. \qquad [18.3.18]$$

Clearly, since the regressors and error term $u_t$ are all $I(0)$ for this regression under the null hypothesis, $\hat{\alpha}_T$ and $\hat{\gamma}_T$ both converge at rate $\sqrt{T}$ to Gaussian variables. Any $t$ or $F$ test based on [18.3.18] has the usual limiting Gaussian or $\chi^2$ distribution.

A third approach, analyzed by Blough (1992), is to estimate [18.3.15] with Cochrane-Orcutt adjustment for first-order serial correlation of the residuals. We will see in Proposition 19.4 in the following chapter that if $\hat{u}_t$ denotes the sample residual from $OLS$ estimation of [18.3.15], then the estimated autoregressive coefficient $\hat{\rho}_T$ from an $OLS$ regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ converges in probability to unity. Blough showed that the Cochrane-Orcutt $GLS$ regression is then asymptotically equivalent to the differenced regression [18.3.18].

Because the specification [18.3.18] avoids the spurious regression problem as well as the nonstandard distributions for certain hypotheses associated with the levels regression [18.3.15], many researchers recommend routinely differencing apparently nonstationary variables before estimating regressions. While this is the ideal cure for the problem discussed in this section, there are two different situations in which it might be inappropriate. First, if the data are really stationary (for example, if the true value of $\phi$ in [18.3.17] is 0.9 rather than than unity), then differencing the data can result in a misspecified regression. Second, even if both $y_{1t}$ and $y_{2t}$ are truly $I(1)$ processes, there is an interesting class of models for which the bivariate dynamic relation between $y_1$ and $y_2$ will be misspecified if the researcher simply differences both $y_1$ and $y_2$. This class of models, known as *cointegrated processes*, is discussed in the following chapter.

## APPENDIX 18.A. *Proofs of Chapter 18 Propositions*

■ **Proof of Proposition 18.1.**

(a) This follows from [18.1.7] and [18.1.8] with $r = 1$.
(b) The derivation is identical to that in [11.A.3].
(c) This follows from Proposition 10.2(d).
(d) Note first in a generalization of [17.1.10] and [17.1.11] that

$$\sum_{t=1}^{T} \boldsymbol{\xi}_t \boldsymbol{\xi}_t' = \sum_{t=1}^{T} (\boldsymbol{\xi}_{t-1} + \mathbf{u}_t)(\boldsymbol{\xi}_{t-1} + \mathbf{u}_t)' = \sum_{t=1}^{T} (\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}' + \boldsymbol{\xi}_{t-1}\mathbf{u}_t' + \mathbf{u}_t\boldsymbol{\xi}_{t-1}' + \mathbf{u}_t\mathbf{u}_t'),$$

so that

$$\begin{aligned} \sum_{t=1}^{T} (\boldsymbol{\xi}_{t-1}\mathbf{u}_t' + \mathbf{u}_t\boldsymbol{\xi}_{t-1}') &= \sum_{t=1}^{T} \boldsymbol{\xi}_t\boldsymbol{\xi}_t' - \sum_{t=1}^{T} (\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}') - \sum_{t=1}^{T} (\mathbf{u}_t\mathbf{u}_t') \\ &= \boldsymbol{\xi}_T\boldsymbol{\xi}_T' - \boldsymbol{\xi}_0\boldsymbol{\xi}_0' - \sum_{t=1}^{T} (\mathbf{u}_t\mathbf{u}_t') \qquad [18.A.1] \\ &= \boldsymbol{\xi}_T\boldsymbol{\xi}_T' - \sum_{t=1}^{T} (\mathbf{u}_t\mathbf{u}_t'). \end{aligned}$$

Dividing by $T$,

$$T^{-1}\sum_{t=1}^{T}(\xi_{t-1}\mathbf{u}_t' + \mathbf{u}_t\xi_{t-1}') = T^{-1}\xi_T\xi_T' - T^{-1}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t'. \qquad [18.A.2]$$

But from [18.1.7], $\xi_T = T\cdot\mathbf{X}_T(1)$. Hence, from [18.1.8] and the continuous mapping theorem,

$$T^{-1}\xi_T\xi_T' = [\sqrt{T}\cdot\mathbf{X}_T(1)]\,[\sqrt{T}\cdot\mathbf{X}_T(1)]' \xrightarrow{L} \Lambda\cdot[\mathbf{W}(1)]\cdot[\mathbf{W}(1)]'\cdot\Lambda'. \qquad [18.A.3]$$

Substituting this along with result (c) into [18.A.2] produces

$$T^{-1}\sum_{t=1}^{T}(\xi_{t-1}\mathbf{u}_t' + \mathbf{u}_t\xi_{t-1}') \xrightarrow{L} \Lambda\cdot[\mathbf{W}(1)]\cdot[\mathbf{W}(1)]'\cdot\Lambda' - \Gamma_0, \qquad [18.A.4]$$

which establishes result (d) for $s = 0$.

For $s > 0$, we have

$$\begin{aligned}
&T^{-1}\sum_{t=s+1}^{T}(\xi_{t-1}\mathbf{u}_{t-s}' + \mathbf{u}_{t-s}\xi_{t-1}')\\
&\quad = T^{-1}\sum_{t=s+1}^{T}[(\xi_{t-s-1} + \mathbf{u}_{t-s} + \mathbf{u}_{t-s+1} + \cdots + \mathbf{u}_{t-1})\mathbf{u}_{t-s}'\\
&\qquad + \mathbf{u}_{t-s}(\xi_{t-s-1}' + \mathbf{u}_{t-s}' + \mathbf{u}_{t-s+1}' + \cdots + \mathbf{u}_{t-1}')]\\
&\quad = T^{-1}\sum_{t=s+1}^{T}(\xi_{t-s-1}\mathbf{u}_{t-s}' + \mathbf{u}_{t-s}\xi_{t-s-1}')\\
&\qquad + T^{-1}\sum_{t=s+1}^{T}[(\mathbf{u}_{t-s}\mathbf{u}_{t-s}') + (\mathbf{u}_{t-s+1}\mathbf{u}_{t-s}') + \cdots + (\mathbf{u}_{t-1}\mathbf{u}_{t-s}')\\
&\qquad + (\mathbf{u}_{t-s}\mathbf{u}_{t-s}') + (\mathbf{u}_{t-s}\mathbf{u}_{t-s+1}') + \cdots + (\mathbf{u}_{t-s}\mathbf{u}_{t-1}')]\\
&\quad \xrightarrow{L} \Lambda\cdot[\mathbf{W}(1)]\cdot[\mathbf{W}(1)]'\cdot\Lambda' - \Gamma_0\\
&\qquad + [\Gamma_0 + \Gamma_1 + \cdots + \Gamma_{s-1} + \Gamma_0 + \Gamma_{-1} + \cdots + \Gamma_{-s+1}],
\end{aligned}$$

by virtue of [18.A.4] and result (c).

(e) See Phillips (1988).

(f) Define $\xi_t^* \equiv \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$ and $E(\varepsilon_t\varepsilon_t') = \mathbf{PP}'$. Notice that result (e) implies that

$$T^{-1}\sum_{t=1}^{T}\xi_{t-1}^*\varepsilon_t' \xrightarrow{L} \mathbf{P}\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\mathbf{P}'. \qquad [18.A.5]$$

For $\xi_t = \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_t$, equation [18.1.6] establishes that

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\xi_{t-1}\varepsilon_t' &= T^{-1}\sum_{t=1}^{T}\{\Psi(1)\cdot\xi_{t-1}^* + \eta_{t-1} - \eta_0\}\cdot\varepsilon_t'\\
&= \Psi(1)\cdot T^{-1}\sum_{t=1}^{T}\xi_{t-1}^*\varepsilon_t' + T^{-1}\sum_{t=1}^{T}(\eta_{t-1} - \eta_0)\cdot\varepsilon_t'.
\end{aligned} \qquad [18.A.6]$$

But each column of $\{(\eta_{t-1} - \eta_0)\cdot\varepsilon_t'\}_{t=1}^{T}$ is a martingale difference sequence with finite variance, and so, from Example 7.11 of Chapter 7,

$$T^{-1}\sum_{t=1}^{T}(\eta_{t-1} - \eta_0)\cdot\varepsilon_t' \xrightarrow{p} \mathbf{0}. \qquad [18.A.7]$$

Substituting [18.A.5] and [18.A.7] into [18.A.6] produces

$$T^{-1}\sum_{t=1}^{T}\xi_{t-1}\varepsilon_t' \xrightarrow{L} \Psi(1)\cdot\mathbf{P}\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\mathbf{P}',$$

as claimed.

(g) This was shown in [18.1.9].

(h) As in [17.3.17], we have

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1} = T^{-1/2}\sum_{t=1}^{T}\mathbf{u}_t - T^{-3/2}\sum_{t=1}^{T}t\mathbf{u}_t$$

or

$$T^{-3/2}\sum_{t=1}^{T} t\mathbf{u}_t = T^{-1/2}\sum_{t=1}^{T}\mathbf{u}_t - T^{-3/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{t-1} \xrightarrow{L} \Lambda\cdot\mathbf{W}(1) - \Lambda\cdot\int_0^1 \mathbf{W}(r)\,dr, \qquad [18.A.8]$$

from results (a) and (g). This establishes result (h) for $s = 0$. The asymptotic distribution is the same for any $s$, from simple adaptation of the proof of Proposition 17.3(g).

(i) As in [17.3.22],

$$T^{-2}\sum_{t=1}^{T}\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}'_{t-1} = \int_0^1 [\sqrt{T}\cdot\mathbf{X}_T(r)]\cdot[\sqrt{T}\cdot\mathbf{X}_T(r)]'\,dr$$

$$\xrightarrow{L} \Lambda\cdot\left\{\int_0^1 [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\Lambda'.$$

(j), (k), and (l) parallel Proposition 17.3(i), (j), and (k). ■

■ **Proof of Proposition 18.2.** The asymptotic distributions are easier to calculate if we work with the following transformed variables:

$$y^*_{1t} \equiv y_{1t} - \boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}^{-1}_{22}\mathbf{y}_{2t} \qquad [18.A.9]$$
$$\mathbf{y}^*_{2t} \equiv \mathbf{L}'_{22}\mathbf{y}_{2t}. \qquad [18.A.10]$$

Note that the inverses $\boldsymbol{\Sigma}^{-1}_{22}$, $(\sigma^*_1)^{-1}$, and $\mathbf{L}^{-1}_{22}$ all exist, since $\Lambda\Lambda'$ is symmetric positive definite. An $OLS$ regression of $y^*_{1t}$ on a constant and $\mathbf{y}^*_{2t}$,

$$y^*_{1t} = \alpha^* + \boldsymbol{\gamma}^{*\prime}\mathbf{y}^*_{2t} + u^*_t, \qquad [18.A.11]$$

would yield estimates

$$\begin{bmatrix}\hat{\alpha}^*_T\\ \hat{\boldsymbol{\gamma}}^*_T\end{bmatrix} = \begin{bmatrix} T & \Sigma\mathbf{y}^{*\prime}_{2t}\\ \Sigma\mathbf{y}^*_{2t} & \Sigma\mathbf{y}^*_{2t}\mathbf{y}^{*\prime}_{2t}\end{bmatrix}^{-1}\begin{bmatrix}\Sigma y^*_{1t}\\ \Sigma\mathbf{y}^*_{2t}y^*_{1t}\end{bmatrix}. \qquad [18.A.12]$$

Clearly, the residuals from $OLS$ estimation of [18.A.11] are identical to those from $OLS$ estimation of [18.3.1]:

$$\begin{aligned} y_{1t} - \hat{\alpha}_T - \hat{\boldsymbol{\gamma}}'_T\mathbf{y}_{2t} &= y^*_{1t} - \hat{\alpha}^*_T - \hat{\boldsymbol{\gamma}}^{*\prime}_T\mathbf{y}^*_{2t}\\ &= (y_{1t} - \boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}^{-1}_{22}\mathbf{y}_{2t}) - \hat{\alpha}^*_T - \hat{\boldsymbol{\gamma}}^{*\prime}_T(\mathbf{L}'_{22}\mathbf{y}_{2t})\\ &= y_{1t} - \hat{\alpha}^*_T - \{\hat{\boldsymbol{\gamma}}^{*\prime}_T\mathbf{L}'_{22} + \boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}^{-1}_{22}\}\mathbf{y}_{2t}.\end{aligned}$$

The $OLS$ estimates for the transformed regression [18.A.11] are thus related to those of the original regression [18.3.1] by

$$\begin{aligned}\hat{\alpha}_T &= \hat{\alpha}^*_T\\ \hat{\boldsymbol{\gamma}}_T &= \mathbf{L}_{22}\hat{\boldsymbol{\gamma}}^*_T + \boldsymbol{\Sigma}^{-1}_{22}\boldsymbol{\Sigma}_{21},\end{aligned} \qquad [18.A.13]$$

implying that

$$\begin{aligned}\hat{\boldsymbol{\gamma}}^*_T &= \mathbf{L}^{-1}_{22}\hat{\boldsymbol{\gamma}}_T - \mathbf{L}^{-1}_{22}\boldsymbol{\Sigma}^{-1}_{22}\boldsymbol{\Sigma}_{21}\\ &= \mathbf{L}^{-1}_{22}\hat{\boldsymbol{\gamma}}_T - \mathbf{L}^{-1}_{22}(\mathbf{L}_{22}\mathbf{L}'_{22})\boldsymbol{\Sigma}_{21}\\ &= \mathbf{L}^{-1}_{22}\hat{\boldsymbol{\gamma}}_T - \mathbf{L}'_{22}\boldsymbol{\Sigma}_{21}.\end{aligned} \qquad [18.A.14]$$

The usefulness of this transformation is as follows. Notice that

$$\begin{bmatrix} y^*_{1t}/\sigma^*_1\\ \mathbf{y}^*_{2t}\end{bmatrix} = \begin{bmatrix}(1/\sigma^*_1) & (-1/\sigma^*_1)\cdot\boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}^{-1}_{22}\\ \mathbf{0} & \mathbf{L}'_{22}\end{bmatrix}\begin{bmatrix} y_{1t}\\ \mathbf{y}_{2t}\end{bmatrix} \equiv \mathbf{L}'\mathbf{y}_t$$

for

$$\mathbf{L}' \equiv \begin{bmatrix}(1/\sigma^*_1) & (-1/\sigma^*_1)\cdot\boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}^{-1}_{22}\\ \mathbf{0} & \mathbf{L}'_{22}\end{bmatrix}.$$

Moreover,

$$\mathbf{L'}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\mathbf{L} = \begin{bmatrix} (1/\sigma_1^*) & (-1/\sigma_1^*)\cdot\boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{L}_{22}' \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} (1/\sigma_1^*) & \mathbf{0}' \\ (-1/\sigma_1^*)\cdot\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{L}_{22} \end{bmatrix}$$
$$= \begin{bmatrix} (1/\sigma_1^*)\cdot(\Sigma_{11} - \boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) & \mathbf{0}' \\ \mathbf{L}_{22}'\boldsymbol{\Sigma}_{21} & \mathbf{L}_{22}'\boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} (1/\sigma_1^*) & \mathbf{0}' \\ (-1/\sigma_1^*)\cdot\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{L}_{22} \end{bmatrix}$$
$$= \begin{bmatrix} (\Sigma_{11} - \boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})/(\sigma_1^*)^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{L}_{22}'\boldsymbol{\Sigma}_{22}\mathbf{L}_{22} \end{bmatrix}. \qquad [18.A.15]$$

But [18.3.7] implies that

$$\boldsymbol{\Sigma}_{22} = (\mathbf{L}_{22}\mathbf{L}_{22}')^{-1} = (\mathbf{L}_{22}')^{-1}\mathbf{L}_{22}^{-1},$$

from which

$$\mathbf{L}_{22}'\boldsymbol{\Sigma}_{22}\mathbf{L}_{22} = \mathbf{L}_{22}'\{(\mathbf{L}_{22}')^{-1}\mathbf{L}_{22}^{-1}\}\mathbf{L}_{22} = \mathbf{I}_g.$$

Substituting this and [18.3.6] into [18.A.15] results in

$$\mathbf{L'}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\mathbf{L} = \mathbf{I}_n. \qquad [18.A.16]$$

One of the implications is that if $\mathbf{W}(r)$ is $n$-dimensional standard Brownian motion, then the $n$-dimensional process $\mathbf{W}^*(r)$ defined by

$$\mathbf{W}^*(r) \equiv \mathbf{L'}\boldsymbol{\Lambda}\cdot\mathbf{W}(r) \qquad [18.A.17]$$

is Brownian motion with variance matrix $\mathbf{L'}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\mathbf{L} = \mathbf{I}_n$. In other words, $\mathbf{W}^*(r)$ could also be described as standard Brownian motion. Since result (g) of Proposition 18.1 implies that

$$T^{-3/2}\sum_{t=1}^{T}\mathbf{y}_t \xrightarrow{L} \boldsymbol{\Lambda}\cdot\int_0^1 \mathbf{W}(r)\,dr,$$

it follows that

$$\begin{bmatrix} T^{-3/2}\Sigma y_{1t}^*/\sigma_1^* \\ T^{-3/2}\Sigma \mathbf{y}_{2t}^* \end{bmatrix} = T^{-3/2}\sum_{t=1}^{T}\mathbf{L'}\mathbf{y}_t \xrightarrow{L} \mathbf{L'}\boldsymbol{\Lambda}\cdot\int_0^1 \mathbf{W}(r)\,dr = \int_0^1 \mathbf{W}^*(r)\,dr. \qquad [18.A.18]$$

Similarly, result (i) of Proposition 18.1 gives

$$\begin{bmatrix} T^{-2}\Sigma(y_{1t}^*)^2/(\sigma_1^*)^2 & T^{-2}\Sigma y_{1t}^*\mathbf{y}_{2t}^{*\prime}/\sigma_1^* \\ T^{-2}\Sigma\mathbf{y}_{2t}^*y_{1t}^*/\sigma_1^* & T^{-2}\Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}$$
$$= \mathbf{L'}\cdot T^{-2}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\cdot\mathbf{L}$$
$$\xrightarrow{L} \mathbf{L'}\boldsymbol{\Lambda}\cdot\left\{\int_0^1 [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\boldsymbol{\Lambda}'\mathbf{L} \qquad [18.A.19]$$
$$= \int_0^1 [\mathbf{W}^*(r)]\cdot[\mathbf{W}^*(r)]'\,dr.$$

It is now straightforward to prove the claims in Proposition 18.2.

**Proof of (a).** If [18.A.12] is divided by $\sigma_1^*$ and premultiplied by the matrix

$$\begin{bmatrix} T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix},$$

the result is

$$\begin{bmatrix} T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix} \begin{bmatrix} \hat{\alpha}_T^*/\sigma_1^* \\ \hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix}$$

$$= \begin{bmatrix} T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix} \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}^{*\prime} \\ \Sigma\mathbf{y}_{2t}^* & \Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2} & \mathbf{0}' \\ \mathbf{0} & T^{-2}\mathbf{I}_g \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2} & \mathbf{0}' \\ \mathbf{0} & T^{-2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} \Sigma y_{1t}^*/\sigma_1^* \\ \Sigma\mathbf{y}_{2t}^* y_{1t}^*/\sigma_1^* \end{bmatrix}$$

$$= \left( \begin{bmatrix} T^{-3/2} & \mathbf{0}' \\ \mathbf{0} & T^{-2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}^{*\prime} \\ \Sigma\mathbf{y}_{2t}^* & \Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix} \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} T^{-3/2} & \mathbf{0}' \\ \mathbf{0} & T^{-2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} \Sigma y_{1t}^*/\sigma_1^* \\ \Sigma\mathbf{y}_{2t}^* y_{1t}^*/\sigma_1^* \end{bmatrix} \right)$$

or

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T^*/\sigma_1^* \\ \hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}^{*\prime} \\ T^{-3/2}\Sigma\mathbf{y}_{2t}^* & T^{-2}\Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2}\Sigma y_{1t}^*/\sigma_1^* \\ T^{-2}\Sigma\mathbf{y}_{2t}^* y_{1t}^*/\sigma_1^* \end{bmatrix}. \qquad [18.A.20]$$

Partition $\mathbf{W}^*(r)$ as

$$\underset{(n\times 1)}{\mathbf{W}^*(r)} = \begin{bmatrix} \underset{(1\times 1)}{W_1^*(r)} \\ \underset{(g\times 1)}{\mathbf{W}_2^*(r)} \end{bmatrix}.$$

Applying [18.A.18] and [18.A.19] to [18.A.20] results in

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T^*/\sigma_1^* \\ \hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \int [\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int [\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1} \begin{bmatrix} \int W_1^*(r)\,dr \\ \int \mathbf{W}_2^*(r)\cdot W_1^*(r)\,dr \end{bmatrix} \qquad [18.A.21]$$

$$= \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix}.$$

Recalling the relation between the transformed estimates and the original estimates given in [18.A.14], this establishes that

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T/\sigma_1^* \\ (1/\sigma_1^*)\cdot[\mathbf{L}_{22}^{-1}\hat{\boldsymbol{\gamma}}_T - \mathbf{L}_{22}'\boldsymbol{\Sigma}_{21}] \end{bmatrix} \xrightarrow{L} \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix}.$$

Premultiplying by

$$\begin{bmatrix} \sigma_1^* & \mathbf{0}' \\ \mathbf{0} & \sigma_1^*\mathbf{L}_{22} \end{bmatrix}$$

and recalling [18.3.7] produces [18.3.8].

**Proof of (b).** Again we exploit the fact that *OLS* estimation of [18.A.11] would produce the identical residuals that would result from *OLS* estimation of [18.3.1]. Recall the expression for the residual sum of squares in [4.A.6]:

$$RSS_T = \Sigma(y_{1t}^*)^2 - \left\{ [\Sigma y_{1t}^* \quad \Sigma y_{1t}^*\mathbf{y}_{2t}^{*\prime}] \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}^{*\prime} \\ \Sigma\mathbf{y}_{2t}^* & \Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_{1t}^* \\ \Sigma y_{1t}^*\mathbf{y}_{2t}^* \end{bmatrix} \right\}$$

$$= \Sigma(y_{1t}^*)^2 - \left\{ [\Sigma y_{1t}^* \quad \Sigma y_{1t}^*\mathbf{y}_{2t}^{*\prime}] \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix} \right.$$

$$\left. \times \left( \begin{bmatrix} T^{-3/2} & \mathbf{0}' \\ \mathbf{0} & T^{-2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}^{*\prime} \\ \Sigma\mathbf{y}_{2t}^* & \Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix} \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix} \right)^{-1} \begin{bmatrix} T^{-3/2} & \mathbf{0}' \\ \mathbf{0} & T^{-2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} \Sigma y_{1t}^* \\ \Sigma y_{1t}^*\mathbf{y}_{2t}^* \end{bmatrix} \right\}.$$

[18.A.22]

If both sides of [18.A.22] are divided by $(T\cdot\sigma_1^*)^2$, the result is

$$T^{-2}\cdot RSS_T/(\sigma_1^*)^2$$

$$= T^{-2}\Sigma(y_{1t}^*/\sigma_1^*)^2 - \Bigg\{[T^{-3/2}\Sigma(y_{1t}^*/\sigma_1^*) \quad T^{-2}\Sigma(y_{1t}^*/\sigma_1^*)\mathbf{y}_{2t}^{*\prime}]$$

$$\times \begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}^{*\prime} \\ T^{-3/2}\Sigma\mathbf{y}_{2t}^* & T^{-2}\Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2}\Sigma y_{1t}^*/\sigma_1^* \\ T^{-2}\Sigma\mathbf{y}_{2t}^* y_{1t}/\sigma_1^* \end{bmatrix}\Bigg\}$$

$$\xrightarrow{L} \int [W_1^*(r)]^2\,dr - \Bigg\{\Bigg[\int W_1^*(r)\,dr \quad \int [W_1^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr\Bigg]$$

$$\times \begin{bmatrix} 1 & \int [\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int [\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1} \begin{bmatrix} \int W_1^*(r)\,dr \\ \int [\mathbf{W}_2^*(r)]\cdot[W_1^*(r)]\,dr \end{bmatrix}\Bigg\}.$$

**Proof of (c).** Note that an $F$ test of the hypothesis $H_0$: $\mathbf{R}\boldsymbol{\gamma} = \mathbf{r}$ for the original regression [18.3.1] would produce exactly the same value as an $F$ test of $\mathbf{R}^*\boldsymbol{\gamma}^* = \mathbf{r}^*$ for $OLS$ estimation of [18.A.11], where, from [18.A.13],

$$\mathbf{R}\boldsymbol{\gamma} - \mathbf{r} = \mathbf{R}\{\mathbf{L}_{22}\boldsymbol{\gamma}^* + \Sigma_{22}^{-1}\Sigma_{21}\} - \mathbf{r} = \mathbf{R}^*\boldsymbol{\gamma}^* - \mathbf{r}^*$$

for

$$\mathbf{R}^* \equiv \mathbf{R}\cdot\mathbf{L}_{22} \qquad [18.A.23]$$

$$\mathbf{r}^* \equiv \mathbf{r} - \mathbf{R}\Sigma_{22}^{-1}\Sigma_{21}. \qquad [18.A.24]$$

The $OLS$ $F$ test of $\mathbf{R}^*\boldsymbol{\gamma}^* = \mathbf{r}^*$ is given by

$$F_T = \{\mathbf{R}^*\hat{\boldsymbol{\gamma}}_T^* - \mathbf{r}^*\}'$$

$$\times \left\{[s_T^*]^2\cdot[\mathbf{0} \quad \mathbf{R}^*]\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}^{*\prime} \\ \Sigma\mathbf{y}_{2t}^* & \Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}^{*\prime} \end{bmatrix}\right\}^{-1}\{\mathbf{R}^*\hat{\boldsymbol{\gamma}}_T^* - \mathbf{r}^*\} \div m,$$

from which

$$T^{-1}\cdot F_T = \{\mathbf{R}^*\hat{\boldsymbol{\gamma}}_T^* - \mathbf{r}^*\}'$$

$$\times \left\{T^{-1}\cdot[s_T^*]^2\cdot[\mathbf{0} \quad \mathbf{R}^*]\begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T\cdot\mathbf{I}_g \end{bmatrix}\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}^{*\prime} \\ \Sigma\mathbf{y}_{2t}^* & \Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1}\right.$$

$$\left.\times \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T\cdot\mathbf{I}_g \end{bmatrix}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}^{*\prime} \end{bmatrix}\right\}^{-1}\{\mathbf{R}^*\hat{\boldsymbol{\gamma}}_T^* - \mathbf{r}^*\} \div m \qquad [18.A.25]$$

$$= \{\mathbf{R}^*\hat{\boldsymbol{\gamma}}_T^* - \mathbf{r}^*\}'\Bigg\{T^{-1}\cdot[s_T^*]^2\cdot[\mathbf{0} \quad \mathbf{R}^*]$$

$$\times \begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}^{*\prime} \\ T^{-3/2}\Sigma\mathbf{y}_{2t}^* & T^{-2}\Sigma\mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}^{*\prime} \end{bmatrix}\Bigg\}^{-1}\{\mathbf{R}^*\hat{\boldsymbol{\gamma}}_T^* - \mathbf{r}^*\} \div m.$$

But

$$[s_T^*]^2 = (T-n)^{-1}\sum_{t=1}^{T}(\hat{u}_t^*)^2 = (T-n)^{-1}\sum_{t=1}^{T}\hat{u}_t^2,$$

and so, from result (b),

$$T^{-1}\cdot[s_T^*]^2 = [T/(T-n)]\cdot T^{-2}\cdot RSS_T \xrightarrow{L} (\sigma_1^*)^2\cdot H. \qquad [18.A.26]$$

Moreover, [18.A.18] and [18.A.19] imply that

$$\begin{bmatrix} 1 & T^{-3/2}\Sigma \mathbf{y}_{2t}^{*\prime} \\ T^{-3/2}\Sigma \mathbf{y}_{2t}^* & T^{-2}\Sigma \mathbf{y}_{2t}^*\mathbf{y}_{2t}^{*\prime} \end{bmatrix}^{-1} \xrightarrow{L} \begin{bmatrix} 1 & \int [\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int [\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1}, \qquad [18.A.27]$$

while from [18.A.21],

$$\hat{\boldsymbol{\gamma}}_T^* \xrightarrow{L} \sigma_1^*\cdot\mathbf{h}_2. \qquad [18.A.28]$$

Snbstituting [18.A.26] through [18.A.28] into [18.A.25], we conclude that

$$T^{-1}\cdot F_T \xrightarrow{L} \{\sigma_1^*\cdot\mathbf{R}^*\mathbf{h}_2 - \mathbf{r}^*\}' \times \left\{(\sigma_1^*)^2\cdot H[\mathbf{0} \quad \mathbf{R}^*] \times \begin{bmatrix} 1 & \int [\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int [\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}' \\ \mathbf{R}^{*\prime} \end{bmatrix}\right\}^{-1} \{\sigma_1^*\cdot\mathbf{R}^*\mathbf{h}_2 - \mathbf{r}^*\} \div m. \quad \blacksquare$$

## Chapter 18 Exercises

18.1. Consider $OLS$ estimation of

$$y_{it} = \boldsymbol{\zeta}_{i1}'\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_{i2}'\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{i,p-1}'\Delta\mathbf{y}_{t-p+1} + \alpha_i + \boldsymbol{\rho}_i'\mathbf{y}_{t-1} + \varepsilon_{it},$$

where $y_{it}$ is the $i$th element of the $(n \times 1)$ vector $\mathbf{y}_t$ and $\varepsilon_{it}$ is the $i$th element of the $(n \times 1)$ vector $\boldsymbol{\varepsilon}_t$. Assume that $\boldsymbol{\varepsilon}_t$ is i.i.d. with mean zero, positive definite variance $\boldsymbol{\Omega}$, and finite fourth moments and that $\Delta\mathbf{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t$, where the sequence of $(n \times n)$ matrices $\{s\cdot\boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ is asolutely summable and $\boldsymbol{\Psi}(1)$ is nonsingular. Let $k = np + 1$ denote the number of regressors, and define

$$\mathbf{x}_t \equiv (\Delta\mathbf{y}_{t-1}', \Delta\mathbf{y}_{t-2}', \ldots, \Delta\mathbf{y}_{t-p+1}', 1, \mathbf{y}_{t-1}')'.$$

Let $\mathbf{b}_T$ denote the $(k \times 1)$ vector of estimated coefficients:

$$\mathbf{b}_T = (\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}(\Sigma\mathbf{x}_t y_{it}),$$

where $\Sigma$ denotes summation over $t$ from 1 to $T$. Consider any null hypothesis $H_0$: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ that involves only the coefficients on $\Delta\mathbf{y}_{t-s}$—that is, $\mathbf{R}$ is of the form

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

Let $\chi_T^2$ be the Wald form of the $OLS$ $\chi^2$ test of $H_0$:

$$\chi_T^2 \equiv (\mathbf{R}\mathbf{b}_T - \mathbf{r})'[s_T^2\mathbf{R}(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}_T - \mathbf{r}),$$

where

$$s_T^2 \equiv (T-k)^{-1}\Sigma(y_{it} - \mathbf{b}_T'\mathbf{x}_t)^2.$$

Under the maintained hypothesis that $\alpha_i = 0$ and $\boldsymbol{\rho}_i' = \mathbf{e}_i'$ (where $\mathbf{e}_i'$ denotes the $i$th row of $\mathbf{I}_n$), show that $\chi_T^2 \xrightarrow{L} \chi^2(m)$.

18.2. Suppose that the regression model

$$y_{it} = \boldsymbol{\zeta}_{i1}'\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_{i2}'\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{i,p-1}'\Delta\mathbf{y}_{t-p+1} + \alpha_i + \boldsymbol{\rho}_i'\mathbf{y}_{t-1} + \varepsilon_{it}$$

satisfies the conditions of Exercise 18.1. Partition this regression as in [18.2.37]:

$$\begin{aligned} y_{it} = {} & \boldsymbol{\beta}_1'\Delta\mathbf{y}_{1,t-1} + \boldsymbol{\gamma}_1'\Delta\mathbf{y}_{2,t-1} + \boldsymbol{\beta}_2'\Delta\mathbf{y}_{1,t-2} + \boldsymbol{\gamma}_2'\Delta\mathbf{y}_{2,t-2} + \cdots \\ & + \boldsymbol{\beta}_{p-1}'\Delta\mathbf{y}_{1,t-p+1} + \boldsymbol{\gamma}_{p-1}'\Delta\mathbf{y}_{2,t-p+1} + \alpha_i + \boldsymbol{\eta}'\mathbf{y}_{1,t-1} \\ & + \boldsymbol{\delta}'\mathbf{y}_{2,t-1} + \varepsilon_{it}, \end{aligned}$$

where $\mathbf{y}_{1t}$ is an $(n_1 \times 1)$ vector and $\mathbf{y}_{2t}$ is an $(n_2 \times 1)$ vector with $n_1 + n_2 = n$. Consider the null hypothesis $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \cdots = \boldsymbol{\gamma}_{p-1} = \boldsymbol{\delta} = \mathbf{0}$. Describe the asymptotic distribution of the Wald form of the *OLS* $\chi^2$ test of this null hypothesis.

18.3. Consider *OLS* estimation of

$$y_{1t} = \gamma\Delta y_{2t} + \alpha + \phi y_{1,t-1} + \eta y_{2,t-1} + u_t,$$

where $y_{1t}$ and $y_{2t}$ are independent random walks as specified in [18.3.13] and [18.3.14]. Note that the fitted valnes of this regression are identical to those for [18.3.17] with $\hat{\alpha}_T$, $\hat{\gamma}_T$, and $\hat{\phi}_T$ the same for both regressions and $\hat{\delta}_T = \hat{\eta}_T - \hat{\gamma}_T$.

(a) Show that

$$\begin{bmatrix} T^{1/2}\hat{\gamma}_T \\ T^{1/2}\hat{\alpha}_T \\ T(\hat{\phi}_T - 1) \\ T\hat{\eta}_T \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{bmatrix},$$

where $\nu_1 \sim N(0, \sigma_1^2/\sigma_2^2)$ and $(\nu_2, \nu_3, \nu_4)'$ has a nonstandard limiting distribution. Conclude that $\hat{\gamma}_T$, $\hat{\alpha}_T$, $\hat{\phi}_T$, and $\hat{\eta}_T$ are consistent estimates of 0, 0, 1, and 0, respectively, meaning that all of the estimated coefficients in [18.3.17] are consistent.

(b) Show that the $t$ test of the null hypothesis that $\gamma = 0$ is asymptotically $N(0, 1)$.

(c) Show that the $t$ test of the null hypothesis that $\delta = 0$ in the regression model of [18.3.17] is also asymptotically $N(0, 1)$.

## *Chapter 18 References*


Blough, Stephen R. 1992. "Spurious Regressions, with AR(1) Correction and Unit Root Pretest." Johns Hopkins University. Mimeo.

Chan, N. H., and C. Z. Wei. 1988. "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes." *Annals of Statistics* 16:367–401.

Granger, C. W. J., and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2:111–20.

Ohanian, Lee E. 1988. "The Spurious Effects of Unit Roots on Vector Autoregressions: A Monte Carlo Study." *Journal of Econometrics* 39:251–66.

Park, Joon Y., and Peter C. B. Phillips. 1988. "Statistical Inference in Regressions with Integrated Processes: Part 1." *Econometric Theory* 4:468–97.

——— and ———. 1989. "Statistical Inference in Regressions with Integrated Processes: Part 2." *Econometric Theory* 5:95–131.

Phillips, Peter C. B. 1986. "Understanding Spurious Regressions in Econometrics." *Journal of Econometrics* 33:311–40.

———. 1988. "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations." *Econometric Theory* 4:528–33.

——— and S. N. Durlauf. 1986. "Multiple Time Series Regression with Integrated Processes." *Review of Economic Studies* 53:473–95.

——— and Victor Solo. 1992. "Asymptotics for Linear Processes." *Annals of Statistics* 20:971–1001.

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.

Toda, H. Y., and P. C. B. Phillips. 1993a. "The Spurious Effect of Unit Roots on Exogeneity

Tests in Vector Autoregressions: An Analytical Study." *Journal of Econometrics* 59:229–55.

——— and ———. 1993b. "Vector Autoregressions and Causality." *Econometrica* forthcoming.

West, Kenneth D. 1988. "Asymptotic Normality, When Regressors Have a Unit Root." *Econometrica* 56:1397–1417.