and $X_T(r)$ defined in [17.5.4], we have

$$T^{-1}\{\xi_1^2/T + \xi_2^2/T + \cdots + \xi_{T-1}^2/T\} = \int_0^1 [\sqrt{T}\cdot X_T(r)]^2\, dr \xrightarrow{L} [\sigma\cdot\psi(1)]^2\cdot\int_0^1 [W(r)]^2\, dr,$$

by virtue of [17.5.5].

(i) As in [17.3.23],

$$\begin{aligned}
T^{-5/2}\sum_{t=1}^{T} t\xi_{t-1} &= T^{1/2}\sum_{t=1}^{T}(t/T)\cdot(\xi_{t-1}/T^2)\\
&= T^{1/2}\int_0^1 \{([Tr]^* + 1)/T\}\cdot\{(u_1 + u_2 + \cdots + u_{[Tr]^*})/T\}\, dr\\
&= T^{1/2}\int_0^1 \{([Tr]^* + 1)/T\}\cdot X_T(r)\, dr\\
&\xrightarrow{L} \sigma\cdot\psi(1)\cdot\int_0^1 rW(r)\, dr,
\end{aligned}$$

from [17.5.5] and the continuous mapping theorem.

(j) From the same argument as in (i),

$$\begin{aligned}
T^{-3}\sum_{t=1}^{T} t\xi_{t-1}^2 &= \sum_{t=1}^{T}(t/T)(\xi_{t-1}^2/T^2)\\
&= T\int_0^1 \{([Tr]^* + 1)/T\}\cdot\{(u_1 + u_2 + \cdots + u_{[Tr]^*})/T\}^2\, dr\\
&= T\int_0^1 \{([Tr]^* + 1)/T\}\cdot[X_T(r)]^2\, dr\\
&\xrightarrow{L} [\sigma\cdot\psi(1)]^2\cdot\int_0^1 r[W(r)]^2\, dr.
\end{aligned}$$

(k) This is identical to result (h) from Proposition 17.1, repeated in this proposition for the reader's convenience. ■

Chapter 17 Exercises

17.1. Let $\{u_t\}$ be an i.i.d. sequence with mean zero and variance σ^2, and let $y_t = u_1 + u_2 + \cdots + u_t$ with $y_0 = 0$. Deduce from [17.3.17] and [17.3.18] that

$$\begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-3/2}\Sigma y_{t-1}\end{bmatrix} \xrightarrow{L} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \sigma^2\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{3}\end{bmatrix}\right),$$

where Σ indicates summation over t from 1 to T. Comparing this result with Proposition 17.1, argue that

$$\begin{bmatrix} W(1) \\ \int W(r)\, dr\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{3}\end{bmatrix}\right),$$

where the integral sign denotes integration over r from 0 to 1.

17.2. *Phillips (1987) generalization of case 1.* Suppose that data are generated from the process $y_t = y_{t-1} + u_t$, where $u_t = \psi(L)\varepsilon_t$, $\Sigma_{j=0}^{\infty} j\cdot|\psi_j| < \infty$, and ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment. Consider OLS estimation of the autoregression $y_t = \rho y_{t-1} + u_t$. Let $\hat{\rho}_T = (\Sigma y_{t-1}^2)^{-1}(\Sigma y_{t-1}y_t)$ be the OLS estimate of ρ, $s_T^2 = (T-1)^{-1}\times \Sigma\hat{u}_t^2$ the OLS estimate of the variance of the regression error, $\hat{\sigma}_{\hat{\rho}_T}^2 = s_T^2\cdot(\Sigma y_{t-1}^2)^{-1}$ the OLS estimate of the variance of $\hat{\rho}_T$, and $t_T = (\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T}$ the OLS t test of $\rho = 1$, and define

$\lambda \equiv \sigma \cdot \psi(1)$. Use Proposition 17.3 to show that

(a) $$T(\hat{\rho}_T - 1) \xrightarrow{L} \frac{\tfrac{1}{2}\{\lambda^2 \cdot [W(1)]^2 - \gamma_0\}}{\lambda^2 \cdot \int [W(r)]^2\, dr};$$

(b) $$T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \xrightarrow{L} \frac{\gamma_0}{\lambda^2 \cdot \int [W(r)]^2\, dr};$$

(c) $$t_T \xrightarrow{L} (\lambda^2/\gamma_0)^{1/2} \left\{ \frac{\tfrac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int [W(r)]^2\, dr\right\}^{1/2}} + \frac{\tfrac{1}{2}(\lambda^2 - \gamma_0)}{\lambda^2 \left\{\int [W(r)]^2\, dr\right\}^{1/2}} \right\};$$

(d) $$T(\hat{\rho}_T - 1) - \tfrac{1}{2}(T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2)(\lambda^2 - \gamma_0) \xrightarrow{L} \frac{\tfrac{1}{2}\{[W(1)]^2 - 1\}}{\int [W(r)]^2\, dr};$$

(e) $$(\gamma_0/\lambda^2)^{1/2} \cdot t_T - \{\tfrac{1}{2}(\lambda^2 - \gamma_0)/\lambda\} \times \{T \cdot \hat{\sigma}_{\hat{\rho}_T} \div s_T\} \xrightarrow{L} \frac{\tfrac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int [W(r)]^2\, dr\right\}^{1/2}}.$$

Suggest estimates of γ_0 and λ^2 that could be used to construct the statistics in (d) and (e), and indicate where one could find critical values for these statistics.

17.3. *Phillips and Perron (1988) generalization of case 4.* Suppose that data are generated from the process $y_t = \alpha + y_{t-1} + u_t$, where $u_t = \psi(L)\varepsilon_t$, $\sum_{j=0}^{\infty} j \cdot |\psi_j| < \infty$, and ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment, and where α can be any value, including zero. Consider *OLS* estimation of

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

As in [17.4.49], note that the fitted values and estimate of ρ from this regression are identical to those from an *OLS* regression of y_t on a constant, time trend, and $\xi_{t-1} \equiv y_{t-1} - \alpha(t-1)$:

$$y_t = \alpha^* + \rho^* \xi_{t-1} + \delta^* t + u_t,$$

where, under the assumed data-generating process, ξ_t satisfies the assumptions of Proposition 17.3. Let $(\hat{\alpha}^*_T, \hat{\rho}^*_T, \hat{\delta}^*_T)'$ be the *OLS* estimates given by equation [17.4.50], $s_T^2 = (T-3)^{-1} \times \sum \hat{u}_t^2$ the *OLS* estimate of the variance of the regression error, $\hat{\sigma}^2_{\hat{\rho}^*_T}$ the *OLS* estimate of the variance of $\hat{\rho}^*_T$ given in [17.4.54], and $t^*_T = (\hat{\rho}^*_T - 1)/\hat{\sigma}_{\hat{\rho}^*_T}$ the *OLS* t test of $\rho = 1$. Recall further that $\hat{\rho}^*_T$, $\hat{\sigma}^2_{\hat{\rho}^*_T}$, and t^*_T are numerically identical to the analogous magnitudes for the original regression, $\hat{\rho}_T$, $\hat{\sigma}^2_{\hat{\rho}_T}$, and t_T. Finally, define $\lambda \equiv \sigma \cdot \psi(1)$. Use Proposition 17.3 to show that

(a) $$\begin{bmatrix} 1 & T^{-3/2}\sum\xi_{t-1} & T^{-2}\sum t \\ T^{-3/2}\sum\xi_{t-1} & T^{-2}\sum\xi^2_{t-1} & T^{-5/2}\sum\xi_{t-1}t \\ T^{-2}\sum t & T^{-5/2}\sum t\xi_{t-1} & T^{-3}\sum t^2 \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\, dr & 1/2 \\ \int W(r)\, dr & \int [W(r)]^2\, dr & \int rW(r)\, dr \\ 1/2 & \int rW(r)\, dr & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

(b) $$\begin{bmatrix} T^{-1/2}\sum u_t \\ T^{-1}\sum\xi_{t-1}u_t \\ T^{-3/2}\sum tu_t \end{bmatrix} \xrightarrow{L} \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} W(1) \\ \tfrac{1}{2}\{[W(1)]^2 - [\gamma_0/\lambda^2]\} \\ W(1) - \int W(r)\, dr \end{bmatrix};$$

(c) $$\begin{bmatrix} T^{1/2}\hat{\alpha}_T^* \\ T(\hat{\rho}_T^* - 1) \\ T^{3/2}(\hat{\delta}_T^* - \alpha_0) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} W(1) \\ \tfrac{1}{2}\{[W(1)]^2 - [\gamma_0/\lambda^2]\} \\ \left\{W(1) - \int W(r)\,dr\right\} \end{bmatrix};$$

(d) $$T^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \xrightarrow{p} (s_T^2/\lambda^2)[0 \quad 1 \quad 0] \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\equiv (s_T^2/\lambda^2) \cdot Q;$$

(e) $t_T \xrightarrow{p} (\lambda^2/\gamma_0)^{1/2} \cdot T(\hat{\rho}_T - 1)/\sqrt{Q}$;

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

$$\xrightarrow{L} [0 \quad 1 \quad 0] \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}$$

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

$$\equiv V;$$

(g) $(\gamma_0/\lambda^2)^{1/2} \cdot t_T - \{\tfrac{1}{2}(\lambda^2 - \gamma_0)/\lambda\} \times \{T \cdot \hat{\sigma}_{\hat{\rho}_T} \div s_T\} \xrightarrow{L} V \div \sqrt{Q}$.

Suggest estimates of γ_0 and λ^2 that could be used to construct the statistics in (f) and (g), and indicate where one could find critical values for these statistics.

17.4. *Generalization of case 1 for autoregression.* Consider *OLS* estimation of

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

where ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment and the roots of $(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1}z^{p-1}) = 0$ are outside the unit circle. Define $\lambda \equiv \sigma/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})$ and $\gamma_j = E\{(\Delta y_t)(\Delta y_{t-j})\}$. Let $\hat{\boldsymbol{\zeta}}_T \equiv (\hat{\zeta}_{1,T}, \hat{\zeta}_{2,T}, \ldots, \hat{\zeta}_{p-1,T})'$ be the $(p - 1) \times 1$ vector of estimated *OLS* coefficients on the lagged changes in y, and let $\boldsymbol{\zeta}$ be the corresponding true value. Show that if the true value of ρ is unity, then

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \\ T(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0}' & \lambda^2 \cdot \int [W(r)]^2\,dr \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_1 \\ \tfrac{1}{2}\sigma\lambda\{[W(1)]^2 - 1\} \end{bmatrix},$$

where $\mathbf{V}$ is the $[(p - 1) \times (p - 1)]$ matrix defined in [17.7.19] and $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V})$. Deduce from this that

(a) $T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{V}^{-1})$;

(b) $$T(\hat{\rho}_T - 1)/(1 - \hat{\zeta}_{1,T} - \hat{\zeta}_{2,T} - \cdots - \hat{\zeta}_{p-1,T}) \xrightarrow{L} \frac{\tfrac{1}{2}\{[W(1)]^2 - 1\}}{\int [W(r)]^2\,dr};$$

(c) $(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T} \xrightarrow{L} \dfrac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\left\{\int [W(r)]^2\, dr\right\}^{1/2}}$.

Where could you find critical values for the statistics in (b) and (c)?

17.5. *Generalization of case 3 for autoregression.* Consider *OLS* estimation of

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

where ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment and the roots of $(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1}z^{p-1}) = 0$ are outside the unit circle.

(a) Show that the fitted values for this regression are identical to those for the following transformed specification:

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

where $u_t \equiv \Delta y_t - \mu$ and $\mu \equiv \alpha/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})$.

(b) Suppose that the true value of ρ is 1 and the true value of α is nonzero. Show that under these assumptions,

$$u_t = [1/(1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_{p-1}L^{p-1})]\varepsilon_t$$
$$y_{t-1} = \mu(t-1) + \xi_{t-1},$$

where

$$\xi_{t-1} \equiv y_0 + u_1 + u_2 + \cdots + u_{t-1}.$$

Conclude that for fixed y_0, the variables u_t and ξ_t satisfy the assumptions of Proposition 17.3 and that y_t is dominated asymptotically by a time trend.

(c) Let $\gamma_j \equiv E(u_t u_{t-j})$, and let $\hat{\boldsymbol{\zeta}}_T \equiv (\hat{\zeta}_{1.T}, \hat{\zeta}_{2.T}, \ldots, \hat{\zeta}_{p-1.T})'$ be the $(p-1) \times 1$ vector of estimated *OLS* coefficients on $(u_{t-1}, u_{t-2}, \ldots, u_{t-p+1})$; these, of course, are identical to the coefficients on $(\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1})$ in the original regression. Show that if $\rho = 1$ and $\alpha \neq 0$,

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\zeta}}_T - \boldsymbol{\zeta}) \\ T^{1/2}(\hat{\mu}_T - \mu) \\ T^{3/2}(\hat{\rho}_T - 1) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} & 0 \\ \mathbf{0}' & 1 & \mu/2 \\ \mathbf{0}' & \mu/2 & \mu^2/3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_1 \\ h_2 \\ h_3 \end{bmatrix},$$

where

$$\begin{bmatrix} \mathbf{h}_1 \\ h_2 \\ h_3 \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{V} & \mathbf{0} & 0 \\ \mathbf{0}' & 1 & \mu/2 \\ \mathbf{0}' & \mu/2 & \mu^2/3 \end{bmatrix} \right)$$

and $\mathbf{V}$ is the matrix in [17.7.19]. Conclude as in the analysis of Section 16.3 that any *OLS* t or F test on the original regression can be compared with the standard t and F tables to give an asymptotically valid inference.

17.6. *Generalization of case 4 for autoregression.* Consider *OLS* estimation of

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

where ε_t is i.i.d. with mean zero, variance σ^2, and finite fourth moment and then roots of $(1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1}z^{p-1}) = 0$ are outside the unit circle.

(a) Show that the fitted values of this regression are numerically identical to those of the following specification:

$$y_t = \zeta_1 u_{t-1} + \zeta_2 u_{t-2} + \cdots + \zeta_{p-1}u_{t-p+1} + \mu^* + \rho \xi_{t-1} + \delta^* t + \varepsilon_t,$$

where $u_t \equiv \Delta y_t - \mu$, $\mu \equiv \alpha/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})$, $\mu^* \equiv (1-\rho)\mu$, $\xi_{t-1} \equiv y_{t-1} - \mu(t-1)$, and $\delta^* \equiv \delta + \rho\mu$. Note that the estimated coefficients $\hat{\boldsymbol{\zeta}}_T$ and $\hat{\rho}_T$ and their standard errors will be identical for the two regressions.

(b) Suppose that the true value of ρ is 1 and the true value of δ is 0. Show that under these assumptions,

$$u_t = [1/(1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_{p-1}L^{p-1})]\varepsilon_t$$
$$\xi_{t-1} = y_0 + u_1 + u_2 + \cdots + u_{t-1}.$$

Conclude that for fixed y_0, the variables u_t and ξ_t satisfy the assumptions of Proposition 17.3.

(c) Again let $\rho = 1$ and $\delta = 0$, and define $\gamma_j \equiv E(u_t u_{t-j})$ and

$$\lambda \equiv \sigma/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1}).$$

Show that

$$\begin{bmatrix} T^{1/2}(\hat{\zeta}_T - \zeta) \\ T^{1/2}\hat{\mu}_T^* \\ T(\hat{\rho}_T - 1) \\ T^{3/2}(\hat{\delta}_T^* - \delta^*) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{V} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \lambda\cdot\int W(r)\,dr & 1/2 \\ \mathbf{0}' & \lambda\cdot\int W(r)\,dr & \lambda^2\cdot\int [W(r)]^2\,dr & \lambda\cdot\int rW(r)\,dr \\ \mathbf{0}' & 1/2 & \lambda\cdot\int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \mathbf{h}_1 \\ \sigma\cdot W(1) \\ \tfrac{1}{2}\sigma\lambda\{[W(1)]^2 - 1\} \\ \sigma\cdot\left\{W(1) - \int W(r)\,dr\right\} \end{bmatrix}$$

where $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma^2\mathbf{V})$ and $\mathbf{V}$ is as defined in [17.7.19].

(d) Deduce from answer (c) that

$$T^{1/2}(\hat{\zeta}_T - \zeta) \xrightarrow{L} N(\mathbf{0}, \sigma^2\mathbf{V}^{-1});$$

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

$$\xrightarrow{L} [0 \quad 1 \quad 0]\begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}\begin{bmatrix} W(1) \\ \tfrac{1}{2}\{[W(1)]^2 - 1\} \\ W(1) - \int W(r)\,dr \end{bmatrix}$$

$$\equiv V;$$

$$(\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T} \xrightarrow{L} V \div \sqrt{Q},$$

where

$$Q \equiv [0 \quad 1 \quad 0]\begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Notice that the distribution of V is the same as the asymptotic distribution of the variable tabulated for case 4 in Table B.5, while the distribution of $V/\sqrt{Q}$ is the same as the asymptotic distribution of the variable tabulated for case 4 in Table B.6.

Chapter 17 References

Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59:817–58.

———. 1993. "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models." *Econometrica* 61:139–65.

Beveridge, Stephen, and Charles R. Nelson. 1981. "A New Approcah to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle.'" *Journal of Monetary Economics* 7:151–74.

Bhargava, Alok. 1986. "On the Theory of Testing for Unit Roots in Observed Time Series." *Review of Economic Studies* 53:369–84.

Billingsley, Patrick. 1968. *Convergence of Probability Measures*. New York: Wiley.

Campbell, John Y., and Pierre Perron. 1991. "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots." *NBER Macroeconomics Annual*. Cambridge, Mass.: MIT Press.

Cecchetti, Stephen G., and Pok-sang Lam. 1991. "What Do We Learn from Variance Ratio Statistics? A Study of Stationary and Nonstationary Models with Breaking Trends." Department of Economics, Ohio State University. Mimeo.

Chan, N. H., and C. Z. Wei. 1987. "Asymptotic Inference for Nearly Nonstationary AR(1) Processes." *Annals of Statistics* 15:1050–63.

——— and ———. 1988. "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes." *Annals of Statistics* 16:367–401.

Cochrane, John H. 1988. "How Big Is the Random Walk in GNP?" *Journal of Political Economy* 96:893–920.

DeJong, David N., and Charles H. Whiteman. 1991. "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series.'" *Journal of Monetary Economics* 28:221–54.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74:427–31.

——— and ———. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49:1057–72.

——— and S. G. Pantula. 1987. "Determining the Order of Differencing in Autoregressive Processes." *Journal of Business and Economic Statistics* 5:455–61.

Evans, G. B. A., and N. E. Savin. 1981. "Testing for Unit Roots: 1." *Econometrica* 49:753–79.

——— and ———. 1984. "Testing for Unit Roots: 2." *Econometrica* 52:1241–69.

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

Hall, Alastair. 1989. "Testing for a Unit Root in the Presence of Moving Average Errors." *Biometrika* 76:49–56.

———. 1991. "Testing for a Unit Root in Time Series with Pretest Data Based Model Selection." Department of Economics, North Carolina State University. Mimeo.

Hall, P., and C. C. Heyde. 1980. *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hansen, Bruce E. 1992. "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes." *Econometrica* 60:967–72.

Jeffreys, H. 1946. "An Invariant Form for the Prior Probability in Estimation Problems." *Proceedings of the Royal Society of London* Series A, 186:453–61.

Kim, Kiwhan, and Peter Schmidt. 1990. "Some Evidence on the Accuracy of Phillips-Perron Tests Using Alternative Estimates of Nuisance Parameters." *Economics Letters* 34:345–50.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54:159–78.

Lo, Andrew W., and A. Craig MacKinlay. 1988. "Stock Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." *Review of Financial Studies* 1:41–66.

——— and ———. 1989. "The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation." *Journal of Econometrics* 40:203–38.

Malliaris, A. G., and W. A. Brock. 1982. *Stochastic Methods in Economics and Finance*. Amsterdam: North-Holland.

Pantula, Sastry G., and Alastair Hall. 1991. "Testing for Unit Roots in Autoregressive Moving Average Models: An Instrumental Variable Approach." *Journal of Econometrics* 48:325–53.

Park, Joon Y., and B. Choi. 1988. "A New Approach to Testing for a Unit Root." Cornell University. Mimeo.

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

——— 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* $0 \le r_1 < r_2 < \cdots < r_k \le 1$, *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

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

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

$$\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'; \tag{18.1.2}$$

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

$$\boldsymbol{\varepsilon}_t = \mathbf{P}\mathbf{v}_t, \tag{18.1.3}$$

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

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

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

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

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

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

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).

$-(\Psi_{s+1} + \Psi_{s+2} + \Psi_{s+3} + \cdots)$, and $\{\alpha_s\}_{s=0}^{\infty}$ 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 $\boldsymbol{\Omega} = \mathbf{PP}'$, 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

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

then it follows from [18.1.6] that

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

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

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

where $\Psi(1)\cdot\mathbf{P}\cdot\mathbf{W}(r)$ is distributed $N(\mathbf{0}, r[\Psi(1)]\cdot\boldsymbol{\Omega}\cdot[\Psi(1)]')$ 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

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

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 ξ_{it}^*, v_{it}, and $W_i(r)$ denote the ith 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

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

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

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

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

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

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 $\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'$ denote the Cholesky factorization of $\boldsymbol{\Omega}$, and define

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

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

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

[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).

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

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

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

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

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

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

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

(d) $$T^{-1} \sum_{t=1}^{T} (\boldsymbol{\xi}_{t-1}\mathbf{u}_{t-s}' + \mathbf{u}_{t-s}\boldsymbol{\xi}_{t-1}')$$
$$\xrightarrow{L} \begin{cases} \mathbf{\Lambda}\cdot[\mathbf{W}(1)]\cdot[\mathbf{W}(1)]'\cdot\mathbf{\Lambda}' - \mathbf{\Gamma}_0 & \textit{for } s = 0 \\ \mathbf{\Lambda}\cdot[\mathbf{W}(1)]\cdot[\mathbf{W}(1)]'\cdot\mathbf{\Lambda}' + \sum_{\nu=-s+1}^{s-1} \mathbf{\Gamma}_\nu & \textit{for } s = 1, 2, \ldots; \end{cases}$$

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

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

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

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

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

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

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

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

18.2. Vector Autoregressions Containing Unit Roots

Suppose that a vector y_t could be described by a vector autoregression in the differences Δy_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 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 ith 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 ith row of $\boldsymbol{\zeta}_s$ for $s = 1, 2, \ldots, p-1$. Similarly, $\boldsymbol{\rho}_i'$ denotes the ith row of $\boldsymbol{\rho}$. Under the null hypothesis [18.2.7], $\boldsymbol{\rho}_i' = \mathbf{e}_i'$, where $\mathbf{e}_i'$ is the ith 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 Σ denotes summation over $t = 1$ through T. In the case of *OLS* estimation of [18.2.13],

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

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

$$= \begin{bmatrix} \Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{u}_{t-1} & \Sigma \mathbf{u}_{t-1}\mathbf{y}_{t-1}' \\ \Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{u}_{t-2} & \Sigma \mathbf{u}_{t-2}\mathbf{y}_{t-1}' \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{u}_{t-p+1} & \Sigma \mathbf{u}_{t-p+1}\mathbf{y}_{t-1}' \\ \Sigma \mathbf{u}_{t-1}' & \Sigma \mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{u}_{t-p+1}' & T & \Sigma \mathbf{y}_{t-1}' \\ \Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-1}' & \Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-2}' & \cdots & \Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-p+1}' & \Sigma \mathbf{y}_{t-1} & \Sigma \mathbf{y}_{t-1}\mathbf{y}_{t-1}' \end{bmatrix} \qquad [18.2.16]$$

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

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

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

$$\begin{matrix} T^{-1}\Sigma \mathbf{u}_{t-1}\mathbf{u}_{t-p+1}' & T^{-1}\Sigma \mathbf{u}_{t-1} & T^{-3/2}\Sigma \mathbf{u}_{t-1}\mathbf{y}_{t-1}' \\ T^{-1}\Sigma \mathbf{u}_{t-2}\mathbf{u}_{t-p+1}' & T^{-1}\Sigma \mathbf{u}_{t-2} & T^{-3/2}\Sigma \mathbf{u}_{t-2}\mathbf{y}_{t-1}' \\ \vdots & \vdots & \vdots \\ T^{-1}\Sigma \mathbf{u}_{t-p+1}\mathbf{u}_{t-p+1}' & T^{-1}\Sigma \mathbf{u}_{t-p+1} & T^{-3/2}\Sigma \mathbf{u}_{t-p+1}\mathbf{y}_{t-1}' \\ T^{-1}\Sigma \mathbf{u}_{t-p+1}' & 1 & T^{-3/2}\Sigma \mathbf{y}_{t-1}' \\ T^{-3/2}\Sigma \mathbf{y}_{t-1}\mathbf{u}_{t-p+1}' & T^{-3/2}\Sigma \mathbf{y}_{t-1} & T^{-2}\Sigma \mathbf{y}_{t-1}\mathbf{y}_{t-1}' \end{matrix} \Bigg]$$

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

where

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

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

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

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

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

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

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

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

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

for $\mathbf{e}_i$ the ith 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:

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

This means that the Wald form of the *OLS* χ^2 test of any linear hypothesis that involves only the coefficients on $\Delta\mathbf{y}_{t-s}$ has the usual asymptotic χ^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:

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

Thus, as in the case of a univariate autoregression, if the goal is to estimate the parameters $\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{i,p-1}$ 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,

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

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 $\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{i,p-1}$ will be dominated asymptotically by the coefficients with the slower rate of convergence, namely, $\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{i,p-1}$, 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

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

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

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

Since $\sqrt{T}(\hat{\zeta}_s - \zeta_s)$ 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 ith 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 ith row of $\boldsymbol{\Phi}_s$. Consider the null hypothesis

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

The Wald form of the OLS χ^2 test of this hypothesis will be numerically identical to the test of

$$H_0\colon \zeta_{i,p_0} = \zeta_{i,p_0+1} = \cdots = \zeta_{i,p-1} = \mathbf{0} \quad [18.2.34]$$

for OLS estimation of

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

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 χ^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 χ^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 nth 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 $\mathbf{y}_t^* \equiv (y_{1t}^*, y_{2t}^*, \ldots, y_{n-1,t}^*)'$ 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 ith 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 ith row of $\boldsymbol{\zeta}_s$ and $\boldsymbol{\rho}_i^{*\prime}$ is the ith 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,

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

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

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

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 χ^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 χ^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 Σ 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

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

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

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

Let $\mathbf{L}_{22}$ denote the Cholesky factor of Σ_{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

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

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

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

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 ε_{1t} is i.i.d. with mean zero and variance σ_1^2, ε_{2t} is i.i.d. with mean zero and variance σ_2^2, and ε_{1t} is independent of $\varepsilon_{2\tau}$ for all t and τ. 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 ξ_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], Δy_{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 σ_1^*, $\mathbf{\Sigma}_{21}$, and $\mathbf{\Sigma}_{22}$, 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 $\mathbf{\Lambda}\cdot\mathbf{\Lambda}'$ 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 $\mathbf{\Lambda}\cdot\mathbf{\Lambda}'$ would just be the scalar $[\psi(1)\cdot\sigma_1]^2$ and the condition that $\mathbf{\Lambda}\cdot\mathbf{\Lambda}'$ 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 Δy_{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 $\mathbf{\Lambda}\cdot\mathbf{\Lambda}'$ 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 $\mathbf{\Lambda}\cdot\mathbf{\Lambda}'$ 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 γ and δ 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 χ^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 ϕ 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 $\xi_t \equiv \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_t$, 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

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

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],

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

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

(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:

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

$$\mathbf{y}_{2t}^* \equiv \mathbf{L}'_{22}\mathbf{y}_{2t}. \quad [18.A.10]$$

Note that the inverses $\boldsymbol{\Sigma}_{22}^{-1}$, $(\sigma_1^*)^{-1}$, and $\mathbf{L}_{22}^{-1}$ all exist, since $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ is symmetric positive definite. An *OLS* regression of y_{1t}^* on a constant and $\mathbf{y}_{2t}^*$,

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

would yield estimates

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

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

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

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

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

implying that

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

The usefulness of this transformation is as follows. Notice that

$$\begin{bmatrix} y_{1t}^*/\sigma_1^*\\ \mathbf{y}_{2t}^*\end{bmatrix} = \begin{bmatrix}(1/\sigma_1^*) & (-1/\sigma_1^*)\cdot\boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}_{22}^{-1}\\ \mathbf{0} & \mathbf{L}'_{22}\end{bmatrix}\begin{bmatrix} y_{1t}\\ \mathbf{y}_{2t}\end{bmatrix} \equiv \mathbf{L}'\mathbf{y}_t$$

for

$$\mathbf{L}' \equiv \begin{bmatrix}(1/\sigma_1^*) & (-1/\sigma_1^*)\cdot\boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}_{22}^{-1}\\ \mathbf{0} & \mathbf{L}'_{22}\end{bmatrix}.$$

Moreover,

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

But [18.3.7] implies that

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

from which

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

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

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

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

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

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

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

it follows that

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

Similarly, result (i) of Proposition 18.1 gives

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

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

Proof of (a). If [18.A.12] is divided by σ_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

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

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

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

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

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

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

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

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

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

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

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

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],

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

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

$$T^{-1}\cdot F_T \xrightarrow{L} \{\sigma_1^*\cdot\mathbf{R}^*\mathbf{h}_2 - \mathbf{r}^*\}' \times \Bigg\{(\sigma_1^*)^2\cdot H[\mathbf{0} \quad \mathbf{R}^*]$$

$$\times \begin{bmatrix} 1 & \int[\mathbf{W}_2^*(r)]'\,dr \\ \int \mathbf{W}_2^*(r)\,dr & \int[\mathbf{W}_2^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}' \\ \mathbf{R}^{*\prime} \end{bmatrix}\Bigg\}^{-1} \{\sigma_1^*\cdot\mathbf{R}^*\mathbf{h}_2 - \mathbf{r}^*\} \div m. \qquad \blacksquare$$

Chapter 18 Exercises

18.1. Consider OLS estimation of

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

where y_{it} is the ith element of the $(n \times 1)$ vector $\mathbf{y}_t$ and ε_{it} is the ith 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 Σ 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 χ_T^2 be the Wald form of the OLS χ^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 ith row of $\mathbf{I}_n$), show that $\chi_T^2 \xrightarrow{L} \chi^2(m)$.

18.2. Suppose that the regression model

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

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* χ^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

$$\begin{bmatrix} T^{1/2}\hat{\gamma}_T \\ T^{1/2}\hat{\alpha}_T \\ T(\hat{\phi}_T - 1) \\ T\hat{\eta}_T \end{bmatrix} \xrightarrow{L} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix},$$

where $v_1 \sim N(0, \sigma_1^2/\sigma_2^2)$ and $(v_2, v_3, v_4)'$ 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.

19 Cointegration

This chapter discusses a particular class of vector unit root processes known as *cointegrated* processes. Such specifications were implicit in the "error-correction" models advocated by Davidson, Hendry, Srba, and Yeo (1978). However, a formal development of the key concepts did not come until the work of Granger (1983) and Engle and Granger (1987).

Section 19.1 introduces the concept of cointegration and develops several alternative representations of a cointegrated system. Section 19.2 discusses tests of whether a vector process is cointegrated. These tests are summarized in Table 19.1. Single-equation methods for estimating a cointegrating vector and testing a hypothesis about its value are presented in Section 19.3. Full-information maximum likelihood estimation is discussed in Chapter 20.

19.1. *Introduction*

Description of Cointegration

An $(n \times 1)$ vector time series $\mathbf{y}_t$ is said to be *cointegrated* if each of the series taken individually is $I(1)$, that is, nonstationary with a unit root, while some linear combination of the series $\mathbf{a}'\mathbf{y}_t$ is stationary, or $I(0)$, for some nonzero $(n \times 1)$ vector $\mathbf{a}$. A simple example of a cointegrated vector process is the following bivariate system:

$$y_{1t} = \gamma y_{2t} + u_{1t} \qquad [19.1.1]$$

$$y_{2t} = y_{2,t-1} + u_{2t}, \qquad [19.1.2]$$

with u_{1t} and u_{2t} uncorrelated white noise processes. The univariate representation for y_{2t} is a random walk,

$$\Delta y_{2t} = u_{2t}, \qquad [19.1.3]$$

while differencing [19.1.1] results in

$$\Delta y_{1t} = \gamma \Delta y_{2t} + \Delta u_{1t} = \gamma u_{2t} + u_{1t} - u_{1,t-1}. \qquad [19.1.4]$$

Recall from Section 4.7 that the right side of [19.1.4] has an $MA(1)$ representation:

$$\Delta y_{1t} = v_t + \theta v_{t-1}, \qquad [19.1.5]$$

where v_t is a white noise process and $\theta \neq -1$ as long as $\gamma \neq 0$ and $E(u_{2t}^2) > 0$. Thus, both y_{1t} and y_{2t} are $I(1)$ processes, though the linear combination

$(y_{1t} - \gamma y_{2t})$ is stationary. Hence, we would say that $\mathbf{y}_t = (y_{1t}, y_{2t})'$ is cointegrated with $\mathbf{a}' = (1, -\gamma)$.

Figure 19.1 plots a sample realization of [19.1.1] and [19.1.2] for $\gamma = 1$ and u_{1t} and u_{2t} independent $N(0, 1)$ variables. Note that either series (y_{1t} or y_{2t}) will wander arbitrarily far from the starting value, though y_{1t} should remain within a fixed distance of γy_{2t}, with this distance determined by the standard deviation of u_{1t}.

Cointegration means that although many developments can cause permanent changes in the individual elements of $\mathbf{y}_t$, there is some long-run equilibrium relation tying the individual components together, represented by the linear combination $\mathbf{a}'\mathbf{y}_t$. An example of such a system is the model of consumption spending proposed by Davidson, Hendry, Srba, and Yeo (1978). Their results suggest that although both consumption and income exhibit a unit root, over the long run consumption tends to be a roughly constant proportion of income, so that the difference between the log of consumption and the log of income appears to be a stationary process.

Another example of an economic hypothesis that lends itself naturally to a cointegration interpretation is the theory of purchasing power parity. This theory holds that, apart from transportation costs, goods should sell for the same effective price in two countries. Let P_t denote an index of the price level in the United States (in dollars per good), P_t^* a price index for Italy (in lire per good), and S_t the rate of exchange between the currencies (in dollars per lira). Then purchasing power parity holds that

$$P_t = S_t P_t^*,$$

or, taking logarithms,

$$p_t = s_t + p_t^*,$$

where $p_t \equiv \log P_t$, $s_t \equiv \log S_t$, and $p_t^* \equiv \log P_t^*$. In practice, errors in measuring prices, transportation costs, and differences in quality prevent purchasing power parity from holding exactly at every date t. A weaker version of the hypothesis is that the variable z_t defined by

$$z_t \equiv p_t - s_t - p_t^* \qquad [19.1.6]$$

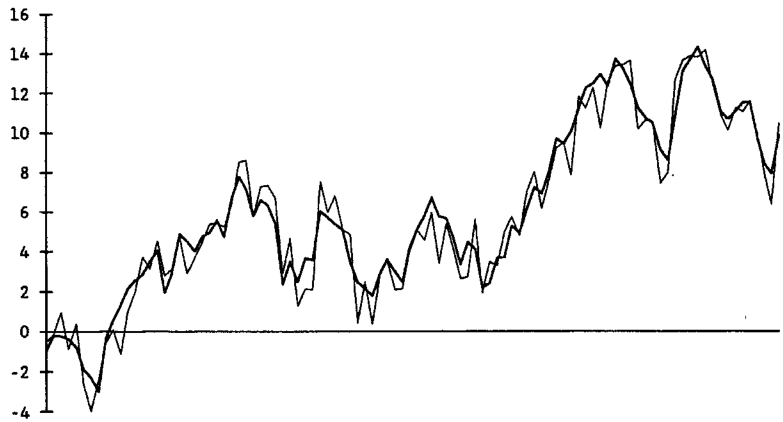


FIGURE 19.1 Sample realization of cointegrated series.

is stationary, even though the individual elements (p_t, s_t, or p_t^*) are all $I(1)$. Empirical tests of this version of the puchasing power parity hypothesis have been explored by Baillie and Selover (1987) and Corbae and Ouliaris (1988).

Many other interesting applications of the idea of cointegration have been investigated. Kremers (1989) suggested that governments are forced politically to maintain their debt at a roughly constant multiple of GNP, so that log(debt) − log(GNP) is stationary even though each component individually is not. Campbell and Shiller (1988a, b) noted that if y_{2t} is $I(1)$ and y_{1t} is a rational forecast of future values of y_2, then y_1 and y_2 will be cointegrated. Other interesting applications include King, Plosser, Stock, and Watson (1991), Ogaki (1992), Ogaki and Park (1992), and Clarida (1991).

It was asserted in the previous chapter that if $\mathbf{y}_t$ is cointegrated, then it is not correct to fit a vector autoregression to the differenced data. We now verify this claim for the particular example of [19.1.1] and [19.1.2]. The issues will then be discussed in terms of a general cointegrated system involving n different variables.

Discussion of the Example of [19.1.1] and [19.1.2]

Returning to the example in [19.1.1] and [19.1.2], notice that $\varepsilon_{2t} \equiv u_{2t}$ is the error in forecasting y_{2t} on the basis of lagged values of y_1 and y_2 while $\varepsilon_{1t} \equiv \gamma u_{2t} + u_{1t}$ is the error in forecasting y_{1t}. The right side of [19.1.4] can be written

$$(\gamma u_{2t} + u_{1t}) - u_{1,t-1} = \varepsilon_{1t} - (\varepsilon_{1,t-1} - \gamma\varepsilon_{2,t-1}) = (1 - L)\varepsilon_{1t} + \gamma L\varepsilon_{2t}.$$

Substituting this into [19.1.4] and stacking it in a vector system along with [19.1.3] produces the vector moving average representation for $(\Delta y_{1t}, \Delta y_{2t})'$,

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \Psi(L)\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \qquad [19.1.7]$$

where

$$\Psi(L) \equiv \begin{bmatrix} 1 - L & \gamma L \\ 0 & 1 \end{bmatrix}. \qquad [19.1.8]$$

A *VAR* for the differenced data, if it existed, would take the form

$$\Phi(L)\Delta\mathbf{y}_t = \boldsymbol{\varepsilon}_t,$$

where $\Phi(L) = [\Psi(L)]^{-1}$. But the matrix polynomial associated with the moving average operator for this process, $\Psi(z)$, has a root at unity,

$$|\Psi(1)| = \begin{vmatrix} (1 - 1) & \gamma \\ 0 & 1 \end{vmatrix} = 0.$$

Hence the matrix moving average operator is noninvertible, and no finite-order vector autoregression could describe $\Delta\mathbf{y}_t$.

The reason a finite-order *VAR* in differences affords a poor approximation to the cointegrated system of [19.1.1] and [19.1.2] is that the *level* of y_2 contains information that is useful for forecasting y_1 beyond that contained in a finite number of lagged *changes* in y_2 alone.

If we are willing to modify the *VAR* by including lagged levels along with lagged changes, a stationary representation similar to a *VAR* for $\Delta\mathbf{y}_t$ is easy to find. Recalling that $u_{1,t-1} = y_{1,t-1} - \gamma y_{2,t-1}$, notice that [19.1.4] and [19.1.3] can be written as

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma u_{2t} + u_{1t} \\ u_{2t} \end{bmatrix}. \qquad [19.1.9]$$

The general principle of which [19.1.9] provides an illustration is that with a cointegrated system, one should include lagged levels along with lagged differences in a vector autoregression explaining $\Delta \mathbf{y}_t$. The lagged levels will appear in the form of those linear combinations of $\mathbf{y}$ that are stationary.

General Characterization of the Cointegrating Vector

Recall that an $(n \times 1)$ vector $\mathbf{y}_t$ is said to be cointegrated if each of its elements individually is $I(1)$ and if there exists a nonzero $(n \times 1)$ vector $\mathbf{a}$ such that $\mathbf{a}'\mathbf{y}_t$ is stationary. When this is the case, $\mathbf{a}$ is called a *cointegrating vector*.

Clearly, the cointegrating vector $\mathbf{a}$ is not unique, for if $\mathbf{a}'\mathbf{y}_t$ is stationary, then so is $b\mathbf{a}'\mathbf{y}_t$ for any nonzero scalar b; if $\mathbf{a}$ is a cointegrating vector, then so is $b\mathbf{a}$. In speaking of the value of the cointegrating vector, an arbitrary normalization must be made, such as that the first element of $\mathbf{a}$ is unity.

If there are more than two variables contained in $\mathbf{y}_t$, then there may be two nonzero $(n \times 1)$ vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ such that $\mathbf{a}_1'\mathbf{y}_t$ and $\mathbf{a}_2'\mathbf{y}_t$ are both stationary, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent (that is, there does not exist a scalar b such that $\mathbf{a}_2 = b\mathbf{a}_1$). Indeed, there may be $h < n$ linearly independent $(n \times 1)$ vectors $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h)$ such that $\mathbf{A}'\mathbf{y}_t$ is a stationary $(h \times 1)$ vector, where $\mathbf{A}'$ is the following $(h \times n)$ matrix:[1]

$$\mathbf{A}' \equiv \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_h' \end{bmatrix}. \qquad [19.1.10]$$

Again, the vectors $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h)$ are not unique; if $\mathbf{A}'\mathbf{y}_t$ is stationary, then for any nonzero $(1 \times h)$ vector $\mathbf{b}'$, the scalar $\mathbf{b}'\mathbf{A}'\mathbf{y}_t$ is also stationary. Then the $(n \times 1)$ vector $\boldsymbol{\pi}$ given by $\boldsymbol{\pi}' = \mathbf{b}'\mathbf{A}'$ could also be described as a cointegrating vector.

Suppose that there exists an $(h \times n)$ matrix $\mathbf{A}'$ whose rows are linearly independent such that $\mathbf{A}'\mathbf{y}_t$ is a stationary $(h \times 1)$ vector. Suppose further that if $\mathbf{c}'$ is any $(1 \times n)$ vector that is linearly independent of the rows of $\mathbf{A}'$, then $\mathbf{c}'\mathbf{y}_t$ is a nonstationary scalar. Then we say that there are exactly h cointegrating relations among the elements of $\mathbf{y}_t$ and that $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h)$ form a *basis* for the space of cointegrating vectors.

Implications of Cointegration for the Vector Moving Average Representation

We now discuss the general implications of cointegration for the moving average and vector autoregressive representations of a vector system.[2] Since it is assumed that $\Delta \mathbf{y}_t$ is stationary, let $\boldsymbol{\delta} \equiv E(\Delta \mathbf{y}_t)$ and define

$$\mathbf{u}_t \equiv \Delta \mathbf{y}_t - \boldsymbol{\delta}. \qquad [19.1.11]$$

Suppose that $\mathbf{u}_t$ has the Wold representation

$$\mathbf{u}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t,$$

[1]If $h = n$ such linearly independent vectors existed, then $\mathbf{y}_t$ would itself be $I(0)$. This claim will become apparent in the triangular representation of a cointegrated system developed in [19.1.20] and [19.1.21].

[2]These results were first derived by Engle and Granger (1987).

where $E(\varepsilon_t) = \mathbf{0}$ and

$$E(\varepsilon_t \varepsilon'_\tau) = \begin{cases} \mathbf{\Omega} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Let $\mathbf{\Psi}(1)$ denote the $(n \times n)$ matrix polynomial $\mathbf{\Psi}(z)$ evaluated at $z = 1$; that is,

$$\mathbf{\Psi}(1) \equiv \mathbf{I}_n + \mathbf{\Psi}_1 + \mathbf{\Psi}_2 + \mathbf{\Psi}_3 + \cdots .$$

We first claim that if $\mathbf{A}'\mathbf{y}_t$ is stationary, then

$$\mathbf{A}'\mathbf{\Psi}(1) = \mathbf{0}. \quad [19.1.12]$$

To verify this claim, note that as long as $\{s \cdot \mathbf{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable, the difference equation [19.1.11] implies that

$$\begin{aligned} \mathbf{y}_t &= \mathbf{y}_0 + \boldsymbol{\delta} \cdot t + \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_t \\ &= \mathbf{y}_0 + \boldsymbol{\delta} \cdot t + \mathbf{\Psi}(1) \cdot (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + \boldsymbol{\eta}_t - \boldsymbol{\eta}_0, \end{aligned} \quad [19.1.13]$$

where the last line follows from [18.1.6] for $\boldsymbol{\eta}_t$ a stationary process. Premultiplying [19.1.13] by $\mathbf{A}'$ results in

$$\mathbf{A}'\mathbf{y}_t = \mathbf{A}'(\mathbf{y}_0 - \boldsymbol{\eta}_0) + \mathbf{A}'\boldsymbol{\delta} \cdot t + \mathbf{A}'\mathbf{\Psi}(1) \cdot (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + \mathbf{A}'\boldsymbol{\eta}_t. \quad [19.1.14]$$

If $E(\varepsilon_t \varepsilon'_t)$ is nonsingular, then $\mathbf{c}'(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t)$ is $I(1)$ for every nonzero $(n \times 1)$ vector $\mathbf{c}$. However, in order for $\mathbf{y}_t$ to be cointegrated with cointegrating vectors given by the rows of $\mathbf{A}'$, expression [19.1.14] is required to be stationary. This could occur only if $\mathbf{A}'\mathbf{\Psi}(1) = \mathbf{0}$. Thus, [19.1.12] is a necessary condition for cointegration, as claimed.

As emphasized by Engle and Yoo (1987) and Ogaki and Park (1992), condition [19.1.12] is not by itself sufficient to ensure that $\mathbf{A}'\mathbf{y}_t$ is stationary. From [19.1.14], stationarity further requires that

$$\mathbf{A}'\boldsymbol{\delta} = \mathbf{0}. \quad [19.1.15]$$

If some of the series exhibit nonzero drift ($\boldsymbol{\delta} \neq \mathbf{0}$), then unless the drift across series satisfies the restriction of [19.1.15], the linear combination $\mathbf{A}'\mathbf{y}_t$ will grow deterministically at rate $\mathbf{A}'\boldsymbol{\delta}$. Thus, if the underlying hypothesis suggesting the possibility of cointegration is that certain linear combinations of $\mathbf{y}_t$ are stable, this requires that both [19.1.12] and [19.1.15] hold.

Note that [19.1.12] implies that certain linear combinations of the rows of $\mathbf{\Psi}(1)$, such as $\mathbf{a}'_1\mathbf{\Psi}(1)$, are zero, meaning that the determinant $|\mathbf{\Psi}(z)| = 0$ at $z = 1$. This in turn means that the matrix operator $\mathbf{\Psi}(L)$ is noninvertible. Thus, a cointegrated system can never be represented by a finite-order vector autoregression in the differenced data $\Delta\mathbf{y}_t$.

For the example of [19.1.1] and [19.1.2], we saw in [19.1.7] and [19.1.8] that

$$\mathbf{\Psi}(z) = \begin{bmatrix} 1 - z & \gamma z \\ 0 & 1 \end{bmatrix}$$

and

$$\mathbf{\Psi}(1) = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix}.$$

This is a singular matrix with $\mathbf{A}'\mathbf{\Psi}(1) = \mathbf{0}$ for $\mathbf{A}' = [1 \quad -\gamma]$.

Phillips's Triangular Representation

Another convenient representation for a cointegrated system was introduced by Phillips (1991). Suppose that the rows of the $(h \times n)$ matrix $\mathbf{A}'$ form a basis for the space of cointegrating vectors. If the (1, 1) element of $\mathbf{A}'$ is nonzero, we can conveniently normalize it to unity. If, instead, the (1, 1) element of $\mathbf{A}'$ is zero, we can reorder the elements of $\mathbf{y}_t$ so that y_{1t} is included in the first cointegrating relation. Hence, without loss of generality, we take

$$\mathbf{A}' = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_h' \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{h1} & a_{h2} & a_{h3} & \cdots & a_{hn} \end{bmatrix}.$$

If a_{21} times the first row of $\mathbf{A}'$ is subtracted from the second row, the resulting row is a new cointegrating vector that is still linearly independent of $\mathbf{a}_1, \mathbf{a}_3, \ldots, \mathbf{a}_n$.[3] Similarly we can subtract a_{31} times the first row of $\mathbf{A}'$ from the third row, and a_{h1} times the first row from the hth row, to deduce that the rows of the following matrix also constitute a basis for the space of cointegrating vectors:

$$\mathbf{A}_1' = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22}^* & a_{23}^* & \cdots & a_{2n}^* \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & a_{h2}^* & a_{h3}^* & \cdots & a_{hn}^* \end{bmatrix}.$$

Next, suppose that a_{22}^* is nonzero; if $a_{22}^* = 0$, we can again switch y_{2t} with some variable $y_{3t}, y_{4t}, \ldots, y_{nt}$ that does appear in the second cointegrating relation. Divide the second row of $\mathbf{A}_1'$ by a_{22}^*. The resulting row can then be multiplied by a_{12} and subtracted from the first row. Similarly, a_{32}^* times the second row of $\mathbf{A}_1'$ can be subtracted from the third row, and a_{h2}^* times the second row can be subtracted from the hth. Thus, the space of cointegrating vectors can also be represented by

$$\mathbf{A}_2' = \begin{bmatrix} 1 & 0 & a_{13}^{**} & \cdots & a_{1n}^{**} \\ 0 & 1 & a_{23}^{**} & \cdots & a_{2n}^{**} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & a_{h3}^{**} & \cdots & a_{hn}^{**} \end{bmatrix}.$$

[3]Since the first and second moments of the $(h \times 1)$ vector

$$\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_h' \end{bmatrix} \mathbf{y}_t$$

do not depend on time, neither will the first and second moments of

$$\begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' - a_{21}\mathbf{a}_1' \\ \vdots \\ \mathbf{a}_h' \end{bmatrix} \mathbf{y}_t.$$

Furthermore, the assumption that $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ are linearly independent means that no linear combination of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ is zero, and so no linear combination of $\mathbf{a}_1, \mathbf{a}_2 - a_{21}\mathbf{a}_1, \ldots, \mathbf{a}_h$ can be zero either. Hence $\mathbf{a}_1, \mathbf{a}_2 - a_{21}\mathbf{a}_1, \ldots, \mathbf{a}_h$ also constitute a basis for the space of cointegrating vectors.

Proceeding through each of the h rows of $\mathbf{A}'$ in this fashion, it follows that given any $(n \times 1)$ vector $\mathbf{y}_t$ that is characterized by exactly h cointegrating relations, it is possible to order the variables $(y_{1t}, y_{2t}, \ldots, y_{nt})$ in such a way that the cointegrating relations can be represented by an $(h \times n)$ matrix $\mathbf{A}'$ of the form

$$\begin{aligned}\mathbf{A}' &= \begin{bmatrix} 1 & 0 & \cdots & 0 & -\gamma_{1,h+1} & -\gamma_{1,h+2} & \cdots & -\gamma_{1,n} \\ 0 & 1 & \cdots & 0 & -\gamma_{2,h+1} & -\gamma_{2,h+2} & \cdots & -\gamma_{2,n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & -\gamma_{h,h+1} & -\gamma_{h,h+2} & \cdots & -\gamma_{h,n} \end{bmatrix} \\ &= [\mathbf{I}_h \quad -\mathbf{\Gamma}'],\end{aligned} \qquad [19.1.16]$$

where $\mathbf{\Gamma}'$ is an $(h \times g)$ matrix of coefficients for $g \equiv n - h$.

Let $\mathbf{z}_t$ denote the residuals associated with the set of cointegrating relations:

$$\underset{(h\times 1)}{\mathbf{z}_t} \equiv \mathbf{A}'\mathbf{y}_t. \qquad [19.1.17]$$

Since $\mathbf{z}_t$ is stationary, the mean $\boldsymbol{\mu}_1^* \equiv E(\mathbf{z}_t)$ exists, and we can define

$$\mathbf{z}_t^* \equiv \mathbf{z}_t - \boldsymbol{\mu}_1^*. \qquad [19.1.18]$$

Partition $\mathbf{y}_t$ as

$$\underset{(n\times 1)}{\mathbf{y}_t} = \begin{bmatrix} \underset{(h\times 1)}{\mathbf{y}_{1t}} \\ \underset{(g\times 1)}{\mathbf{y}_{2t}} \end{bmatrix}. \qquad [19.1.19]$$

Substituting [19.1.16], [19.1.18], and [19.1.19] into [19.1.17] results in

$$\mathbf{z}_t^* + \boldsymbol{\mu}_1^* = [\mathbf{I}_h \quad -\mathbf{\Gamma}']\begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{bmatrix}$$

or

$$\underset{(h\times 1)}{\mathbf{y}_{1t}} = \underset{(h\times g)}{\mathbf{\Gamma}'} \cdot \underset{(g\times 1)}{\mathbf{y}_{2t}} + \underset{(h\times 1)}{\boldsymbol{\mu}_1^*} + \underset{(h\times 1)}{\mathbf{z}_t^*}. \qquad [19.1.20]$$

A representation for $\mathbf{y}_{2t}$ is given by the last g rows of [19.1.11]:

$$\underset{(g\times 1)}{\Delta\mathbf{y}_{2t}} = \underset{(g\times 1)}{\boldsymbol{\delta}_2} + \underset{(g\times 1)}{\mathbf{u}_{2t}}, \qquad [19.1.21]$$

where $\boldsymbol{\delta}_2$ and $\mathbf{u}_{2t}$ represent the last g elements of the $(n \times 1)$ vectors $\boldsymbol{\delta}$ and $\mathbf{u}_t$, respectively. Equations [19.1.20] and [19.1.21] constitute Phillips's (1991) triangular representation of a system with exactly h cointegrating relations. Note that $\mathbf{z}_t^*$ and $\mathbf{u}_{2t}$ represent zero-mean stationary disturbances in this representation.

If a vector $\mathbf{y}_t$ is characterized by exactly h cointegrating relations with the variables ordered so that [19.1.20] and [19.1.21] hold, then the $(g \times 1)$ vector $\mathbf{y}_{2t}$ is $I(1)$ with no cointegrating relations. To verify this last claim, notice that if some linear combination $\mathbf{c}'\mathbf{y}_{2t}$ were stationary, this would mean that $(\mathbf{0}', \mathbf{c}')\mathbf{y}_t$ would be stationary or that $(\mathbf{0}', \mathbf{c}')$ would be a cointegrating vector for $\mathbf{y}_t$. But $(\mathbf{0}', \mathbf{c}')$ is linearly independent of the rows of $\mathbf{A}'$ in [19.1.16], and by the assumption that the rows of $\mathbf{A}'$ constitute a basis for the space of cointegrating vectors, the linear combination $(\mathbf{0}', \mathbf{c}')\mathbf{y}_t$ cannot be stationary.

Expressions [19.1.1] and [19.1.2] are a simple example of a cointegrated system expressed in triangular form. For the purchasing power parity example

[19.1.6], the triangular representation would be

$$p_t = \gamma_1 s_t + \gamma_2 p_t^* + \mu_1^* + z_t^*$$

$$\Delta s_t = \delta_s + u_{st}$$

$$\Delta p_t^* = \delta_{p^*} + u_{p^*,t},$$

where the hypothesized values are $\gamma_1 = \gamma_2 = 1$.

The Stock-Watson Common Trends Representation

Another useful representation for any cointegrated system was proposed by Stock and Watson (1988). Suppose that an $(n \times 1)$ vector $\mathbf{y}_t$ is characterized by exactly h cointegrating relations with $g \equiv n - h$. We have seen that it is possible to order the elements of $\mathbf{y}_t$ in such a way that a triangular representation of the form of [19.1.20] and [19.1.21] exists with $(\mathbf{z}_t^{*\prime}, \mathbf{u}_{2t}')'$ a stationary $(n \times 1)$ vector with zero mean. Suppose that

$$\begin{bmatrix} \mathbf{z}_t^* \\ \mathbf{u}_{2t} \end{bmatrix} = \sum_{s=0}^{\infty} \begin{bmatrix} \mathbf{H}_s \boldsymbol{\varepsilon}_{t-s} \\ \mathbf{J}_s \boldsymbol{\varepsilon}_{t-s} \end{bmatrix}$$

for $\boldsymbol{\varepsilon}_t$ an $(n \times 1)$ white noise process, with $\{s \cdot \mathbf{H}_s\}_{s=0}^{\infty}$ and $\{s \cdot \mathbf{J}_s\}_{s=0}^{\infty}$ absolutely summable sequences of $(h \times n)$ and $(g \times n)$ matrices, respectively. Adapting the result in [18.1.6], equation [19.1.21] implies that

$$\begin{aligned} \mathbf{y}_{2t} &= \mathbf{y}_{2,0} + \boldsymbol{\delta}_2 \cdot t + \sum_{s=1}^{t} \mathbf{u}_{2s} \\ &= \mathbf{y}_{2,0} + \boldsymbol{\delta}_2 \cdot t + \mathbf{J}(1) \cdot (\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t) + \boldsymbol{\eta}_{2t} - \boldsymbol{\eta}_{2,0}, \end{aligned} \qquad [19.1.22]$$

where $\mathbf{J}(1) \equiv (\mathbf{J}_0 + \mathbf{J}_1 + \mathbf{J}_2 + \cdots)$, $\boldsymbol{\eta}_{2t} \equiv \Sigma_{s=0}^{\infty} \boldsymbol{\alpha}_{2s} \boldsymbol{\varepsilon}_{t-s}$, and $\boldsymbol{\alpha}_{2s} \equiv -(\mathbf{J}_{s+1} + \mathbf{J}_{s+2} + \mathbf{J}_{s+3} + \cdots)$. Since the $(n \times 1)$ vector $\boldsymbol{\varepsilon}_t$ is white noise, the $(g \times 1)$ vector $\mathbf{J}(1) \cdot \boldsymbol{\varepsilon}_t$ is also white noise, implying that each element of the $(g \times 1)$ vector $\boldsymbol{\xi}_{2t}$ defined by

$$\boldsymbol{\xi}_{2t} \equiv \mathbf{J}(1) \cdot (\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t) \qquad [19.1.23]$$

is described by a random walk.

Substituting [19.1.23] into [19.1.22] results in

$$\mathbf{y}_{2t} = \tilde{\boldsymbol{\mu}}_2 + \boldsymbol{\delta}_2 \cdot t + \boldsymbol{\xi}_{2t} + \boldsymbol{\eta}_{2t} \qquad [19.1.24]$$

for $\tilde{\boldsymbol{\mu}}_2 \equiv (\mathbf{y}_{2,0} - \boldsymbol{\eta}_{2,0})$. Substituting [19.1.24] into [19.1.20] produces

$$\mathbf{y}_{1t} = \tilde{\boldsymbol{\mu}}_1 + \boldsymbol{\Gamma}'(\boldsymbol{\delta}_2 \cdot t + \boldsymbol{\xi}_{2t}) + \tilde{\boldsymbol{\eta}}_{1t} \qquad [19.1.25]$$

for $\tilde{\boldsymbol{\mu}}_1 \equiv \boldsymbol{\mu}_1^* + \boldsymbol{\Gamma}'\tilde{\boldsymbol{\mu}}_2$ and $\tilde{\boldsymbol{\eta}}_{1t} \equiv \mathbf{z}_t^* + \boldsymbol{\Gamma}'\boldsymbol{\eta}_{2t}$.

Equations [19.1.24] and [19.1.25] give Stock and Watson's (1988) common trends representation. These equations show that the vector $\mathbf{y}_t$ can be described as a stationary component,

$$\begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \tilde{\boldsymbol{\mu}}_2 \end{bmatrix} + \begin{bmatrix} \tilde{\boldsymbol{\eta}}_{1t} \\ \boldsymbol{\eta}_{2t} \end{bmatrix},$$

plus linear combinations of up to g common deterministic trends, as described by the $(g \times 1)$ vector $\boldsymbol{\delta}_2 \cdot t$, and linear combinations of g common random walk variables as described by the $(g \times 1)$ vector $\boldsymbol{\xi}_{2t}$.

Implications of Cointegration for the Vector Autoregressive Representation

Although a *VAR* in differences is not consistent with a cointegrated system, a *VAR* in levels could be. Suppose that the level of $\mathbf{y}_t$ can be represented as a nonstationary pth-order vector autoregression:

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad [19.1.26]$$

or

$$\boldsymbol{\Phi}(L)\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t, \qquad [19.1.27]$$

where

$$\boldsymbol{\Phi}(L) \equiv \mathbf{I}_n - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \cdots - \boldsymbol{\Phi}_p L^p. \qquad [19.1.28]$$

Suppose that $\Delta\mathbf{y}_t$ has the Wold representation

$$(1 - L)\mathbf{y}_t = \boldsymbol{\delta} + \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t. \qquad [19.1.29]$$

Premultiplying [19.1.29] by $\boldsymbol{\Phi}(L)$ results in

$$(1 - L)\boldsymbol{\Phi}(L)\mathbf{y}_t = \boldsymbol{\Phi}(1)\boldsymbol{\delta} + \boldsymbol{\Phi}(L)\boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t. \qquad [19.1.30]$$

Substituting [19.1.27] into [19.1.30], we have

$$(1 - L)\boldsymbol{\varepsilon}_t = \boldsymbol{\Phi}(1)\boldsymbol{\delta} + \boldsymbol{\Phi}(L)\boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t, \qquad [19.1.31]$$

since $(1 - L)\boldsymbol{\alpha} = \mathbf{0}$. Now, equation [19.1.31] has to hold for all realizations of $\boldsymbol{\varepsilon}_t$, which requires that

$$\boldsymbol{\Phi}(1)\boldsymbol{\delta} = \mathbf{0} \qquad [19.1.32]$$

and that $(1 - L)\mathbf{I}_n$ and $\boldsymbol{\Phi}(L)\boldsymbol{\Psi}(L)$ represent the identical polynomials in L. This means that

$$(1 - z)\mathbf{I}_n = \boldsymbol{\Phi}(z)\boldsymbol{\Psi}(z) \qquad [19.1.33]$$

for all values of z. In particular, for $z = 1$, equation [19.1.33] implies that

$$\boldsymbol{\Phi}(1)\boldsymbol{\Psi}(1) = \mathbf{0}. \qquad [19.1.34]$$

Let $\boldsymbol{\pi}'$ denote any row of $\boldsymbol{\Phi}(1)$. Then [19.1.34] and [19.1.32] state that $\boldsymbol{\pi}'\boldsymbol{\Psi}(1) = \mathbf{0}'$ and $\boldsymbol{\pi}'\boldsymbol{\delta} = 0$. Recalling [19.1.12] and [19.1.15], this means that $\boldsymbol{\pi}$ is a cointegrating vector. If $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ form a basis for the space of cointegrating vectors, then it must be possible to express $\boldsymbol{\pi}$ as a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$—that is, there exists an $(h \times 1)$ vector b such that

$$\boldsymbol{\pi} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_h]\mathrm{b}$$

or

$$\boldsymbol{\pi}' = \mathrm{b}'\mathbf{A}'$$

for $\mathbf{A}'$ the $(h \times n)$ matrix whose ith row is $\mathbf{a}_i'$. Applying this reasoning to each of the rows of $\boldsymbol{\Phi}(1)$, it follows that there exists an $(n \times h)$ matrix $\mathbf{B}$ such that

$$\boldsymbol{\Phi}(1) = \mathbf{B}\mathbf{A}'. \qquad [19.1.35]$$

Note that [19.1.34] implies that $\boldsymbol{\Phi}(1)$ is a singular $(n \times n)$ matrix—linear combinations of the columns of $\boldsymbol{\Phi}(1)$ of the form $\boldsymbol{\Phi}(1)\mathbf{x}$ are zero for $\mathbf{x}$ any column of $\boldsymbol{\Psi}(1)$. Thus, the determinant $|\boldsymbol{\Phi}(z)|$ contains a unit root:

$$|\mathbf{I}_n - \boldsymbol{\Phi}_1 z^1 - \boldsymbol{\Phi}_2 z^2 - \cdots - \boldsymbol{\Phi}_p z^p| = 0 \qquad \text{at } z = 1.$$

Indeed, in the light of the Stock-Watson common trends representation in [19.1.24] and [19.1.25], we could say that $\mathbf{\Phi}(z)$ contains $g = n - h$ unit roots.

Error-Correction Representation

A final representation for a cointegrated system is obtained by recalling from equation [18.2.5] that any *VAR* in the form of [19.1.26] can equivalently be written as

$$\mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\rho}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \qquad [19.1.36]$$

where

$$\boldsymbol{\rho} \equiv \mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \cdots + \mathbf{\Phi}_p \qquad [19.1.37]$$

$$\boldsymbol{\zeta}_s \equiv -[\mathbf{\Phi}_{s+1} + \mathbf{\Phi}_{s+2} + \cdots + \mathbf{\Phi}_p] \qquad \text{for } s = 1, 2, \ldots, p-1. \qquad [19.1.38]$$

Subtracting $\mathbf{y}_{t-1}$ from both sides of [19.1.36] produces

$$\Delta\mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\zeta}_0\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \qquad [19.1.39]$$

where

$$\boldsymbol{\zeta}_0 \equiv \boldsymbol{\rho} - \mathbf{I}_n = -(\mathbf{I}_n - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \cdots - \mathbf{\Phi}_p) = -\mathbf{\Phi}(1). \qquad [19.1.40]$$

Note that if $\mathbf{y}_t$ has h cointegrating relations, then substitution of [19.1.35] and [19.1.40] into [19.1.39] results in

$$\Delta\mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} - \mathbf{B}\mathbf{A}'\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \qquad [19.1.41]$$

Define $\mathbf{z}_t \equiv \mathbf{A}'\mathbf{y}_t$, noticing that $\mathbf{z}_t$ is a stationary $(h \times 1)$ vector. Then [19.1.41] can be written

$$\Delta\mathbf{y}_t = \boldsymbol{\zeta}_1 \Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} - \mathbf{B}\mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t. \qquad [19.1.42]$$

Expression [19.1.42] is known as the *error-correction* representation of the cointegrated system. For example, the first equation takes the form

$$\begin{aligned}
\Delta y_{1t} = {} & \zeta_{11}^{(1)}\Delta y_{1,t-1} + \zeta_{12}^{(1)}\Delta y_{2,t-1} + \cdots + \zeta_{1n}^{(1)}\Delta y_{n,t-1} \\
& + \zeta_{11}^{(2)}\Delta y_{1,t-2} + \zeta_{12}^{(2)}\Delta y_{2,t-2} + \cdots + \zeta_{1n}^{(2)}\Delta y_{n,t-2} + \cdots \\
& + \zeta_{11}^{(p-1)}\Delta y_{1,t-p+1} + \zeta_{12}^{(p-1)}\Delta y_{2,t-p+1} + \cdots + \zeta_{1n}^{(p-1)}\Delta y_{n,t-p+1} \\
& + \alpha_1 - b_{11}z_{1,t-1} - b_{12}z_{2,t-1} - \cdots - b_{1h}z_{h,t-1} + \varepsilon_{1t},
\end{aligned}$$

where $\zeta_{ij}^{(s)}$ indicates the row i, column j element of the matrix $\boldsymbol{\zeta}_s$, b_{ij} indicates the row i, column j element of the matrix $\mathbf{B}$, and z_{it} represents the ith element of $\mathbf{z}_t$. Thus, in the error-correction form, changes in each variable are regressed on a constant, $(p-1)$ lags of the variable's own changes, $(p-1)$ lags of changes in each of the other variables, and the levels of each of the h elements of $\mathbf{z}_{t-1}$.

For example, recall from [19.1.9] that the system of [19.1.1] and [19.1.2] can be written in the form

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma u_{2t} + u_{1t} \\ u_{2t} \end{bmatrix}.$$

Note that this is a special case of [19.1.39] with $p = 1$,

$$\boldsymbol{\zeta}_0 = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix},$$

$\varepsilon_{1t} = \gamma u_{2t} + u_{1t}$, $\varepsilon_{2t} = u_{2t}$, and all other parameters in [19.1.39] equal to zero.

The error-correction form is

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} z_{t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix},$$

where $z_t \equiv y_{1t} - \gamma y_{2t}$.

An economic interpretation of an error-correction representation was proposed by Davidson, Hendry, Srba, and Yeo (1978), who examined a relation between the log of consumption spending (denoted c_t) and the log of income (y_t) of the form

$$(1 - L^4)c_t = \beta_1(1 - L^4)y_t + \beta_2(1 - L^4)y_{t-1} + \beta_3(c_{t-4} - y_{t-4}) + u_t. \quad [19.1.43]$$

This equation was fitted to quarterly data, so that $(1 - L^4)c_t$ denotes the percentage change in consumption over its value in the comparable quarter of the preceding year. The authors argued that seasonal differences $(1 - L^4)$ provided a better description of the data than would simple quarterly differences $(1 - L)$. Their claim was that seasonally differenced consumption $(1 - L^4)c_t$ could not be described using only its own lags or those of seasonally differenced income. In addition to these factors, [19.1.43] includes the "error-correction" term $\beta_3(c_{t-4} - y_{t-4})$. One could argue that there is a long run, historical average ratio of consumption to income, in which case the difference between the logs of consumption and income, $c_t - y_t$, would be a stationary random variable, even though log consumption or log income viewed by itself exhibits a unit root. For $\beta_3 < 0$, equation [19.1.43] asserts that if consumption had previously been a larger-than-normal share of income (so that $c_{t-4} - y_{t-4}$ is larger than normal), then that causes c_t to be lower for any given values of the other explanatory variables. The term $(c_{t-4} - y_{t-4})$ is viewed as the "error" from the long-run equilibrium relation, and β_3 gives the "correction" to c_t caused by this error.

Restrictions on the Constant Term in the VAR *Representation*

Notice that all the variables appearing in the error-correction representation [19.1.42] are stationary. Taking expectations of both sides of that equation results in

$$(\mathbf{I}_n - \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2 - \cdots - \boldsymbol{\zeta}_{p-1})\boldsymbol{\delta} = \boldsymbol{\alpha} - \mathbf{B}\boldsymbol{\mu}_1^*, \quad [19.1.44]$$

where $\boldsymbol{\delta} = E(\Delta \mathbf{y}_t)$ and $\boldsymbol{\mu}_1^* = E(\mathbf{z}_t)$. Assuming that the roots of

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

are all outside the unit circle, the matrix $(\mathbf{I}_n - \boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2 - \cdots - \boldsymbol{\zeta}_{p-1})$ is nonsingular. Thus, in order to represent a system in which there is no drift in any of the variables $(\boldsymbol{\delta} = \mathbf{0})$, we would have to impose the restriction

$$\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\mu}_1^*. \quad [19.1.45]$$

In the absence of any restriction on $\boldsymbol{\alpha}$, the system of [19.1.42] implies that there are g separate time trends that account for the trend in $\mathbf{y}_t$.

Granger Representation Theorem

For convenience, some of the preceding results are now summarized in the form of a proposition.

Proposition 19.1: *(Granger representation theorem). Consider an $(n \times 1)$ vector $\mathbf{y}_t$ where $\Delta\mathbf{y}_t$ satisfies [19.1.29] for $\boldsymbol{\varepsilon}_t$ white noise with positive definite variance-covariance matrix and $\{s \cdot \boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ absolutely summable. Suppose that there are exactly h cointegrating relations among the elements of $\mathbf{y}_t$. Then there exists an $(h \times n)$ matrix $\mathbf{A}'$ whose rows are linearly independent such that the $(h \times 1)$ vector $\mathbf{z}_t$ defined by*

$$\mathbf{z}_t \equiv \mathbf{A}'\mathbf{y}_t$$

is stationary. The matrix $\mathbf{A}'$ has the property that

$$\mathbf{A}'\boldsymbol{\Psi}(1) = \mathbf{0}.$$

If, moreover, the process can be represented as the pth-order VAR in levels as in equation [19.1.26], then there exists an $(n \times h)$ matrix $\mathbf{B}$ such that

$$\boldsymbol{\Phi}(1) = \mathbf{B}\mathbf{A}',$$

and there further exist $(n \times n)$ matrices $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1}$ such that

$$\Delta\mathbf{y}_t = \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} - \mathbf{B}\mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t.$$

19.2. *Testing the Null Hypothesis of No Cointegration*

This section discusses tests for cointegration. The approach will be to test the null hypothesis that there is no cointegration among the elements of an $(n \times 1)$ vector $\mathbf{y}_t$; rejection of the null is then taken as evidence of cointegration.

Testing for Cointegration When the Cointegrating Vector Is Known

Often when theoretical considerations suggest that certain variables will be cointegrated, or that $\mathbf{a}'\mathbf{y}_t$ is stationary for some $(n \times 1)$ cointegrating vector $\mathbf{a}$, the theory is based on a particular known value for $\mathbf{a}$. In the purchasing power parity example [19.1.6], $\mathbf{a} = (1, -1, -1)'$. The Davidson, Hendry, Srba, and Yeo hypothesis (1978) that consumption is a stable fraction of income implies a cointegrating vector of $\mathbf{a} = (1, -1)'$, as did Kremers's assertion (1989) that government debt is a stable multiple of GNP.

If the interest in cointegration is motivated by the possibility of a particular known cointegrating vector $\mathbf{a}$, then by far the best method is to use this value directly to construct a test for cointegration. To implement this approach, we first test whether each of the elements of $\mathbf{y}_t$ is individually $I(1)$. This can be done using any of the tests discussed in Chapter 17. Assuming that the null hypothesis of a unit root in each series individually is accepted, we next construct the scalar $z_t = \mathbf{a}'\mathbf{y}_t$. Notice that if $\mathbf{a}$ is truly a cointegrating vector, then $\mathbf{a}'\mathbf{y}_t$ will be $I(0)$. If $\mathbf{a}$ is not a cointegrating vector, then $\mathbf{a}'\mathbf{y}_t$ will be $I(1)$. Thus, a test of the null hypothesis that z_t is $I(1)$ is equivalent to a test of the null hypothesis that $\mathbf{y}_t$ is not cointegrated. If the null hypothesis that z_t is $I(1)$ is rejected, we would conclude that $z_t = \mathbf{a}'\mathbf{y}_t$ is stationary, or that $\mathbf{y}_t$ is cointegrated with cointegrating vector $\mathbf{a}$. The null hypothesis that z_t is $I(1)$ can also be tested using any of the approaches in Chapter 17.

For example, Figure 19.2 plots monthly data from 1973:1 to 1989:10 for the consumer price indexes for the United States (p_t) and Italy (p_t^*), along with the

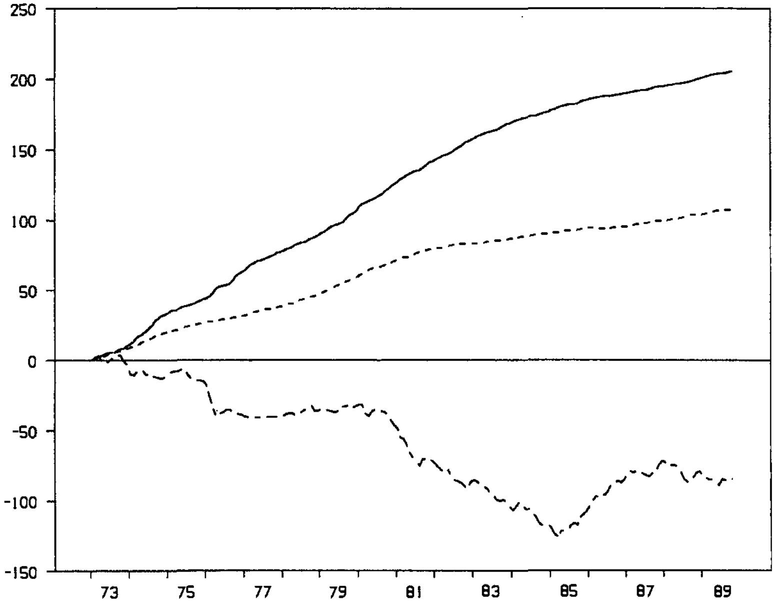


FIGURE 19.2 One hundred times the log of the price level in the United States (p_t), the dollar-lira exchange rate (s_t), and the price level in Italy (p_t^*), monthly, 1973–89. Key: ---- p_t; ——— s_t; ——— p_t^*.

exchange rate (s_t), where s_t is in terms of the number of U.S. dollars needed to purchase an Italian lira. Natural logs of the raw data were taken and multiplied by 100, and the initial value for 1973:1 was then subtracted, as in

$$p_t = 100\cdot[\log(P_t) - \log(P_{1973:1})].$$

The purpose of subtracting the constant $\log(P_{1973:1})$ from each observation is to normalize each series to be zero for 1973:1 so that the graph is easier to read. Multiplying the log by 100 means that p_t is approximately the percentage difference between P_t and its starting value $P_{1973:1}$. The graph shows that Italy experienced about twice the average inflation rate of the United States over this period and that the lira dropped in value relative to the dollar (that is, s_t fell) by roughly this same proportion.

Figure 19.3 plots the real exchange rate,

$$z_t \equiv p_t - s_t - p_t^*.$$

It appears that the trends are eliminated by this transformation, though deviations of the real exchange rate from its historical mean can persist for several years.

To test for cointegration, we first verify that p_t, p_t^*, and s_t are each individually $I(1)$. Certainly, we anticipate the average inflation rate to be positive ($E(\Delta p_t) > 0$), so that the natural null hypothesis is that p_t is a unit root process with positive drift, while the alternative is that p_t is stationary around a deterministic time trend. With monthly data it is a good idea to include at least twelve lags in the regression. Thus, the following model was estimated by OLS for the U.S. data for t = 1974:2

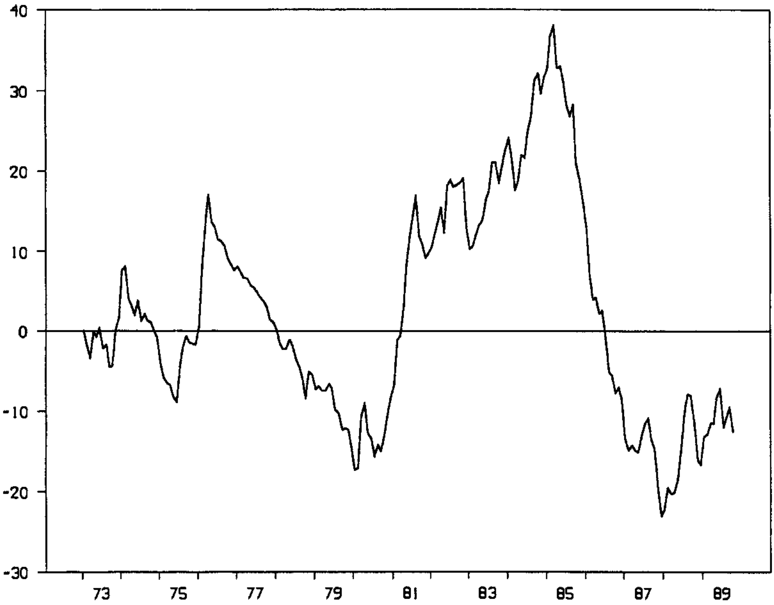


FIGURE 19.3 The real dollar-lira exchange rate, monthly, 1973–89.

through 1989:10 (standard errors in parentheses):

$$\begin{aligned} p_t = {} & \underset{(0.08)}{0.55}\,\Delta p_{t-1} - \underset{(0.09)}{0.06}\,\Delta p_{t-2} + \underset{(0.08)}{0.07}\,\Delta p_{t-3} + \underset{(0.08)}{0.06}\,\Delta p_{t-4} \\ & - \underset{(0.08)}{0.08}\,\Delta p_{t-5} - \underset{(0.07)}{0.05}\,\Delta p_{t-6} + \underset{(0.07)}{0.17}\,\Delta p_{t-7} - \underset{(0.07)}{0.07}\,\Delta p_{t-8} \\ & + \underset{(0.07)}{0.24}\,\Delta p_{t-9} - \underset{(0.07)}{0.11}\,\Delta p_{t-10} + \underset{(0.07)}{0.12}\,\Delta p_{t-11} + \underset{(0.07)}{0.05}\,\Delta p_{t-12} \\ & + \underset{(0.09)}{0.14} + \underset{(0.00307)}{0.99400}\,p_{t-1} + \underset{(0.0018)}{0.0029}\,t. \end{aligned} \qquad [19.2.1]$$

The t statistic for testing the null hypothesis that ρ (the coefficient on p_{t-1}) is unity is

$$t = (0.99400 - 1.0)/(0.00307) = -1.95.$$

Comparing this with the 5% crtical value from the case 4 section of Table B.6 for a sample of size $T = 189$, we see that $-1.95 > -3.44$. Thus, the null hypothesis of a unit root is accepted. The F test of the joint null hypothesis that $\rho = 1$ and $\delta = 0$ (for ρ the coefficient on p_{t-1} and δ the coefficient on the time trend) is 2.41. Comparing this with the critical value of 6.40 from the case 4 section of Table B.7, the null hypothesis is again accepted, further confirming the impression that U.S. prices follow a unit root process with drift.

If p_t in [19.2.1] is replaced by p_t^*, the augmented Dickey-Fuller t and F tests are calculated to be -0.13 and 4.25, respectively, so that the null hypothesis that the Italian price level follows an $I(1)$ process is again accepted. When p_t in [19.2.1] is replaced by s_t, the t and F tests are -1.58 and 1.49, so that the exchange rate likewise admits an $ARIMA(12, 1, 0)$ representation. Thus, each of the three series individually could reasonably be described as a unit root process with drift.

The next step is to test whether $z_t = p_t - s_t - p_t^*$ is stationary. According to the theory, there should not be any trend in z_t, and none appears evident in Figure 19.3. Thus, the augmented Dickey-Fuller test without trend might be used. The following estimates were obtained by *OLS*:

$$\begin{aligned} z_t = {} & \underset{(0.07)}{0.32}\,\Delta z_{t-1} - \underset{(0.08)}{0.01}\,\Delta z_{t-2} + \underset{(0.08)}{0.01}\,\Delta z_{t-3} + \underset{(0.08)}{0.02}\,\Delta z_{t-4} \\ & + \underset{(0.08)}{0.08}\,\Delta z_{t-5} - \underset{(0.08)}{0.00}\,\Delta z_{t-6} + \underset{(0.08)}{0.03}\,\Delta z_{t-7} + \underset{(0.08)}{0.08}\,\Delta z_{t-8} \\ & - \underset{(0.08)}{0.05}\,\Delta z_{t-9} + \underset{(0.08)}{0.08}\,\Delta z_{t-10} + \underset{(0.08)}{0.05}\,\Delta z_{t-11} - \underset{(0.08)}{0.01}\,\Delta z_{t-12} \\ & + \underset{(0.18)}{0.00} + \underset{(0.01410)}{0.97124}\, z_{t-1}. \end{aligned} \qquad [19.2.2]$$

Here the augmented Dickey-Fuller t test is

$$t = (0.97124 - 1.0)/(0.01410) = -2.04.$$

Comparing this with the 5% critical value for case 2 of Table B.6, we see that $-2.04 > -2.88$, and so the null hypothesis of a unit root is accepted. The F test of the joint null hypothesis that $\rho = 1$ and that the constant term is zero is 2.19 < 4.66, which is again accepted. Thus, we could accept the null hypothesis that the series are not cointegrated.

Alternatively, the null hypothesis that z_t is nonstationary could be tested using the Phillips-Perron tests. *OLS* estimation gives

$$z_t = \underset{(0.178)}{-0.030} + \underset{(0.01275)}{0.98654}\, z_{t-1} + \hat{u}_t$$

with

$$\begin{aligned} s^2 &= (T-2)^{-1} \sum_{t=1}^{T} \hat{u}_t^2 = (2.49116)^2 \\ \hat{c}_j &= T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j} \\ \hat{c}_0 &= 6.144 \\ \hat{\lambda}^2 &= \hat{c}_0 + 2\cdot \sum_{j=1}^{12} [1 - (j/13)]\hat{c}_j = 13.031. \end{aligned}$$

The Phillips-Perron Z_ρ test is then

$$\begin{aligned} Z_\rho &= T(\hat{\rho} - 1) - \tfrac{1}{2}\{T\cdot\hat{\sigma}_{\hat{\rho}} \div s\}^2(\hat{\lambda}^2 - \hat{c}_0) \\ &= (201)(0.98654 - 1) \\ &\quad - \tfrac{1}{2}\{(201)(0.01275) \div (2.49116)\}^2(13.031 - 6.144) \\ &= -6.35. \end{aligned}$$

Since $-6.35 > -13.9$, the null hypothesis of no cointegration is again accepted. Similarly, the Phillips-Perron Z_t test is

$$\begin{aligned} Z_t &= (\hat{c}_0/\hat{\lambda}^2)^{1/2}(\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}} - \tfrac{1}{2}\{T\cdot\hat{\sigma}_{\hat{\rho}} \div s\}(\hat{\lambda}^2 - \hat{c}_0)/\hat{\lambda} \\ &= (6.144/13.031)^{1/2}(0.98654 - 1)/(0.01275) \\ &\quad - \tfrac{1}{2}\{(201)(0.01275) \div (2.49116)\}(13.031 - 6.144)/(13.031)^{1/2} \\ &= -1.71, \end{aligned}$$

which, since $-1.71 > -2.88$, gives the same conclusion as the other tests.

Clearly, the comments about the observational equivalence of $I(0)$ and $I(1)$ processes are also applicable to testing for cointegration. There exist both $I(0)$ and $I(1)$ representations that are perfectly capable of describing the observed data for z_t plotted in Figure 19.3. Another way of describing the results is to calculate how long a deviation from purchasing power parity is likely to persist. The regression of [19.2.2] implies an autoregression in levels of the form

$$z_t = \alpha + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_{13} z_{t-13} + \varepsilon_t,$$

for which the impulse-response function,

$$\psi_j = \frac{\partial z_{t+j}}{\partial \varepsilon_t},$$

can be calculated using the methods described in Chapter 1. Figure 19.4 plots the estimated impulse-response coefficients as a function of j. An unanticipated increase in z_t would cause us to revise upward our forecast of z_{t+j} by 25% even 3 years into the future ($\psi_{36} = 0.27$). Hence, any forces that restore z_t to its historical value must operate relatively slowly. The same conclusion might have been gleaned from Figure 19.3 directly, in that it is clear that deviations of z_t from its historical norm can persist for a number of years.

Estimating the Cointegrating Vector

If the theoretical model of the system dynamics does not suggest a particular value for the cointegrating vector $\mathbf{a}$, then one approach to testing for cointegration is first to estimate $\mathbf{a}$ by *OLS*. To see why this produces a reasonable initial estimate,

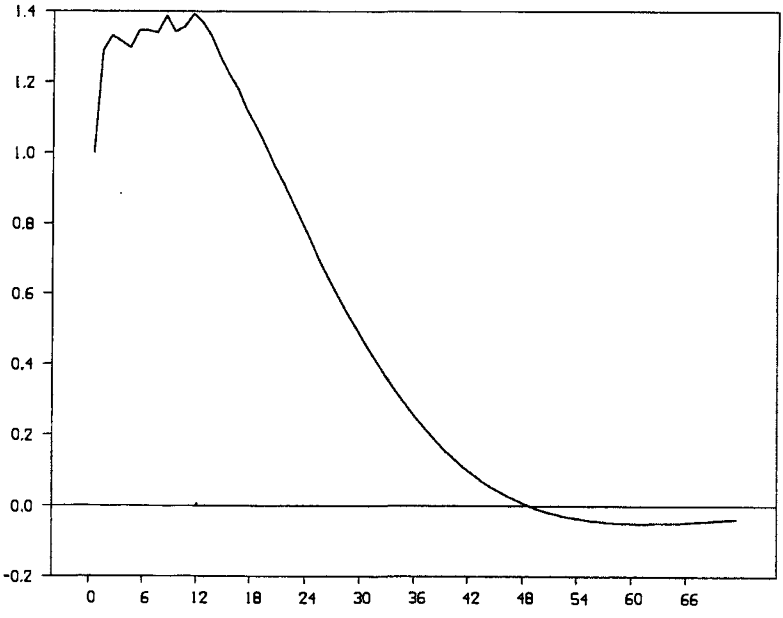


FIGURE 19.4 Impulse-response function for the real dollar-lira exchange rate. Graph shows $\psi_j = \partial(p_{t+j} - s_{t+j} - p^*_{t+j})/\varepsilon_t$ as a function of j.

note that if $z_t = \mathbf{a}'\mathbf{y}_t$ is stationary and ergodic for second moments, then

$$T^{-1}\sum_{t=1}^{T} z_t^2 = T^{-1}\sum_{t=1}^{T} (\mathbf{a}'\mathbf{y}_t)^2 \xrightarrow{p} E(z_t^2). \qquad [19.2.3]$$

By contrast, if $\mathbf{a}$ is not a cointegrating vector, then $z_t = \mathbf{a}'\mathbf{y}_t$ is $I(1)$, and so, from result (h) of Proposition 17.3,

$$T^{-2}\sum_{t=1}^{T} (\mathbf{a}'\mathbf{y}_t)^2 \xrightarrow{L} \lambda^2 \cdot \int_0^1 [W(r)]^2\, dr, \qquad [19.2.4]$$

where $W(r)$ is standard Brownian motion and λ is a parameter determined by the autocovariances of $(1 - L)z_t$. Hence, if $\mathbf{a}$ is not a cointegrating vector, the statistic in [19.2.3] would diverge to $+\infty$.

This suggests that we can obtain a consistent estimate of a cointegrating vector by choosing $\mathbf{a}$ so as to minimize [19.2.3] subject to some normalization condition on $\mathbf{a}$. Indeed, such an estimator turns out to be superconsistent, converging at rate T rather than $T^{1/2}$.

If it is known for certain that the cointegrating vector has a nonzero coefficient for the first element of $\mathbf{y}_t$ ($a_1 \neq 0$), then a particularly convenient normalization is to set $a_1 = 1$ and represent subsequent entries of $\mathbf{a}$ $(a_2, a_3, \ldots, a_n)$ as the negatives of a set of unknown parameters $(\gamma_2, \gamma_3, \ldots, \gamma_n)$:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 1 \\ -\gamma_2 \\ -\gamma_3 \\ \vdots \\ -\gamma_n \end{bmatrix}. \qquad [19.2.5]$$

In this case, the objective is to choose $(\gamma_2, \gamma_3, \ldots, \gamma_n)$ so as to minimize

$$T^{-1}\sum_{t=1}^{T} (\mathbf{a}'\mathbf{y}_t)^2 = T^{-1}\sum_{t=1}^{T} (y_{1t} - \gamma_2 y_{2t} - \gamma_3 y_{3t} - \cdots - \gamma_n y_{nt})^2. \qquad [19.2.6]$$

This minimization is, of course, achieved by an OLS regression of the first element of $\mathbf{y}_t$ on all of the others:

$$y_{1t} = \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t. \qquad [19.2.7]$$

Consistent estimates of $\gamma_2, \gamma_3, \ldots, \gamma_n$ are also obtained when a constant term is included in [19.2.7], as in

$$y_{1t} = \alpha + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t \qquad [19.2.8]$$

or

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + u_t,$$

where $\boldsymbol{\gamma}' \equiv (\gamma_2, \gamma_3, \ldots, \gamma_n)$ and $\mathbf{y}_{2t} \equiv (y_{2t}, y_{3t}, \ldots, y_{nt})'$.

These points were first analyzed by Phillips and Durlauf (1986) and Stock (1987) and are formally summarized in the following proposition.

Proposition 19.2: *Let y_{1t} be a scalar and $\mathbf{y}_{2t}$ be a $(g \times 1)$ vector. Let $n \equiv g + 1$, and suppose that the $(n \times 1)$ vector $(y_{1t}, \mathbf{y}_{2t}')'$ is characterized by exactly one cointegrating relation ($h = 1$) that has a nonzero coefficient on y_{1t}. Let the triangular* *representation for the system be*

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + z_t^* \qquad [19.2.9]$$

$$\Delta\mathbf{y}_{2t} = \mathbf{u}_{2t}. \qquad [19.2.10]$$

Suppose that

$$\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t} \end{bmatrix} = \boldsymbol{\Psi}^*(L)\boldsymbol{\varepsilon}_t, \qquad [19.2.11]$$

where $\boldsymbol{\varepsilon}_t$ is an $(n \times 1)$ i.i.d. vector with mean zero, finite fourth moments, and positive definite variance-covariance matrix $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{PP}'$. Suppose further that the sequence of $(n \times n)$ matrices $\{s\cdot\boldsymbol{\Psi}_s^*\}_{s=0}^{\infty}$ is absolutely summable and that the rows of $\boldsymbol{\Psi}^*(1)$ are linearly independent. Let $\hat{\alpha}_T$ and $\hat{\boldsymbol{\gamma}}_T$ be estimates based on OLS estimation of [19.2.9],

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\boldsymbol{\gamma}}_T \end{bmatrix} = \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_{1t} \\ \Sigma\mathbf{y}_{2t}y_{1t} \end{bmatrix}, \qquad [19.2.12]$$

where Σ indicates summation over t from 1 to T. Partition $\boldsymbol{\Psi}^*(1)\cdot\mathbf{P}$ as

$$\underset{(n\times n)}{\boldsymbol{\Psi}^*(1)\cdot\mathbf{P}} = \begin{bmatrix} \underset{(1\times n)}{\boldsymbol{\lambda}_1^{*\prime}} \\ \underset{(g\times n)}{\boldsymbol{\Lambda}_2^*} \end{bmatrix}.$$

Then

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \left\{\int [\mathbf{W}(r)]'\,dr\right\}\cdot\boldsymbol{\Lambda}_2^{*\prime} \\ \boldsymbol{\Lambda}_2^*\cdot\int \mathbf{W}(r)\,dr & \boldsymbol{\Lambda}_2^*\cdot\left\{\int [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\boldsymbol{\Lambda}_2^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix}, \qquad [19.2.13]$$

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

$$h_1 \equiv \boldsymbol{\lambda}_1^{*\prime}\cdot\mathbf{W}(1)$$

$$\mathbf{h}_2 \equiv \boldsymbol{\Lambda}_2^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\boldsymbol{\lambda}_1^* + \sum_{v=0}^{\infty} E(\mathbf{u}_{2t}z_{t+v}^*).$$

Note that the *OLS* estimate of the cointegrating vector is consistent even though the error term u_t in [19.2.8] may be serially correlated and correlated with $\Delta y_{2t}, \Delta y_{3t}, \ldots, \Delta y_{nt}$. The latter correlation would contribute a bias in the limiting distribution of $T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma})$, for then the random variable $\mathbf{h}_2$ would not have mean zero. However, the bias in $\hat{\boldsymbol{\gamma}}_T$ is $O_p(T^{-1})$.

Since the *OLS* estimates are consistent, the average squared sample residual converges to

$$T^{-1}\sum_{t=1}^{T} \hat{u}_{t,T}^2 \xrightarrow{p} E(u_t^2),$$

whereas the sample variance of y_{1t},

$$T^{-1}\sum_{t=1}^{T} (y_{1t} - \bar{y}_1)^2,$$

diverges to $+\infty$. Hence, the R^2 for the regression of [19.2.8] will converge to unity as the sample size grows.

Cointegration can be viewed as a structural assumption under which certain behavioral relations of interest can be estimated from the data by *OLS*. Consider the supply-and-demand example in equations [9.1.2] and [9.1.1]:

$$q_t^s = \gamma p_t + \varepsilon_t^s \qquad [19.2.14]$$

$$q_t^d = \beta p_t + \varepsilon_t^d. \qquad [19.2.15]$$

We noted in equation [9.1.6] that if ε_t^d and ε_t^s are i.i.d. with $\text{Var}(\varepsilon_t^s)$ finite, then as the variance of ε_t^d goes to infinity, *OLS* estimation of [19.2.14] produces a consistent estimate of the supply elasticity γ despite the potential simultaneous equations bias. This is because the large shifts in the demand curve effectively trace out the supply curve in the sample; see Figure 9.3. More generally, if ε_t^s is $I(0)$ and ε_t^d is $I(1)$, then [19.2.14] and [19.2.15] imply that $(q_t, p_t)'$ is cointegrated with cointegrating vector $(1, -\gamma)'$. In this case the cointegrating vector can be consistently estimated by *OLS* for essentially the same reason as in Figure 9.3. The hypothesis that a certain structural relation involving $I(1)$ variables is characterized by an $I(0)$ disturbance amounts to a structural assumption that can help identify the parameters of the structural relation.

Although the estimates based on [19.2.8] are consistent, there often exist alternative estimates that are superior. These will be discussed in Section 19.3. *OLS* estimation of [19.2.8] is proposed only as a quick way to obtain an initial estimate of the cointegrating vector.

It was assumed in Proposition 19.2 that $\Delta\mathbf{y}_{2t}$ had mean zero. If, instead, $E(\Delta\mathbf{y}_{2t}) = \boldsymbol{\delta}_2$, it is straightforward to generalize Proposition 19.2 using a rotation of variables as in [18.2.43]; for details, see Hansen (1992). As long as there is no time trend in the true cointegrating relation [19.2.9], the estimate $\hat{\boldsymbol{\gamma}}_T$ based on *OLS* estimation of [19.2.8] will be superconsistent regardless of whether the $I(1)$ vector $\mathbf{y}_{2t}$ includes a deterministic time trend or not.

The Role of Normalization

The *OLS* estimate of the cointegrating vector was obtained by normalizing the first element of the cointegrating vector $\mathbf{a}$ to be unity. The proposal was then to regress the first element of $\mathbf{y}_t$ on the others. For example, with $n = 2$, we would regress y_{1t} on y_{2t}:

$$y_{1t} = \alpha + \gamma y_{2t} + u_t.$$

Obviously, we might equally well have normalized $a_2 = 1$ and used the same argument to suggest a regression of y_{2t} on y_{1t}:

$$y_{2t} = \theta + \aleph y_{1t} + v_t.$$

The *OLS* estimate $\hat{\aleph}$ is not simply the inverse of $\hat{\gamma}$, meaning that these two regressions will give different estimates of the cointegrating vector:

$$\begin{bmatrix} 1 \\ -\hat{\gamma} \end{bmatrix} \neq -\hat{\gamma}\begin{bmatrix} -\hat{\aleph} \\ 1 \end{bmatrix}.$$

Only in the limiting case where the R^2 is 1 would the two estimates coincide.

Thus, choosing which variable to call y_1 and which to call y_2 might end up making a material difference for the estimate of $\mathbf{a}$ as well as for the evidence one finds for cointegration among the series. One approach that avoids this normalization problem is the full-information maximum likelihood estimate proposed by Johansen (1988, 1991). This will be discussed in detail in Chapter 20.

What Is the Regression Estimating When There Is More Than One Cointegrating Relation?

The limiting distribution of the *OLS* estimate in Proposition 19.2 was derived under the assumption that there is just one cointegrating relation ($h = 1$). In the more general case with $h > 1$, *OLS* estimation of [19.2.8] should still provide a consistent estimate of a cointegrating vector by virtue of the argument given in [19.2.3] and [19.2.4]. But which cointegrating vector is it?

Consider the general triangular representation for a vector with h cointegrating relations given in [19.1.20] and [19.1.21]:

$$\mathbf{y}_{1t} = \boldsymbol{\mu}_1^* + \boldsymbol{\Gamma}'\mathbf{y}_{2t} + \mathbf{z}_t^* \qquad [19.2.16]$$

$$\Delta\mathbf{y}_{2t} = \boldsymbol{\delta}_2 + \mathbf{u}_{2t}, \qquad [19.2.17]$$

where the $(h \times 1)$ vector $\mathbf{y}_{1t}$ contains the first h elements of $\mathbf{y}_t$ and $\mathbf{y}_{2t}$ contains the remaining g elements. Since $\mathbf{z}_t^* \equiv (z_{1t}^*, z_{2t}^*, \ldots, z_{ht}^*)'$ is covariance-stationary with mean zero, we can define $\beta_2, \beta_3, \ldots, \beta_h$ to be the population coefficients associated with a linear projection of z_{1t}^* on $z_{2t}^*, z_{3t}^*, \ldots, z_{ht}^*$:

$$z_{1t}^* = \beta_2 z_{2t}^* + \beta_3 z_{3t}^* + \cdots + \beta_h z_{ht}^* + u_t, \qquad [19.2.18]$$

where u_t by construction has mean zero and is uncorrelated with $z_{2t}^*, z_{3t}^*, \ldots, z_{ht}^*$.

The following proposition, adapted from Wooldridge (1991), shows that the sample residual $\hat{u}_t$ resulting from *OLS* estimation of [19.2.8] converges in probability to the population residual u_t associated with the linear projection in [19.2.18]. In other words, among the set of possible cointegrating relations, *OLS* estimation of [19.2.8] selects the relation whose residuals are uncorrelated with any other $I(1)$ linear combinations of $(y_{2t}, y_{3t}, \ldots, y_{nt})$.

Proposition 19.3: *Let $\mathbf{y}_t = (\mathbf{y}_{1t}', \mathbf{y}_{2t}')'$ satisfy [19.2.16] and [19.2.17] with $\mathbf{y}_{1t}$ an $(h \times 1)$ vector with $h > 1$, and let $\beta_2, \beta_3, \ldots, \beta_h$ denote the linear projection coefficients in [19.2.18]. Suppose that*

$$\begin{bmatrix} \mathbf{z}_t^* \\ \mathbf{u}_{2t} \end{bmatrix} = \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s^* \boldsymbol{\varepsilon}_{t-s},$$

where $\{s \cdot \boldsymbol{\Psi}_s^*\}_{s=0}^{\infty}$ is absolutely summable and $\boldsymbol{\varepsilon}_t$ is an i.i.d. $(n \times 1)$ vector with mean zero, variance $\mathbf{PP}'$, and finite fourth moments. Suppose further that the rows of $\boldsymbol{\Psi}^*(1) \cdot \mathbf{P}$ are linearly independent. Then the coefficient estimates associated with OLS estimation of

$$y_{1t} = \alpha + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t \qquad [19.2.19]$$

converge in probability to

$$\hat{\alpha}_T \xrightarrow{p} [1 \quad -\boldsymbol{\beta}']\boldsymbol{\mu}_1^*, \qquad [19.2.20]$$

where

$$\underset{(h-1)\times 1}{\boldsymbol{\beta}} \equiv (\beta_2, \beta_3, \ldots, \beta_h)'$$

and

$$\begin{bmatrix} \hat{\gamma}_{2,T} \\ \hat{\gamma}_{3,T} \\ \vdots \\ \hat{\gamma}_{n,T} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma}_2 \end{bmatrix} \qquad [19.2.21]$$

where

$$\underset{(g\times 1)}{\boldsymbol{\gamma}_2} \equiv \boldsymbol{\Gamma}\begin{bmatrix} 1 \\ -\boldsymbol{\beta} \end{bmatrix}.$$

Proposition 19.3 establishes that the sample residuals associated with *OLS* estimation of [19.2.19] converge in probability to

$$\begin{aligned} y_{1t} - \hat{\alpha}_T - \hat{\gamma}_{2,T}y_{2t} - \hat{\gamma}_{3,T}y_{3t} - \cdots - \hat{\gamma}_{n,T}y_{nt} & \\ \xrightarrow{p} y_{1t} - [1 \quad -\boldsymbol{\beta}']\boldsymbol{\mu}_1^* - \boldsymbol{\beta}'\begin{bmatrix} y_{2t} \\ y_{3t} \\ \vdots \\ y_{ht} \end{bmatrix} - [1 \quad -\boldsymbol{\beta}']\boldsymbol{\Gamma}'\begin{bmatrix} y_{h+1,t} \\ y_{h+2,t} \\ \vdots \\ y_{nt} \end{bmatrix} & \\ = [1 \quad -\boldsymbol{\beta}']\cdot\{\mathbf{y}_{1t} - \boldsymbol{\mu}_1^* - \boldsymbol{\Gamma}'\mathbf{y}_{2t}\} & \\ = [1 \quad -\boldsymbol{\beta}']\cdot\mathbf{z}_t^*, & \end{aligned}$$

with the last equality following from [19.2.16]. But from [19.2.18] these are the same as the population residuals associated with the linear projection of z_{1t}^* on $z_{2t}^*, z_{3t}^*, \ldots, z_{ht}^*$.

This is an illustration of a general property observed by Wooldridge (1991). Consider a regression model of the form

$$y_t = \alpha + \mathbf{x}_t'\boldsymbol{\beta} + u_t. \qquad [19.2.22]$$

If y_t and $\mathbf{x}_t$ are $I(0)$, then $\alpha + \mathbf{x}_t'\boldsymbol{\beta}$ was said to be the linear projection of y_t on $\mathbf{x}_t$ and a constant if the population residual $u_t = y_t - \alpha - \mathbf{x}_t'\boldsymbol{\beta}$ has mean zero and is uncorrelated with $\mathbf{x}_t$. We saw that in such a case *OLS* estimation of [19.2.22] would typically yield consistent estimates of these linear projection coefficients. In the more general case where y_t can be $I(0)$ or $I(1)$ and elements of $\mathbf{x}_t$ can be $I(0)$ or $I(1)$, the analogous condition is that the residual $u_t = y_t - \alpha - \mathbf{x}_t'\boldsymbol{\beta}$ is a zero-mean stationary process that is uncorrelated with all $I(0)$ linear combinations of $\mathbf{x}_t$. Then $\alpha + \mathbf{x}_t'\boldsymbol{\beta}$ can be viewed as the $I(1)$ generalization of a population linear projection of y_t on a constant and $\mathbf{x}_t$. As long as there is some value for $\boldsymbol{\beta}$ such that $y_t - \mathbf{x}_t'\boldsymbol{\beta}$ is $I(0)$, such a linear projection $\alpha + \mathbf{x}_t'\boldsymbol{\beta}$ exists, and *OLS* estimation of [19.2.22] should give a consistent estimate of this projection.

What Is the Regression Estimating When There Is No Cointegrating Relation?

We have seen that if there is at least one cointegrating relation involving y_{1t}, then *OLS* estimation of [19.2.19] gives a consistent estimate of a cointegrating vector. Let us now consider the properties of *OLS* estimation when there is no cointegrating relation. Then [19.2.19] is a regression of an $I(1)$ variable on a set of $(n - 1)$ $I(1)$ variables for which no coefficients produce an $I(0)$ error term. The regression is therefore subject to the spurious regression problem described in Section 18.3. The coefficients $\hat{\alpha}_T$ and $\hat{\gamma}_T$ do not provide consistent estimates of any population parameters, and the *OLS* sample residuals $\hat{u}_t$ will be nonstationary. However, this last property can be exploited to test for cointegration. If there is no cointegration, then a regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ should yield a unit coefficient. If there is cointegration, then a regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ should yield a coefficient that is less than 1.

The proposal is thus to estimate [19.2.19] by *OLS* and then construct one of the standard unit root tests on the estimated residuals, such as the augmented Dickey-Fuller t test or the Phillips Z_ρ or Z_t test. Although these test statistics are constructed in the same way as when they are applied to an individual series y_t, when the tests are applied to the residuals $\hat{u}_t$ from a spurious regression, the critical values that are used to interpret the test statistics are different from those employed in Chapter 17.

Specifically, let $\mathbf{y}_t$ be an $(n \times 1)$ vector partitioned as

$$\underset{(n\times 1)}{\mathbf{y}_t} = \begin{bmatrix} \underset{(1\times 1)}{y_{1t}} \\ \underset{(g\times 1)}{\mathbf{y}_{2t}} \end{bmatrix} \qquad [19.2.23]$$

for $g \equiv (n - 1)$. Consider the regression

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + u_t. \qquad [19.2.24]$$

Let $\hat{u}_t$ be the sample residual associated with *OLS* estimation of [19.2.24] in a sample of size T:

$$\hat{u}_t = y_{1t} - \hat{\alpha}_T - \hat{\boldsymbol{\gamma}}_T'\mathbf{y}_{2t} \qquad \text{for } t = 1, 2, \ldots, T, \qquad [19.2.25]$$

where

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

and where Σ indicates summation over t from 1 to T. The residual $\hat{u}_t$ can then be regressed on its own lagged value $\hat{u}_{t-1}$ without a constant term:

$$\hat{u}_t = \rho\hat{u}_{t-1} + e_t \qquad \text{for } t = 2, 3, \ldots, T, \qquad [19.2.26]$$

yielding the estimate

$$\hat{\rho}_T = \frac{\sum_{t=2}^{T} \hat{u}_{t-1}\hat{u}_t}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}. \qquad [19.2.27]$$

Let s_T^2 be the *OLS* estimate of the variance of e_t for the regression of [19.2.26]:

$$s_T^2 = (T - 2)^{-1} \sum_{t=2}^{T} (\hat{u}_t - \hat{\rho}_T\hat{u}_{t-1})^2, \qquad [19.2.28]$$

and let $\hat{\sigma}_{\hat{\rho}_T}$ be the standard error of $\hat{\rho}_T$ as calculated by the usual *OLS* formula:

$$\hat{\sigma}_{\hat{\rho}_T}^2 = s_T^2 \div \left\{\sum_{t=2}^{T} \hat{u}_{t-1}^2\right\}. \qquad [19.2.29]$$

Finally, let $\hat{c}_{j,T}$ be the jth sample autocovariance of the estimated residuals associated with [19.2.26]:

$$\hat{c}_{j,T} = (T-1)^{-1} \sum_{t=j+2}^{T} \hat{e}_t \hat{e}_{t-j} \qquad \text{for } j = 0, 1, 2, \ldots, T-2 \qquad [19.2.30]$$

for $\hat{e}_t \equiv \hat{u}_t - \hat{\rho}_T \hat{u}_{t-1}$; and let the square of $\hat{\lambda}_T$ be given by

$$\hat{\lambda}_T^2 = \hat{c}_{0,T} + 2 \cdot \sum_{j=1}^{q} [1 - j/(q+1)] \hat{c}_{j,T}, \qquad [19.2.31]$$

where q is the number of autocovariances to be used. Phillips's Z_ρ statistic (1987) can be calculated just as in [17.6.8]:

$$Z_{\rho,T} = (T-1)(\hat{\rho}_T - 1) - (1/2) \cdot \{(T-1)^2 \cdot \hat{\sigma}_{\hat{\rho}_T}^2 \div s_T^2\} \cdot \{\hat{\lambda}_T^2 - \hat{c}_{0,T}\}. \qquad [19.2.32]$$

However, the asymptotic distribution of this statistic is not the expression in [17.6.8] but instead is a distribution that will be described in Proposition 19.4.

If the vector $\mathbf{y}_t$ is not cointegrated, then [19.2.24] will be a spurious regression and $\hat{\rho}_T$ should be near 1. On the other hand, if we find that $\hat{\rho}_T$ is well below 1—that is, if calculation of [19.2.32] yields a negative number that is sufficiently large in absolute value—then the null hypothesis that [19.2.24] is a spurious regression should be rejected, and we would conclude that the variables are cointegrated.

Similarly, Phillips's Z_t statistic associated with the residual autoregression [19.2.26] would be

$$Z_{t,T} = (\hat{c}_{0,T}/\hat{\lambda}_T^2)^{1/2} \cdot t_T - (1/2) \cdot \{(T-1) \cdot \hat{\sigma}_{\hat{\rho}_T} \div s_T\} \cdot \{\hat{\lambda}_T^2 - \hat{c}_{0,T}\}/\hat{\lambda}_T \qquad [19.2.33]$$

for t_T the usual *OLS* t statistic for testing the hypothesis $\rho = 1$:

$$t_T = (\hat{\rho}_T - 1)/\hat{\sigma}_{\hat{\rho}_T}.$$

Alternatively, lagged changes in the residuals could be added to the regression of [19.2.26] as in the augmented Dickey-Fuller test with no constant term:

$$\hat{u}_t = \zeta_1 \Delta\hat{u}_{t-1} + \zeta_2 \Delta\hat{u}_{t-2} + \cdots + \zeta_{p-1} \Delta\hat{u}_{t-p+1} + \rho\hat{u}_{t-1} + e_t. \qquad [19.2.34]$$

Again, this is estimated by *OLS* for $t = p+1, p+2, \ldots, T$, and the *OLS* t test of $\rho = 1$ is calculated using the standard *OLS* formula [8.1.26]. If this t statistic or the Z_t statistic in [19.2.33] is negative and sufficiently large in absolute value, this again casts doubt on the null hypothesis of no cointegration.

The following proposition, adapted from Phillips and Ouliaris (1990), provides a formal statement of the asymptotic distributions of these three test statistics.

Proposition 19.4: *Consider an $(n \times 1)$ vector $\mathbf{y}_t$ such that*

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

for $\boldsymbol{\varepsilon}_t$ an i.i.d. sequence with mean zero, variance $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{PP}'$, and finite fourth moments, and where $\{s \cdot \mathbf{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable. Let $g \equiv n - 1$ and $\mathbf{\Lambda} \equiv \mathbf{\Psi}(1) \cdot \mathbf{P}$. Suppose that the $(n \times n)$ matrix $\mathbf{\Lambda\Lambda}'$ is nonsingular, and let $\mathbf{L}$ denote the Cholesky factor of $(\mathbf{\Lambda\Lambda}')^{-1}$:

$$(\mathbf{\Lambda\Lambda}')^{-1} = \mathbf{LL}'. \qquad [19.2.35]$$

Then the following hold:

(*a*) *The statistic $\hat{\rho}_T$ defined in [19.2.27] satisfies*

$$(T-1)(\hat{\rho}_T - 1) \xrightarrow{L} \left\{\frac{1}{2}\left\{[1 \quad -\mathbf{h}_2']\cdot[\mathbf{W}^*(1)]\cdot[\mathbf{W}^*(1)]'\begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix}\right\}\right.$$
$$- h_1[\mathbf{W}^*(1)]'\begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix} \tag{19.2.36}$$
$$\left. - \frac{1}{2}[1 \quad -\mathbf{h}_2']\mathbf{L}'\{E(\Delta\mathbf{y}_t)(\Delta\mathbf{y}_t')\}\mathbf{L}\begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix}\right\} \div H_n.$$

Here, $\mathbf{W}^*(r)$ *denotes n-dimensional standard Brownian motion partitioned as*

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

h_1 *is a scalar and* $\mathbf{h}_2$ *a* $(g \times 1)$ *vector given by*

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

where the integral sign indicates integration over r from 0 *to* 1; *and*

$$H_n \equiv \int [W_1^*(r)]^2\,dr - \left[\int W_1^*(r)\,dr \quad \int [W_1^*(r)]\cdot[\mathbf{W}_2^*(r)]'\,dr\right]\begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix}.$$

(*b*) *If* $q \to \infty$ *as* $T \to \infty$ *but* $q/T \to 0$, *then the statistic* $Z_{\rho,T}$ *in* [19.2.32] *satisfies*

$$Z_{\rho,T} \xrightarrow{L} Z_n, \tag{19.2.37}$$

where

$$Z_n \equiv \left\{\frac{1}{2}\left\{[1 \quad -\mathbf{h}_2']\cdot[\mathbf{W}^*(1)]\cdot[\mathbf{W}^*(1)]'\begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix}\right\}\right. \tag{19.2.38}$$
$$\left. - h_1[\mathbf{W}^*(1)]'\begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix} - \frac{1}{2}(1 + \mathbf{h}_2'\mathbf{h}_2)\right\} \div H_n.$$

(*c*) *If* $q \to \infty$ *as* $T \to \infty$ *but* $q/T \to 0$, *then the statistic* $Z_{t,T}$ *in* [19.2.33] *satisfies*

$$Z_{t,T} \xrightarrow{L} Z_n\cdot\sqrt{H_n} \div (1 + \mathbf{h}_2'\mathbf{h}_2)^{1/2}. \tag{19.2.39}$$

(*d*) *If, in addition to the preceding assumptions,* $\Delta\mathbf{y}_t$ *follows a zero-mean stationary vector ARMA process and if* $p \to \infty$ *as* $T \to \infty$ *but* $p/T^{1/3} \to 0$, *then the augmented Dickey-Fuller t test associated with* [19.2.34] *has the same limiting distribution* Z_n *as the test statistic* $Z_{\rho,T}$ *described in* [19.2.37].

Result (a) implies that $\hat{\rho}_T \xrightarrow{p} 1$. Hence, when the estimated "cointegrating" regression [19.2.24] is spurious, the estimated residuals from this regression behave like a unit root process in the sense that if $\hat{u}_t$ is regressed on $\hat{u}_{t-1}$, the estimated coefficient should tend to unity as the sample size grows. No linear combination of $\mathbf{y}_t$ is stationary, and so the residuals from the spurious regression cannot be stationary.

Note that since $W_1^*(r)$ and $\mathbf{W}_2^*(r)$ are standard Brownian motion, the distributions of the terms h_1, $\mathbf{h}_2$, H_n, and Z_n in Proposition 19.4 depend only on the number of stochastic explanatory variables included in the cointegrating regression $(n - 1)$ and on whether a constant term appears in that regression but are not affected by the variances, correlations, and dynamics of $\Delta\mathbf{y}_t$.

In the special case when $\Delta\mathbf{y}_t$ is i.i.d., then $\mathbf{\Psi}(L) = \mathbf{I}_n$ and the matrix $\mathbf{\Lambda\Lambda}' = E[(\Delta\mathbf{y}_t)(\Delta\mathbf{y}_t')]$. Since $\mathbf{LL}' = (\mathbf{\Lambda\Lambda}')^{-1}$, it follows that $(\mathbf{\Lambda\Lambda}') = (\mathbf{L}')^{-1}(\mathbf{L})^{-1}$. Hence, for this special case,

$$\mathbf{L}'\{E[(\Delta\mathbf{y}_t)(\Delta\mathbf{y}_t')]\}\mathbf{L} = \mathbf{L}'(\mathbf{\Lambda\Lambda}')\mathbf{L} = \mathbf{L}'\{(\mathbf{L}')^{-1}(\mathbf{L})^{-1}\}\mathbf{L} = \mathbf{I}_n. \quad [19.2.40]$$

If [19.2.40] is substituted into [19.2.36], the result is that when $\Delta\mathbf{y}_t$ is i.i.d.,

$$(T - 1)(\hat{\rho}_T - 1) \xrightarrow{L} Z_n$$

for Z_n defined in [19.2.38].

In the more general case when $\Delta\mathbf{y}_t$ is serially correlated, the limiting distribution of $T(\hat{\rho}_T - 1)$ depends on the nature of this correlation as captured by the elements of $\mathbf{L}$. However, the corrections for autocorrelation implicit in Phillips's Z_ρ and Z_t statistics or the augmented Dickey-Fuller t test turn out to generate variables whose distributions do not depend on any nuisance parameters.

Although the distributions of Z_ρ, Z_t, and the augmented Dickey-Fuller t test do not depend on nuisance parameters, the distributions when these statistics are calculated from the residuals $\hat{u}_t$ are not the same as the distributions these statistics would have if calculated from the raw data $\mathbf{y}_t$. Moreover, different values for $n - 1$ (the number of stochastic explanatory variables in the cointegrating regression of [19.2.24]) imply different characterizations of the limiting statistics h_1, $\mathbf{h}_2$, H_n, and Z_n, meaning that a different critical value must be used to interpret Z_ρ for each value of $n - 1$. Similarly, the asymptotic distributions of $\mathbf{h}_2$, H_n, and Z_n are different depending on whether a constant term is included in the cointegrating regression [19.2.24].

The section labeled Case 1 in Table B.8 refers to the case when the cointegrating regression is estimated without a constant term:

$$y_{1t} = \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t. \quad [19.2.41]$$

The table reports Monte Carlo estimates of the critical values for the test statistic Z_ρ described in [19.2.32], for $\hat{u}_t$ the date t residual from OLS estimation of [19.2.41]. The values were calculated by generating a sample of size $T = 500$ for y_{1t}, y_{2t}, $\ldots$, y_{nt} independent Gaussian random walks, estimating [19.2.41] and [19.2.26] by OLS, and tabulating the distribution of $(T - 1)(\hat{\rho}_T - 1)$. For example, the table indicates that if we were to regress a random walk y_{1t} on three other random walks (y_{2t}, y_{3t}, and y_{4t}), then in 95% of the samples, $(T - 1)(\hat{\rho}_T - 1)$ would be greater than -27.9, that is, $\hat{\rho}_T$ should exceed 0.94 in a sample of size $T = 500$. If the estimate $\hat{\rho}_T$ is below 0.94, then this might be taken as evidence that the series are cointegrated.

The section labeled Case 2 in Table B.8 gives critical values for $Z_{\rho,T}$ when a constant term is included in the cointegrating regression:

$$y_{1t} = \alpha + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t. \quad [19.2.42]$$

For this case, [19.2.26] is estimated with $\hat{u}_t$ now interpreted as the residual from

OLS estimation of [19.2.42]. Note that the different cases (1 and 2) refer to whether a constant term is included in the cointegrating regression [19.2.42] and not to whether a constant term is included in the residual regression [19.2.26]. In each case, the autoregression for the residuals is estimated in the form of [19.2.26] with no constant term.

Critical values for the Z_t statistic or the augmented Dickey-Fuller t statistic are reported in Table B.9. Again, if no constant term is included in the cointegrating regression as in [19.2.41], the case 1 entries are appropriate, whereas if a constant term is included in the cointegrating regression as in [19.2.42], the case 2 entries should be used. If the value for the Z_t or augmented Dickey-Fuller t statistic is negative and large in absolute value, this is evidence against the null hypothesis that $\mathbf{y}_t$ is not cointegrated.

When the corrections for serial correlation implicit in the Z_ρ, Z_t, or augmented Dickey-Fuller test are used, the justification for using the critical values in Table B.8 or B.9 is asymptotic, and accordingly these tables describe only the large-sample distribution. Small-sample critical values tabulated by Engle and Yoo (1987) and Haug (1992) can differ somewhat from the large-sample critical values.

Testing for Cointegration Among Trending Series

It was assumed in Proposition 19.4 that $E(\Delta \mathbf{y}_t) = \mathbf{0}$, in which case none of the series would exhibit nonzero drift. Bruce Hansen (1992) described how the results change if instead $E(\Delta \mathbf{y}_t)$ contains one or more nonzero elements.

Consider first the case $n = 2$, a regression of one scalar on another:

$$y_{1t} = \alpha + \gamma y_{2t} + u_t. \qquad [19.2.43]$$

Suppose that

$$\Delta y_{2t} = \delta_2 + u_{2t}$$

with $\delta_2 \neq 0$. Then

$$y_{2t} = y_{2,0} + \delta_2 \cdot t + \sum_{s=1}^{t} u_{2s},$$

which is asymptotically dominated by the deterministic time trend $\delta_2 \cdot t$. Thus, estimates $\hat{\alpha}_T$ and $\hat{\gamma}_T$ based on *OLS* estimation of [19.2.43] have the same asymptotic distribution as the coefficients in a regression of an $I(1)$ series on a constant and a time trend. If

$$\Delta y_{1t} = \delta_1 + u_{1t}$$

(where δ_1 may be zero), then the *OLS* estimate $\hat{\gamma}_T$ based on [19.2.43] gives a consistent estimate of (δ_1/δ_2), and the first difference of the residuals from that regression converges to $u_{1t} - (\delta_1/\delta_2)u_{2t}$; see Exercise 19.1.

If, in fact, [19.2.43] were a simple time trend regression of the form

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

then an augmented Dickey-Fuller test on the residuals,

$$\hat{u}_t = \zeta_1 \Delta\hat{u}_{t-1} + \zeta_2 \Delta\hat{u}_{t-2} + \cdots + \zeta_{p-1}\Delta\hat{u}_{t-p+1} + \rho\hat{u}_{t-1} + e_t, \qquad [19.2.44]$$

would be asymptotically equivalent to an augmented Dickey-Fuller test on the original series y_{1t} that included a constant term and a time trend:

$$\begin{aligned} y_{1t} = {} & \zeta_1 \Delta y_{1,t-1} + \zeta_2 \Delta y_{1,t-2} + \cdots + \zeta_{p-1}\Delta y_{1,t-p+1} \\ & + \alpha + \rho y_{1,t-1} + \delta t + u_t. \end{aligned} \qquad [19.2.45]$$

Since the residuals from OLS estimation of [19.2.43] behave like the residuals from a regression of $[y_{1t} - (\delta_1/\delta_2)y_{2t}]$ on a time trend, Hansen (1992) showed that when y_{2t} has a nonzero trend, the t test of $\rho = 1$ in [19.2.44] for $\hat{u}_t$ the residual from OLS estimation of [19.2.43] has the same asymptotic distribution as the usual augmented Dickey-Fuller t test for a regression of the form of [19.2.45] with y_{1t} replaced by $[y_{1t} - (\delta_1/\delta_2)y_{2t}]$. Thus, if the cointegrating regression involves a single variable y_{2t} with nonzero drift, we estimate the regression [19.2.43] and calculate the Z_t or augmented Dickey-Fuller t statistic in exactly the same manner that was specified in equation [19.2.33] or [19.2.34]. However, rather than compare these statistics with the $(n - 1) = 1$ entry for case 2 from Table B.9, we instead compare these statistics with the case 4 section of Table B.6.

For convenience, the values for a sample of size $T = 500$ for the univariate case 4 section of Table B.6 are reproduced in the $(n - 1) = 1$ row of the section labeled Case 3 in Table B.9. This is described as case 3 in the multivariate tabulations for the following reason. In the univariate analysis, "case 3" referred to a regression in which the single variable y_t had a nonzero trend but no trend term was included in the regression. The multivariate generalization obtains when the explanatory variable y_{2t} has a nonzero trend but no trend is included in the regression [19.2.43]. The asymptotic distribution that describes the residuals from that regression is the same as that for a univariate regression in which a trend is included.

Similarly, if y_{2t} has a nonzero trend, we can estimate [19.2.43] by OLS and construct Phillips's Z_ρ statistic exactly as in equation [19.2.32] and compare this with the values tabulated in the case 4 portion of Table B.5. These numbers are reproduced in row $(n - 1) = 1$ of the case 3 section of Table B.8.

More generally, consider a regression involving $n - 1$ stochastic explanatory variables of the form of [19.2.42]. Let δ_i denote the trend in the ith variable:

$$E(\Delta y_{it}) = \delta_i.$$

Suppose that at least one of the explanatory variables has a nonzero trend component; for illustration, call this the nth variable:

$$\delta_n \neq 0.$$

Whether or not other explanatory variables or the dependent variable also have nonzero trends turns out not to matter for the asymptotic distribution; that is, the values of $\delta_1, \delta_2, \ldots, \delta_{n-1}$ are irrelevant given that $\delta_n \neq 0$.

Note that the fitted values of [19.2.42] are identical to the fitted values from OLS estimation of

$$y_{1t}^* = \alpha^* + \gamma_2^* y_{2t}^* + \gamma_3^* y_{3t}^* + \cdots + \gamma_{n-1}^* y_{n-1,t}^* + \gamma_n^* y_{nt} + u_t, \quad [19.2.46]$$

where

$$y_{it}^* \equiv y_{it} - (\delta_i/\delta_n)y_{nt} \qquad \text{for } i = 1, 2, \ldots, n - 1.$$

As in the analysis of [18.2.44], moments involving y_{nt} are dominated by the time trend $\delta_n t$, while the y_{it}^* are driftless $I(1)$ variables for $i = 1, 2, \ldots, n - 1$. Thus, the residuals from [19.2.46] have the same asymptotic properties as the residuals from OLS estimation of

$$y_{1t}^* = \alpha^* + \gamma_2^* y_{2t}^* + \gamma_3^* y_{3t}^* + \cdots + \gamma_{n-1}^* y_{n-1,t}^* + \gamma_n^* \delta_n t + u_t. \quad [19.2.47]$$

The appropriate critical values for statistics constructed when $\hat{u}_t$ denotes the residual from OLS estimation of [19.2.42] can therefore be calculated from those for an OLS regression of an $I(1)$ variable on a constant, $(n - 2)$ other $I(1)$ variables, and a time trend. The appropriate critical values are tabulated under the heading Case 3 in Tables B.8 and B.9.

Of course, we could instead imagine including a time trend directly in the regression, as in

$$y_{1t} = \alpha + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + \delta t + u_t. \qquad [19.2.48]$$

Since [19.2.48] is in the same form as the regression of [19.2.47], critical values for such a regression could be found by treating this as if it were a regression involving $(n + 1)$ variables and looking in the case 3 section of Table B.8 or B.9 for the critical values that would be appropriate if we actually had $(n + 1)$ rather than n total variables. Clearly, the specification in [19.2.42] has more power to reject a false null hypothesis than [19.2.48], since we would use the same table of critical values for [19.2.42] or [19.2.48] with one more degree of freedom used up by [19.2.48]. Conceivably, we might still want to estimate the regression in the form of [19.2.48] to cover the case when we are not sure whether any of the elements of $\mathbf{y}_t$ have a nonzero trend or not.

Summary of Residual-Based Tests for Cointegration

The Phillips-Ouliaris-Hansen procedure for testing for cointegration is summarized in Table 19.1.

To illustrate this approach, consider again the purchasing power parity example where p_t is the log of the U.S. price level, s_t is the log of the dollar-lira exchange rate, and p_t^* is the log of the Italian price level. We have already seen that the vector $\mathbf{a} = (1, -1, -1)'$ does not appear to be a cointegrating vector for $\mathbf{y}_t = (p_t, s_t, p_t^*)'$. Let us now ask whether there is any cointegrating relation among these variables.

The following regression was estimated by *OLS* for t = 1973:1 to 1989:10 (standard errors in parentheses):

$$p_t = \underset{(0.37)}{2.71} + \underset{(0.012)}{0.051}\, s_t + \underset{(0.0067)}{0.5300}\, p_t^* + \hat{u}_t. \qquad [19.2.49]$$

The number of observations used to estimate [19.2.49] is $T = 202$. When the sample residuals $\hat{u}_t$ are regressed on their own lagged values, the result is

$$\hat{u}_t = \underset{(0.01172)}{0.98331}\, \hat{u}_{t-1} + \hat{e}_t$$

$$s^2 = (T - 2)^{-1} \sum_{t=2}^{T} \hat{e}_t^2 = (0.40374)^2$$

$$\hat{c}_0 = 0.1622$$

$$\hat{c}_j = (T - 1)^{-1} \sum_{t=j+2}^{T} \hat{e}_t \hat{e}_{t-j}$$

$$\hat{\lambda}^2 = \hat{c}_0 + 2 \cdot \sum_{j=1}^{12} [1 - (j/13)]\hat{c}_j = 0.4082.$$

The Phillips-Ouliaris Z_ρ test is

$$\begin{aligned} Z_\rho &= (T - 1)(\hat{\rho} - 1) - (1/2)\{(T - 1)\cdot\hat{\sigma}_{\hat{\rho}} \div s\}^2(\hat{\lambda}^2 - \hat{c}_0) \\ &= (201)(0.98331 - 1) \\ &\quad - \tfrac{1}{2}\{(201)(0.01172) \div (0.40374)\}^2(0.4082 - 0.1622) \\ &= -7.54. \end{aligned}$$

Given the evidence of nonzero drift in the explanatory variables, this is to be compared with the case 3 section of Table B.8. For $(n - 1) = 2$, the 5% critical

TABLE 19.1
Summary of Phillips-Ouliaris-Hansen Tests for Cointegration

Case 1:

Estimated cointegrating regression:

$$y_{1t} = \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t$$

True process for $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$:

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

Z_ρ has the same asymptotic distribution as the variable described under the heading Case 1 in Table B.8.

Z_t and the augmented Dickey-Fuller t test have the same asymptotic distribution as the variable described under Case 1 in Table B.9.

Case 2:

Estimated cointegrating regression:

$$y_{1t} = \alpha + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t$$

True process for $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$:

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

Z_ρ has the same asymptotic distribution as the variable described under Case 2 in Table B.8.

Z_t and the augmented Dickey-Fuller t test have the same asymptotic distribution as the variable described under Case 2 in Table B.9.

Case 3:

Estimated cointegrating regression:

$$y_{1t} = \alpha + \gamma_2 y_{2t} + \gamma_3 y_{3t} + \cdots + \gamma_n y_{nt} + u_t$$

True process for $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$:

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

with at least one element of $\delta_2, \delta_3, \ldots, \delta_n$ nonzero.

Z_ρ has the same asymptotic distribution as the variable described under Case 3 in Table B.8.

Z_t and the augmented Dickey-Fuller t test have the same asymptotic distribution as the variable described under Case 3 in Table B.9.

Notes to Table 19.1

Estimated cointegrating regression indicates the form in which the regression that could describe the cointegrating relation is estimated, using observations $t = 1, 2, \ldots, T$.

True process describes the null hypothesis under which the distribution is calculated. In each case, $\boldsymbol{\varepsilon}_t$ is assumed to be i.i.d. with mean zero, positive definite variance-covariance matrix, and finite fourth moments, and the sequence $\{s \cdot \boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable. The matrix $\boldsymbol{\Psi}(1)$ is assumed to be nonsingular, meaning that the vector $\mathbf{y}_t$ is not cointegrated under the null hypothesis. If the test statistic is below the indicated critical value (that is, if Z_ρ, Z_t, or t is negative and sufficiently large in absolute value), then the null hypothesis of no cointegration is rejected.

Z_ρ is the following statistic,

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

where $\hat{\rho}_T$ is the estimate of ρ based on OLS estimation of $\hat{u}_t = \rho \hat{u}_{t-1} + e_t$ for $\hat{u}_t$ the OLS sample residual value for Z_ρ is -27.1. Since $-7.54 > -27.1$, the null hypothesis of no cointegration is accepted. Similarly, the Phillips-Ouliaris Z_t statistic is

$$\begin{aligned} Z_t &= (\hat{c}_0/\hat{\lambda}^2)^{1/2}(\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}} - (1/2)\{(T-1)\cdot\hat{\sigma}_{\hat{\rho}} \div s\}(\hat{\lambda}^2 - \hat{c}_0)/\hat{\lambda} \\ &= \{(0.1622)/(0.4082)\}^{1/2}(0.98331 - 1)/(0.01172) \\ &\quad - \tfrac{1}{2}\{(201)(0.01172) \div (0.40374)\}(0.4082 - 0.1622)/(0.4082)^{1/2} \\ &= -2.02. \end{aligned}$$

Comparing this with the case 3 section of Table B.9, we see that $-2.02 > -3.80$, so that the null hypothesis of no cointegration is also accepted by this test. An *OLS* regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ and twelve lags of $\Delta\hat{u}_{t-j}$ produces an *OLS* t test of $\rho = 1$ of -2.73, which is again above -3.80. We thus find little evidence that p_t, s_t, and p_t^* are cointegrated. Indeed, the regression [19.2.49] displays the classic symptoms of a spurious regression—the estimated standard errors are small relative to the coefficient estimates, and the estimated first-order autocorrelation of the residuals is near unity.

As a second example, Figure 19.5 plots 100 times the logs of real quarterly aggregate personal disposable income (y_t) and personal consumption expenditures (c_t) for the United States over 1947:I to 1989:III. In a regression of y_t on a constant, a time trend, y_{t-1}, and Δy_{t-j} for $j = 1, 2, \ldots, 6$, the *OLS* t test that the coefficient on y_{t-1} is unity is -1.28. Similarly, in a regression of c_t on a constant, a time trend, c_{t-1}, and Δc_{t-j} for $j = 1, 2, \ldots, 6$, the *OLS* t test that the coefficient on c_{t-1} is unity is -1.88. Thus, both processes might well be described as $I(1)$ with positive drift.

The *OLS* estimate of the cointegrating relation is

$$c_t = \underset{(2.35)}{0.67} + \underset{(0.0032)}{0.9865}\, y_t + u_t. \qquad [19.2.50]$$

A first-order autoregression fitted to the residuals produces

$$\hat{u}_t = \underset{(0.048)}{0.782}\, \hat{u}_{t-1} + \hat{e}_t,$$

Notes to Table 19.1 (continued).

from the estimated regression. Here,

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

where $\hat{e}_t = \hat{u}_t - \hat{\rho}_T\hat{u}_{t-1}$ is the sample residual from the autoregression describing $\hat{u}_t$ and $\hat{\sigma}_{\hat{\rho}_T}$ is the standard error for $\hat{\rho}_T$ as calculated by the usual *OLS* formula:

$$\hat{\sigma}_{\hat{\rho}_T}^2 = s_T^2 \div \sum_{t=2}^{T} \hat{u}_{t-1}^2.$$

Also,

$$\hat{c}_{j,T} = (T-1)^{-1} \sum_{t=j+2}^{T} \hat{e}_t\hat{e}_{t-j}$$

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

Z_t is the following statistic:

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

Augmented Dickey-Fuller t statistic is the *OLS* t test of the null hypothesis that $\rho = 1$ in the regression

$$\hat{u}_t = \zeta_1\Delta\hat{u}_{t-1} + \zeta_2\Delta\hat{u}_{t-2} + \cdots + \zeta_{p-1}\Delta\hat{u}_{t-p+1} + \rho\hat{u}_{t-1} + e_t.$$

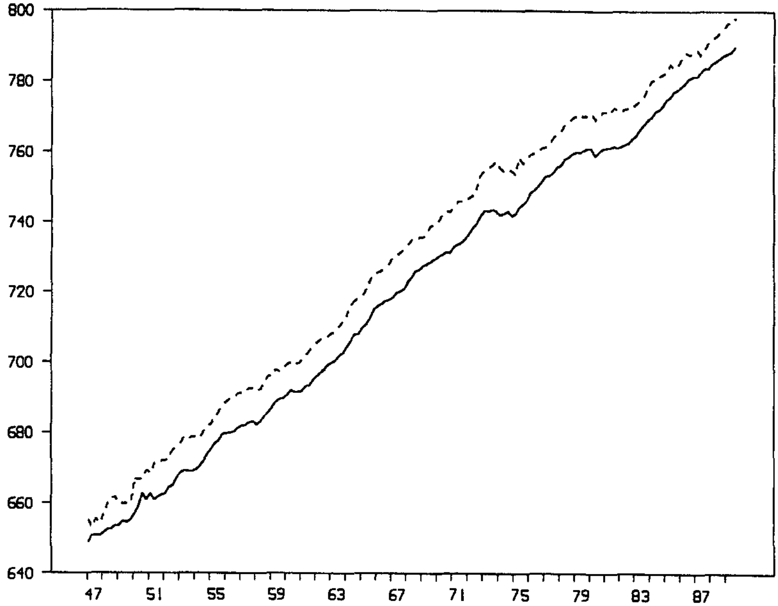


FIGURE 19.5 One hundred times the log of personal consumption expenditures (c_t) and personal disposable income (y_t) for the United States in billions of 1982 dollars, quarterly, 1947–89. Key: ——— c_t; ---- y_t.

for which the corresponding Z_ρ and Z_t statistics for $q = 6$ are -32.0 and -4.28. Since there is again ample evidence that y_t has positive drift, these are to be compared with the case 3 sections of Tables B.8 and B.9, respectively. Since $-32.0 < -21.5$ and $-4.28 < -3.42$, in each case the null hypothesis of no cointegration is rejected at the 5% level. Thus consumption and income appear to be cointegrated.

Other Tests for Cointegration

The tests that have been discussed in this section are based on the residuals from an *OLS* regression of y_{1t} on $(y_{2t}, y_{3t}, \ldots, y_{nt})$. Since these are not the same as the residuals from a regression of y_{2t} on $(y_{1t}, y_{3t}, \ldots, y_{nt})$, the tests can give different answers depending on which variable is labeled y_1. Important tests for cointegration that are invariant to the ordering of variables are the full-information maximum likelihood test of Johansen (1988, 1991) and the related tests of Stock and Watson (1988) and Ahn and Reinsel (1990). These will be discussed in Chapter 20. Other useful tests for cointegration have been proposed by Phillips and Ouliaris (1990), Park, Ouliaris, and Choi (1988), Stock (1990), and Hansen (1990).

19.3. *Testing Hypotheses About the Cointegrating Vector*

The previous section described some ways to test whether a vector $\mathbf{y}_t$ is cointegrated. It was noted that if $\mathbf{y}_t$ is cointegrated, then a consistent estimate of the cointegrating vector can be obtained by *OLS*. This section explores further the distribution theory of this estimate and proposes several alternative estimates that simplify hypothesis testing.

Distribution of the OLS *Estimate for a Special Case*

Let y_{1t} be a scalar and $\mathbf{y}_{2t}$ be a $(g \times 1)$ vector satisfying

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + z_t^* \qquad [19.3.1]$$

$$\mathbf{y}_{2t} = \mathbf{y}_{2,t-1} + \mathbf{u}_{2t}. \qquad [19.3.2]$$

If y_{1t} and $\mathbf{y}_{2t}$ are both $I(1)$ but z_t^* and $\mathbf{u}_{2t}$ are $I(0)$, then, for $n \equiv (g + 1)$, the n-dimensional vector $(y_{1t}, \mathbf{y}_{2t}')'$ is cointegrated with cointegrating relation [19.3.1].

Consider the special case of a Gaussian system for which $\mathbf{y}_{2t}$ follows a random walk and for which z_t^* is white noise and uncorrelated with $\mathbf{u}_{2\tau}$ for all t and τ:

$$\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t} \end{bmatrix} \sim \text{i.i.d. } N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{bmatrix}\right). \qquad [19.3.3]$$

Then [19.3.1] describes a regression in which the explanatory variables $(\mathbf{y}_{2t})$ are independent of the error term (z_τ^*) for all t and τ. The regression thus satisfies Assumption 8.2 in Chapter 8. There it was seen that *conditional* on $(\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})$, the *OLS* estimates have a Gaussian distribution:

$$\begin{aligned} \left[\begin{matrix} (\hat{\alpha}_T - \alpha) \\ (\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{matrix}\middle| (\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})\right] &= \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \Sigma z_t^* \\ \Sigma\mathbf{y}_{2t}z_t^* \end{bmatrix} \\ &\sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \sigma_1^2 \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\right), \end{aligned} \qquad [19.3.4]$$

where Σ indicates summation over t from 1 to T.

Recall further from Chapter 8 that this conditional Gaussian distribution is all that is needed to justify small-sample application of the usual *OLS* t or F tests. Consider a hypothesis test involving m restrictions on α and $\boldsymbol{\gamma}$ of the form

$$\mathbf{R}_\alpha \alpha + \mathbf{R}_\gamma \boldsymbol{\gamma} = \mathbf{r},$$

where $\mathbf{R}_\alpha$ and $\mathbf{r}$ are known $(m \times 1)$ vectors and $\mathbf{R}_\gamma$ is a known $(m \times g)$ matrix describing the restrictions. The Wald form of the *OLS* F test of the null hypothesis is

$$\begin{aligned} (\mathbf{R}_\alpha\hat{\alpha}_T + \mathbf{R}_\gamma\hat{\boldsymbol{\gamma}}_T - \mathbf{r})' &\left\{ s_T^2 [\mathbf{R}_\alpha \;\; \mathbf{R}_\gamma] \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_\alpha' \\ \mathbf{R}_\gamma' \end{bmatrix} \right\}^{-1} \\ &\times (\mathbf{R}_\alpha\hat{\alpha}_T + \mathbf{R}_\gamma\hat{\boldsymbol{\gamma}}_T - \mathbf{r}) \div m, \end{aligned} \qquad [19.3.5]$$

where

$$s_T^2 = (T - n)^{-1} \sum_{t=1}^{T} (y_{1t} - \hat{\alpha}_T - \hat{\boldsymbol{\gamma}}_T'\mathbf{y}_{2t})^2.$$

Result [19.3.4] implies that conditional on $(\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})$, under the null hypothesis the vector $(\mathbf{R}_\alpha\hat{\alpha}_T + \mathbf{R}_\gamma\hat{\boldsymbol{\gamma}}_T - \mathbf{r})$ has a Gaussian distribution with mean $\mathbf{0}$ and variance

$$\sigma_1^2 [\mathbf{R}_\alpha \;\; \mathbf{R}_\gamma] \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_\alpha' \\ \mathbf{R}_\gamma' \end{bmatrix}.$$

It follows that conditional on $(\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})$, the term

$$(\mathbf{R}_\alpha \hat{\alpha}_T + \mathbf{R}_\gamma \hat{\gamma}_T - \mathbf{r})' \left\{ \sigma_1^2 [\mathbf{R}_\alpha \quad \mathbf{R}_\gamma] \begin{bmatrix} T & \Sigma \mathbf{y}_{2t}' \\ \Sigma \mathbf{y}_{2t} & \Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_\alpha' \\ \mathbf{R}_\gamma' \end{bmatrix} \right\}^{-1} \times (\mathbf{R}_\alpha \hat{\alpha}_T + \mathbf{R}_\gamma \hat{\gamma}_T - \mathbf{r}) \qquad [19.3.6]$$

is a quadratic form in a Gaussian vector. Proposition 8.1 establishes that conditional on $(\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})$, the magnitude in [19.3.6] has a $\chi^2(m)$ distribution. Thus, conditional on $(\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})$, the *OLS F* test [19.3.5] could be viewed as the ratio of a $\chi^2(m)$ variable to the independent $\chi^2(T - n)$ variable $(T - n)s_T^2/\sigma_1^2$, with numerator and denominator each divided by its degree of freedom. The *OLS* *F* test thus has an exact $F(m, T - n)$ conditional distribution. Since this is the same distribution for all realizations of $(\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})$, it follows that [19.3.5] has an unconditional $F(m, T - n)$ distribution as well. Hence, despite the $I(1)$ regressors and complications of cointegration, the correct approach for this example would be to estimate [19.3.1] by *OLS* and use standard *t* or *F* statistics to test any hypotheses about the cointegrating vector. No special procedures are needed to estimate the cointegrating vector, and no unusual critical values need be consulted to test a hypothesis about its value.

We now seek to make an analogous statement in terms of the corresponding asymptotic distributions. To do so it will be helpful to rescale the results in [19.3.4] and [19.3.5] so that they define sequences of statistics with nondegenerate asymptotic distributions. If [19.3.4] is premultiplied by the matrix

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

the implication is that the distribution of the *OLS* estimates conditional on $(\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})$ is given by

$$\left[\begin{matrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\gamma}_T - \gamma) \end{matrix} \middle| (\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T}) \right]$$

$$\sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \sigma_1^2 \left\{ \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T\cdot\mathbf{I}_g \end{bmatrix} \begin{bmatrix} T & \Sigma \mathbf{y}_{2t}' \\ \Sigma \mathbf{y}_{2t} & \Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T\cdot\mathbf{I}_g \end{bmatrix} \right\} \right) \qquad [19.3.7]$$

$$= N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \sigma_1^2 \begin{bmatrix} 1 & T^{-3/2}\Sigma \mathbf{y}_{2t}' \\ T^{-3/2}\Sigma \mathbf{y}_{2t} & T^{-2}\Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \right).$$

To analyze the asymptotic distribution, notice that [19.3.1] through [19.3.3] are a special case of the system analyzed in Proposition 19.2 with $\boldsymbol{\Psi}^*(L) = \mathbf{I}_n$ and

$$\mathbf{P} = \begin{bmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix},$$

where $\mathbf{P}_{22}$ is the Cholesky factor of $\boldsymbol{\Omega}_{22}$:

$$\boldsymbol{\Omega}_{22} = \mathbf{P}_{22}\mathbf{P}_{22}'.$$

For this special case,

$$\boldsymbol{\Psi}^*(1)\cdot\mathbf{P} = \begin{bmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix}. \qquad [19.3.8]$$

The terms $\boldsymbol{\lambda}_1^{*\prime}$ and $\boldsymbol{\Lambda}_2^*$ referred to in Proposition 19.2 would then be given by

$$\underset{(1\times n)}{\boldsymbol{\lambda}_1^{*\prime}} = \begin{bmatrix} \underset{(1\times 1)}{\sigma_1} & \underset{(1\times g)}{\mathbf{0}'} \end{bmatrix}$$

$$\underset{(g\times n)}{\boldsymbol{\Lambda}_2^*} = \begin{bmatrix} \underset{(g\times 1)}{\mathbf{0}} & \underset{(g\times g)}{\mathbf{P}_{22}} \end{bmatrix}.$$

Thus, result [19.2.13] of Proposition 19.2 establishes that

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma z_t^* \\ T^{-1}\Sigma\mathbf{y}_{2t}z_t^* \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} 1 & \left\{\int [\mathbf{W}(r)]'\,dr\right\}\begin{bmatrix}\mathbf{0}' \\ \mathbf{P}_{22}'\end{bmatrix} \\ [\mathbf{0} \quad \mathbf{P}_{22}]\int \mathbf{W}(r)\,dr & [\mathbf{0} \quad \mathbf{P}_{22}]\left\{\int [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\begin{bmatrix}\mathbf{0}' \\ \mathbf{P}_{22}'\end{bmatrix} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} [\sigma_1 \quad \mathbf{0}']\mathbf{W}(1) \\ [\mathbf{0} \quad \mathbf{P}_{22}]\left\{\int [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\begin{bmatrix}\sigma_1 \\ \mathbf{0}\end{bmatrix} \end{bmatrix}, \qquad [19.3.9]$$

where the integral sign indicates integration over r from 0 to 1. If the n-dimensional standard Brownian motion $\mathbf{W}(r)$ is partitioned as

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

then [19.3.9] can be written

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \\ \mathbf{P}_{22}\int \mathbf{W}_2(r)\,dr & \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \end{bmatrix}^{-1} \qquad [19.3.10]$$

$$\times \begin{bmatrix} \sigma_1 W_1(1) \\ \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)]\,dW_1(r)\right\}\sigma_1 \end{bmatrix}$$

$$\equiv \sigma_1 \begin{bmatrix} \nu_1 \\ \boldsymbol{\nu}_2 \end{bmatrix},$$

where

$$\begin{bmatrix} \nu_1 \\ \boldsymbol{\nu}_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \\ \mathbf{P}_{22}\int \mathbf{W}_2(r)\,dr & \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \end{bmatrix}^{-1} \qquad [19.3.11]$$

$$\times \begin{bmatrix} W_1(1) \\ \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)]\,dW_1(r)\right\} \end{bmatrix}.$$

Since $W_1(\cdot)$ is independent of $\mathbf{W}_2(\cdot)$, the distribution of $(\nu_1, \boldsymbol{\nu}_2')'$ conditional on $\mathbf{W}_2(\cdot)$ is found by treating $\mathbf{W}_2(r)$ as a deterministic function of r and leaving the process $W_1(\cdot)$ unaffected. Then $\int[\mathbf{W}_2(r)]\,dW_1(r)$ has a simple Gaussian distribution, and [19.3.11] describes a Gaussian vector. In particular, the exact finite-sample result for Gaussian disturbances [19.3.7] implied that

$$\left[\begin{matrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{matrix}\middle| (\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T})\right] = \begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} T^{-1/2}\Sigma z_t^* \\ T^{-1}\Sigma\mathbf{y}_{2t}z_t^* \end{bmatrix}$$

$$\sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \sigma_1^2\begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\right).$$

Comparing this with the limiting distribution [19.3.10], it appears that the vector $(\nu_1, \boldsymbol{\nu}_2')'$ has distribution conditional on $\mathbf{W}_2(\cdot)$ that could be described as

$$\left[\begin{matrix} \nu_1 \\ \boldsymbol{\nu}_2 \end{matrix}\middle| \mathbf{W}_2(\cdot)\right]$$

$$\sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \\ \mathbf{P}_{22}\int \mathbf{W}_2(r)\,dr & \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \end{bmatrix}^{-1}\right). \qquad [19.3.12]$$

Expression [19.3.12] allows the argument that was used to motivate the usual *OLS* t and F tests on the system of [19.3.1] and [19.3.2] with Gaussian disturbances satisfying [19.3.3] to give an asymptotic justification for these same tests in a system with non-Gaussian disturbances whose means and autocovariances are as assumed in [19.3.3]. Consider for illustration a hypothesis that involves only the cointegrating vector, so that $\mathbf{R}_\alpha = \mathbf{0}$. Then, under the null hypothesis, m times the F test in [19.3.5] becomes

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

$$= [\mathbf{R}_\gamma\cdot T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma})]'\left\{s_T^2[\mathbf{0} \quad \mathbf{R}_\gamma\cdot T]\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ T\cdot\mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1}$$
$$\times [\mathbf{R}_\gamma\cdot T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma})]$$

$$= [\mathbf{R}_\gamma\cdot T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma})]'(s_T^2)^{-1}\left\{[\mathbf{0} \quad \mathbf{R}_\gamma]\left(\begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T\cdot\mathbf{I}_g \end{bmatrix}^{-1}\right.\right.$$
$$\left.\left.\times\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}\begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T\cdot\mathbf{I}_g \end{bmatrix}^{-1}\right)^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1}[\mathbf{R}_\gamma\cdot T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma})]$$

$$\xrightarrow{p} [\mathbf{R}_\gamma\sigma_1\boldsymbol{\nu}_2]'(s_T^2)^{-1}\left\{[\mathbf{0} \quad \mathbf{R}_\gamma]\right.$$
$$\left.\times\begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \\ \mathbf{P}_{22}\int \mathbf{W}_2(r)\,dr & \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\mathbf{P}_{22}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1}[\mathbf{R}_\gamma\sigma_1\boldsymbol{\nu}_2]$$

$$= (\sigma_1^2/s_T^2)[\mathbf{R}_\gamma \boldsymbol{\nu}_2]' \left\{ [\mathbf{0} \quad \mathbf{R}_\gamma] \times \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]' \, dr\right\}\mathbf{P}_{22}' \\ \mathbf{P}_{22} \int \mathbf{W}_2(r) \, dr & \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)] \cdot [\mathbf{W}_2(r)]' \, dr\right\}\mathbf{P}_{22}' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix} \right\}^{-1} [\mathbf{R}_\gamma \boldsymbol{\nu}_2]. \tag{19.3.13}$$

Result [19.3.12] implies that conditional on $\mathbf{W}_2(\cdot)$, the vector $\mathbf{R}_\gamma \boldsymbol{\nu}_2$ has a Gaussian distribution with mean $\mathbf{0}$ and variance

$$[\mathbf{0} \quad \mathbf{R}_\gamma] \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]' \, dr\right\}\mathbf{P}_{22}' \\ \mathbf{P}_{22} \int \mathbf{W}_2(r) \, dr & \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)] \cdot [\mathbf{W}_2(r)]' \, dr\right\}\mathbf{P}_{22}' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}.$$

Since s_T^2 provides a consistent estimate of σ_1^2, the limiting distribution of $m \cdot F_T$ conditional on $\mathbf{W}_2(\cdot)$ is thus $\chi^2(m)$, and so the unconditional distribution is $\chi^2(m)$ as well. This means that *OLS* t or F tests involving the cointegrating vector have their standard asymptotic Gaussian or χ^2 distributions.

It is also straightforward to adapt the methods in Section 16.3 to show that the *OLS* χ^2 test of a hypothesis involving just α, or that for a joint hypothesis involving both α and γ, also has a limiting χ^2 distribution.

The analysis to this point applies in the special case when y_{1t} and $\mathbf{y}_{2t}$ follow random walks. The analysis is easily extended to allow for serial correlation in z_t^* or $\mathbf{u}_{2t}$, as long as the critical condition that z_t^* is uncorrelated with $\mathbf{u}_{2\tau}$ for all t and τ is maintained. In particular, suppose that the dynamic process for $(z_t^*, \mathbf{u}_{2t}')'$ is given by

$$\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t} \end{bmatrix} = \mathbf{\Psi}^*(L)\boldsymbol{\varepsilon}_t,$$

with $\{s \cdot \mathbf{\Psi}_s^*\}_{s=0}^\infty$ absolutely summable, $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_\tau') = \mathbf{PP}'$ if $t = \tau$ and $\mathbf{0}$ otherwise, and fourth moments of $\boldsymbol{\varepsilon}_t$ finite. In order for z_t^* to be uncorrelated with $\mathbf{u}_{2\tau}$ for all t and τ, both $\mathbf{\Psi}^*(L)$ and $\mathbf{P}$ must be block-diagonal:

$$\mathbf{\Psi}^*(L) = \begin{bmatrix} \psi_{11}^*(L) & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Psi}_{22}^*(L) \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix},$$

implying that the matrix $\mathbf{\Psi}^*(1) \cdot \mathbf{P}$ is also block-diagonal:

$$\mathbf{\Psi}^*(1) \cdot \mathbf{P} = \begin{bmatrix} \sigma_1 \psi_{11}^*(1) & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Psi}_{22}^*(1) \cdot \mathbf{P}_{22} \end{bmatrix} \equiv \begin{bmatrix} \lambda_1^* & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Lambda}_{22}^* \end{bmatrix}. \tag{19.3.14}$$

Noting the parallel between [19.3.14] and [19.3.8], it is easy to confirm that if $\lambda_1^* \neq 0$ and the rows of $\mathbf{\Lambda}_{22}^*$ are linearly independent, then the analysis of [19.3.10] continues to hold, with σ_1 replaced by λ_1^* and $\mathbf{P}_{22}$ replaced by $\mathbf{\Lambda}_{22}^*$:

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\mathbf{\Lambda}_{22}^{*\prime} \\ \mathbf{\Lambda}_{22}^* \int \mathbf{W}_2(r)\,dr & \mathbf{\Lambda}_{22}^*\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\mathbf{\Lambda}_{22}^{*\prime} \end{bmatrix}^{-1} \times \begin{bmatrix} \lambda_1^* W_1(1) \\ \mathbf{\Lambda}_{22}^*\left\{\int [\mathbf{W}_2(r)]\,dW_1(r)\right\}\lambda_1^* \end{bmatrix}. \qquad [19.3.15]$$

Conditional on $\mathbf{W}_2(\cdot)$, this again describes a Gaussian vector with mean zero and variance

$$(\lambda_1^*)^2 \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\mathbf{\Lambda}_{22}^{*\prime} \\ \mathbf{\Lambda}_{22}^* \int \mathbf{W}_2(r)\,dr & \mathbf{\Lambda}_{22}^*\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\mathbf{\Lambda}_{22}^{*\prime} \end{bmatrix}^{-1}$$

The same calculations as in [19.3.13] further indicate that m times the *OLS F* test of m restrictions involving α or $\boldsymbol{\gamma}$ converges to $(\lambda_1^*)^2/s_T^2$ times a variable that is $\chi^2(m)$ conditional on $\mathbf{W}_2(\cdot)$. Since this distribution does not depend on $\mathbf{W}_2(\cdot)$, the unconditional distribution is also $[(\lambda_1^*)^2/s_T^2]\cdot\chi^2(m)$.

Note that the *OLS* estimate s_T^2 provides a consistent estimate of the variance of z_t^*:

$$s_T^2 \equiv (T - n)^{-1} \sum_{t=1}^{T} (y_{1t} - \hat{\alpha}_T - \hat{\boldsymbol{\gamma}}_T'\mathbf{y}_{2t})^2 \xrightarrow{p} E(z_t^*)^2.$$

However, if z_t^* is serially correlated, this is not the same magnitude as $(\lambda_1^*)^2$. Fortunately, this is simple to correct for. For example, s_T^2 in the usual formula for the F test [19.3.5] could be replaced with

$$(\hat{\lambda}_{1,T}^*)^2 = \hat{c}_{0,T} + 2\cdot\sum_{j=1}^{q} [1 - j/(q+1)]\hat{c}_{j,T} \qquad [19.3.16]$$

for

$$\hat{c}_{j,T} \equiv T^{-1} \sum_{t=j+1}^{T} \hat{u}_t\hat{u}_{t-j} \qquad [19.3.17]$$

with $\hat{u}_t = (y_{1t} - \hat{\alpha}_T - \hat{\boldsymbol{\gamma}}_T'\mathbf{y}_{2t})$ the sample residual resulting from *OLS* estimation of [19.3.1]. If $q \to \infty$ but $q/T \to 0$, then $\hat{\lambda}_{1,T}^* \xrightarrow{p} \lambda_1^*$. It then follows that the test statistic given by

$$(\mathbf{R}_\alpha\hat{\alpha}_T + \mathbf{R}_\gamma\hat{\boldsymbol{\gamma}}_T - \mathbf{r})'\left\{(\hat{\lambda}_{1,T}^*)^2[\mathbf{R}_\alpha \quad \mathbf{R}_\gamma]\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{R}_\alpha' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1} \times (\mathbf{R}_\alpha\hat{\alpha}_T + \mathbf{R}_\gamma\hat{\boldsymbol{\gamma}}_T - \mathbf{r}) \qquad [19.3.18]$$

has an asymptotic $\chi^2(m)$ distribution.

The difficulties with nonstandard distributions for hypothesis tests about the cointegrating vector are thus due to the possibility of nonzero correlations between z_t^* and $\mathbf{u}_{2\tau}$. The basic approach to constructing hypothesis tests will therefore be to transform the regression or the estimates so as to eliminate the effects of this correlation.

Correcting for Correlation by Adding Leads and Lags of $\Delta\mathbf{y}_2$

One correction for the correlation between z_t^* and $\mathbf{u}_{2\tau}$, suggested by Saikkonen (1991), Phillips and Loretan (1991), Stock and Watson (1993), and Wooldridge (1991), is to augment [19.3.1] with leads and lags of $\Delta\mathbf{y}_{2t}$. Specifically, since z_t^* and $\mathbf{u}_{2t}$ are stationary, we can define $\tilde{z}_t$ to be the residual from a linear projection of z_t^* on $\{\mathbf{u}_{2,t-p}, \mathbf{u}_{2,t-p+1}, \ldots, \mathbf{u}_{2,t-1}, \mathbf{u}_{2t}, \mathbf{u}_{2,t+1}, \ldots, \mathbf{u}_{2,t+p}\}$:

$$z_t^* = \sum_{s=-p}^{p} \boldsymbol{\beta}_s' \mathbf{u}_{2,t-s} + \tilde{z}_t,$$

where $\tilde{z}_t$ by construction is uncorrelated with $\mathbf{u}_{2,t-s}$ for $s = -p, -p + 1, \ldots, p$. Recalling from [19.3.2] that $\mathbf{u}_{2t} = \Delta\mathbf{y}_{2t}$, equation [19.3.1] then can be written

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + \sum_{s=-p}^{p} \boldsymbol{\beta}_s' \Delta\mathbf{y}_{2,t-s} + \tilde{z}_t. \qquad [19.3.19]$$

If we are willing to assume that the correlation between z_t^* and $\mathbf{u}_{2,t-s}$ is zero for $|s| > p$, then an F test about the true value of $\boldsymbol{\gamma}$ that has an asymptotic χ^2 distribution is easy to construct using the same approach adopted in [19.3.18].

For a more formal statement, let y_{1t} and $\mathbf{y}_{2t}$ satisfy [19.3.19] and [19.3.2] with

$$\begin{bmatrix} \tilde{z}_t \\ \mathbf{u}_{2t} \end{bmatrix} = \sum_{s=0}^{\infty} \tilde{\boldsymbol{\Psi}}_s \boldsymbol{\varepsilon}_{t-s},$$

where $\{s \cdot \tilde{\boldsymbol{\Psi}}_s\}_{s=0}^{\infty}$ is an absolutely summable sequence of $(n \times n)$ matrices and $\{\boldsymbol{\varepsilon}_t\}_{t=-\infty}^{\infty}$ is an i.i.d. sequence of $(n \times 1)$ vectors with mean zero, variance $\mathbf{PP}'$, and finite fourth moments and with $\tilde{\boldsymbol{\Psi}}(1) \cdot \mathbf{P}$ nonsingular. Suppose that $\tilde{z}_t$ is uncorrelated with $\mathbf{u}_{2\tau}$ for all t and τ, so that

$$\mathbf{P} = \begin{bmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix} \qquad [19.3.20]$$

$$\tilde{\boldsymbol{\Psi}}(L) = \begin{bmatrix} \tilde{\psi}_{11}(L) & \mathbf{0}' \\ \mathbf{0} & \tilde{\boldsymbol{\Psi}}_{22}(L) \end{bmatrix}, \qquad [19.3.21]$$

where $\mathbf{P}_{22}$ and $\tilde{\boldsymbol{\Psi}}_{22}(L)$ are $(g \times g)$ matrices for $g \equiv n - 1$. Define

$$\mathbf{w}_t \equiv (\mathbf{u}_{2,t-p}', \mathbf{u}_{2,t-p+1}', \ldots, \mathbf{u}_{2,t-1}', \mathbf{u}_{2t}', \mathbf{u}_{2,t+1}', \ldots, \mathbf{u}_{2,t+p}')'$$
$$\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_p', \boldsymbol{\beta}_{p-1}', \ldots, \boldsymbol{\beta}_{-p}')',$$

so that the regression model [19.3.19] can be written

$$y_{1t} = \boldsymbol{\beta}'\mathbf{w}_t + \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + \tilde{z}_t. \qquad [19.3.22]$$

The reader is invited to confirm in Exercise 19.2 that the *OLS* estimates of [19.3.22] satisfy

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol\beta}_T - \boldsymbol\beta) \\ T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol\gamma}_T - \boldsymbol\gamma) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{Q}^{-1}\mathbf{h}_1 \\ \tilde{\lambda}_{11}\nu_1 \\ \tilde{\lambda}_{11}\boldsymbol\nu_2 \end{bmatrix}, \qquad [19.3.23]$$

where $\mathbf{Q} \equiv E(\mathbf{w}_t\mathbf{w}_t')$, $T^{-1/2}\Sigma\mathbf{w}_t\tilde{z}_t \xrightarrow{L} \mathbf{h}_1$, $\tilde{\lambda}_{11} \equiv \sigma_1 \cdot \tilde{\psi}_{11}(1)$, and

$$\begin{bmatrix} \nu_1 \\ \boldsymbol\nu_2 \end{bmatrix} = \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol\Lambda}_{22}' \\ \tilde{\boldsymbol\Lambda}_{22}\int \mathbf{W}_2(r)\,dr & \tilde{\boldsymbol\Lambda}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol\Lambda}_{22}' \end{bmatrix}^{-1} \times \begin{bmatrix} W_1(1) \\ \tilde{\boldsymbol\Lambda}_{22}\left\{\int [\mathbf{W}_2(r)]\,dW_1(r)\right\} \end{bmatrix}.$$

Here $\tilde{\boldsymbol\Lambda}_{22} \equiv \tilde{\boldsymbol\Psi}_{22}(1)\cdot\mathbf{P}_{22}$, $W_1(r)$ is univariate standard Brownian motion, $\mathbf{W}_2(r)$ is g-dimensional standard Brownian motion that is independent of $W_1(\cdot)$, and the integral sign denotes integration over r from 0 to 1. Hence, as in [19.3.12],

$$\left[\begin{array}{c|c} \nu_1 & \\ \boldsymbol\nu_2 & \mathbf{W}_2(\cdot) \end{array}\right] \sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol\Lambda}_{22}' \\ \tilde{\boldsymbol\Lambda}_{22}\int \mathbf{W}_2(r)\,dr & \tilde{\boldsymbol\Lambda}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol\Lambda}_{22}' \end{bmatrix}^{-1}\right). \qquad [19.3.24]$$

Moreover, the Wald form of the OLS χ^2 test of the null hypothesis $\mathbf{R}_\gamma\boldsymbol\gamma = \mathbf{r}$, where $\mathbf{R}_\gamma$ is an $(m \times g)$ matrix and $\mathbf{r}$ is an $(m \times 1)$ vector, can be shown to satisfy

$$\begin{aligned} \chi_T^2 &= \{\mathbf{R}_\gamma\hat{\boldsymbol\gamma}_T - \mathbf{r}\}'\left\{s_T^2[\mathbf{0} \quad 0 \quad \mathbf{R}_\gamma]\begin{bmatrix} \Sigma\mathbf{w}_t\mathbf{w}_t' & \Sigma\mathbf{w}_t & \Sigma\mathbf{w}_t\mathbf{y}_{2t}' \\ \Sigma\mathbf{w}_t' & T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t}\mathbf{w}_t' & \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0} \\ \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1} \\ &\quad \times \{\mathbf{R}_\gamma\hat{\boldsymbol\gamma}_T - \mathbf{r}\} \\ &\xrightarrow{p} (\tilde{\lambda}_{11}^2/s_T^2)[\mathbf{R}_\gamma\boldsymbol\nu_2]'\left\{[\mathbf{0} \quad \mathbf{R}_\gamma] \right. \\ &\quad \left. \times \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol\Lambda}_{22}' \\ \tilde{\boldsymbol\Lambda}_{22}\int \mathbf{W}_2(r)\,dr & \tilde{\boldsymbol\Lambda}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol\Lambda}_{22}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1}[\mathbf{R}_\gamma\boldsymbol\nu_2]; \end{aligned} \qquad [19.3.25]$$

see Exercise 19.3. But result [19.3.24] implies that conditional on $\mathbf{W}_2(\cdot)$, the expression in [19.3.25] is $(\tilde{\lambda}_{11}^2/s_T^2)$ times a $\chi^2(m)$ variable. Since this distribution is the same for all $\mathbf{W}_2(\cdot)$, it follows that the unconditional distribution also satisfies

$$\chi_T^2 \xrightarrow{p} (\tilde{\lambda}_{11}^2/s_T^2)\cdot\chi^2(m). \qquad [19.3.26]$$

Result [19.3.26] establishes that in order to test a hypothesis about the value of the cointegrating vector $\boldsymbol{\gamma}$, we can estimate [19.3.19] by *OLS* and calculate a standard F test of the hypothesis that $\mathbf{R}_\gamma\boldsymbol{\gamma} = \mathbf{r}$ using the usual formula. We need only to multiply the *OLS F* statistic by a consistent estimate of $(s_T^2/\tilde{\lambda}_{11}^2)$, and the F statistic can be compared with the usual $F(m, T - k)$ tables for k the number of parameters estimated in [19.3.19] for an asymptotically valid test. Similarly, the *OLS t* statistic could be multiplied by $(s_T^2/\tilde{\lambda}_{11}^2)^{1/2}$ and compared with the standard t tables.

A consistent estimate of $\tilde{\lambda}_{11}^2$ is easy to obtain. Recall that $\tilde{\lambda}_{11} = \sigma_1\cdot\tilde{\psi}_{11}(1)$, where $\tilde{z}_t = \tilde{\psi}_{11}(L)\varepsilon_{1t}$ and $E(\varepsilon_{1t}^2) = \sigma_1^2$. Suppose we approximate $\tilde{\psi}_{11}(L)$ by an $AR(p)$ process, and let $\hat{u}_t$ denote the sample residual resulting from *OLS* estimation of [19.3.19]. If $\hat{u}_t$ is regressed on p of its own lags:

$$\hat{u}_t = \phi_1\hat{u}_{t-1} + \phi_2\hat{u}_{t-2} + \cdots + \phi_p\hat{u}_{t-p} + e_t,$$

then a natural estimate of $\tilde{\lambda}_{11}$ is

$$\hat{\tilde{\lambda}}_{11} = \hat{\sigma}_1/(1 - \hat{\phi}_1 - \hat{\phi}_2 - \cdots - \hat{\phi}_p), \qquad [19.3.27]$$

where

$$\hat{\sigma}_1^2 = (T - p)^{-1}\sum_{t=p+1}^{T}\hat{e}_t^2$$

and where T indicates the number of observations actually used to estimate [19.3.19]. Alternatively, if the dynamics implied by $\tilde{\psi}_{11}(L)$ were to be approximated on the basis of q autocovariances, the Newey-West estimator could be used:

$$\hat{\tilde{\lambda}}_{11}^2 = \hat{c}_0 + 2\cdot\sum_{j=1}^{q}[1 - j/(q + 1)]\hat{c}_j, \qquad [19.3.28]$$

where

$$\hat{c}_j = T^{-1}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}.$$

These results were derived under the assumption that there were no drift terms in any of the elements of $\mathbf{y}_{2t}$. However, it is not hard to show that the same procedure works in exactly the same way when some or all of the elements of $\mathbf{y}_{2t}$ involve deterministic time trends. In addition, there is no problem with adding a time trend to the regression of [19.3.19] and testing a hypothesis about its value using this same factor applied to the usual F test. This allows testing separately the hypotheses that (1) $y_{1t} - \boldsymbol{\gamma}'\mathbf{y}_{2t}$ has no time trend and (2) $y_{1t} - \boldsymbol{\gamma}'\mathbf{y}_{2t}$ is $I(0)$, that is, testing separately the restrictions [19.1.15] and [19.1.12]. The reader is invited to verify these claims in Exercises 19.4 and 19.5.

Illustration—Testing Hypotheses About the Cointegrating Relation Between Consumption and Income

As an illustration of this approach, consider again the relation between consumption c_t and income y_t, for which evidence of cointegration was found earlier.

The following regression was estimated for t = 1948:II to 1988:III by *OLS*, with the usual *OLS* formulas for standard deviations given in parentheses:

$$\begin{aligned} c_t = & \underset{(2.34)}{-4.52} + \underset{(0.00306)}{0.99216}\, y_t + \underset{(0.12)}{0.15}\, \Delta y_{t+4} + \underset{(0.12)}{0.29}\, \Delta y_{t+3} + \underset{(0.11)}{0.26}\, \Delta y_{t+2} \\ & + \underset{(0.12)}{0.49}\, \Delta y_{t+1} - \underset{(0.12)}{0.24}\, \Delta y_t - \underset{(0.11)}{0.01}\, \Delta y_{t-1} + \underset{(0.11)}{0.07}\, \Delta y_{t-2} \\ & + \underset{(0.11)}{0.04}\, \Delta y_{t-3} + \underset{(0.11)}{0.02}\, \Delta y_{t-4} + \hat{u}_t \\ s^2 = & (T - 11)^{-1} \sum_{t=1}^{T} \hat{u}_t^2 = (1.516)^2. \end{aligned} \tag{19.3.29}$$

Here T, the number of observations actually used to estimate [19.3.29], is 162. To test the null hypothesis that the cointegrating vector is $\mathbf{a} = (1, -1)'$, we start with the usual *OLS* t test of this hypothesis,

$$t = (0.99216 - 1)/0.00306 = -2.562.$$

A second-order autoregression fitted to the residuals of [19.3.29] by *OLS* produced

$$\hat{u}_t = 0.7180\, \hat{u}_{t-1} + 0.2057\, \hat{u}_{t-2} + \hat{e}_t, \tag{19.3.30}$$

where

$$\hat{\sigma}_1^2 = (T - 2)^{-1} \sum_{t=3}^{T} \hat{e}_t^2 = 0.38092.$$

Thus, the estimate of λ_{11} suggested in [19.3.27] is

$$\hat{\lambda}_{11} = (0.38092)^{1/2}/(1 - 0.7180 - 0.2057) = 8.089.$$

Hence, a test of the null hypothesis that $\mathbf{a} = (1, -1)'$ can be based on

$$t \cdot (s/\hat{\lambda}_{11}) = (-2.562)(1.516)/(8.089) = -0.48.$$

Since -0.48 is above the 5% critical value of -1.96 for a $N(0, 1)$ variable, we accept the null hypothesis that $\mathbf{a} = (1, -1)'$.

To test the restrictions implied by cointegration for the time trend and stochastic component separately, the regression of [19.3.29] was reestimated with a time trend included:

$$\begin{aligned} c_t = & \underset{(15.0)}{198.9} + \underset{(0.0229)}{0.6812}\, y_t + \underset{(0.0197)}{0.2690}\, t + \underset{(0.08)}{0.03}\, \Delta y_{t+4} + \underset{(0.08)}{0.17}\, \Delta y_{t+3} \\ & + \underset{(0.08)}{0.15}\, \Delta y_{t+2} + \underset{(0.08)}{0.40}\, \Delta y_{t+1} - \underset{(0.08)}{0.05}\, \Delta y_t + \underset{(0.08)}{0.13}\, \Delta y_{t-1} \\ & + \underset{(0.08)}{0.23}\, \Delta y_{t-2} + \underset{(0.08)}{0.20}\, \Delta y_{t-3} + \underset{(0.07)}{0.19}\, \Delta y_{t-4} + \hat{u}_t \\ s^2 = & (T - 12)^{-1} \sum_{t=1}^{T} \hat{u}_t^2 = (1.017)^2. \end{aligned} \tag{19.3.31}$$

A second-order autoregression fitted to the residuals of [19.3.31] produced

$$\hat{u}_t = 0.6872\, \hat{u}_{t-1} + 0.1292\, \hat{u}_{t-2} + \hat{e}_t,$$

where

$$\hat{\sigma}_1^2 = (T - 2)^{-1} \sum_{t=3}^{T} \hat{e}_t^2 = 0.34395$$

and

$$\hat{\lambda}_{11} = (0.34395)^{1/2}/(1 - 0.6872 - 0.1292) = 3.194.$$

A test of the hypothesis that the time trend does not contribute to [19.3.31] is thus given by

$$[(0.2690)/(0.0197)]\cdot[(1.017)/(3.194)] = 4.35.$$

Since $4.35 > 1.96$, we reject the null hypothesis that the coefficient on the time trend is zero.

The *OLS* results in [19.3.29] are certainly consistent with the hypothesis that consumption and income are cointegrated with cointegrating vector $\mathbf{a} = (1, -1)'$. However, [19.3.31] indicates that this result is dominated by the deterministic time trend common to c_t and y_t. It appears that while $\mathbf{a} = (1, -1)'$ is sufficient to eliminate the trend components of c_t and y_t, the residual $c_t - y_t$ contains a stochastic component that could be viewed as $I(1)$. Figure 19.6 provides a plot of $c_t - y_t$. It is indeed the case that this transformation seems to have eliminated the trend, though stochastic shocks to $c_t - y_t$ do not appear to die out within a period as short as 2 years.

Further Remarks and Extensions

It was assumed throughout the derivations in this section that $\tilde{z}_t$ is $I(0)$, so that $\mathbf{y}_t$ is cointegrated with the cointegrating vector having a nonzero coefficient on y_{1t}. If $\mathbf{y}_t$ were not cointegrated, then [19.3.19] would be a spurious regression and the tests that were described would not be valid. For this reason estimation of [19.3.19] would usually be undertaken after an initial investigation suggested the presence of a cointegrating relation.

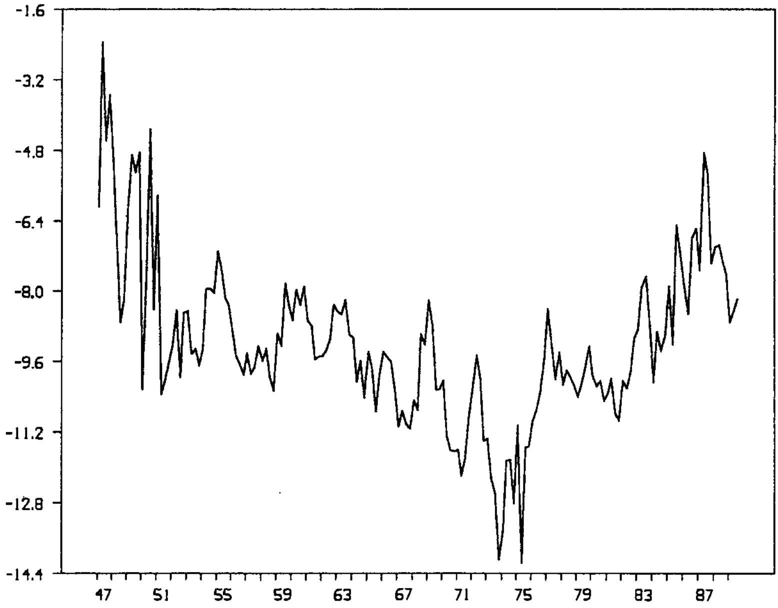


FIGURE 19.6 One hundred times the difference between the log of personal consumption expenditures (c_t) and the log of personal disposable income (y_t) for the United States, quarterly, 1947–89.

It was also assumed that Λ_{22} is nonsingular, meaning that there are no cointegrating relations among the variables in $\mathbf{y}_{2t}$. Suppose instead that we are interested in estimating $h > 1$ different cointegrating vectors, as represented by a system of the form

$$\underset{(h\times 1)}{\mathbf{y}_{1t}} = \underset{(h\times g)}{\mathbf{\Gamma}'} \cdot \underset{(g\times 1)}{\mathbf{y}_{2t}} + \underset{(h\times 1)}{\boldsymbol{\mu}_1^*} + \underset{(h\times 1)}{\mathbf{z}_t^*} \qquad [19.3.32]$$

$$\underset{(g\times 1)}{\Delta\mathbf{y}_{2t}} = \underset{(g\times 1)}{\boldsymbol{\delta}_2} + \underset{(g\times 1)}{\mathbf{u}_{2t}} \qquad [19.3.33]$$

with

$$\begin{bmatrix} \mathbf{z}_t^* \\ \mathbf{u}_{2t} \end{bmatrix} = \mathbf{\Psi}^*(L)\boldsymbol{\varepsilon}_t$$

and $\mathbf{\Psi}^*(1)$ nonsingular. Here the generalization of the previous approach would be to augment [19.3.32] with leads and lags of $\Delta\mathbf{y}_{2t}$:

$$\mathbf{y}_{1t} = \boldsymbol{\mu}_1^* + \mathbf{\Gamma}'\mathbf{y}_{2t} + \sum_{s=-p}^{p} \mathbf{B}_s'\Delta\mathbf{y}_{2,t-s} + \tilde{\mathbf{z}}_t, \qquad [19.3.34]$$

where $\mathbf{B}_s'$ denotes an $(h \times g)$ matrix of coefficients and it is assumed that $\tilde{\mathbf{z}}_t$ is uncorrelated with $\mathbf{u}_{2\tau}$ for all t and τ. Expression [19.3.34] describes a set of h equations. The ith equation regresses y_{it} on a constant, on the current value of all the elements of $\mathbf{y}_{2t}$, and on past, present, and future changes of all the elements of $\mathbf{y}_{2t}$. This equation could be estimated by *OLS*, with the usual F statistics multiplied by $[s_T^{(i)}/\hat{\lambda}_{11}^{(i)}]^2$, where $s_T^{(i)}$ is the standard error of the regression and $\hat{\lambda}_{11}^{(i)}$ could be estimated from the autocovariances of the residuals $\hat{\tilde{z}}_{it}$ for the regression.

The approach just described estimated the relation in [19.3.19] by *OLS* and made adjustments to the usual t or F statistics so that they could be compared with the standard t and F tables. Stock and Watson (1993) also suggested the more efficient approach of first estimating [19.3.19] by *OLS*, then using the residuals to construct a consistent estimate of the autocorrelation of u_t as in [19.3.27] or [19.3.28], and finally reestimating the equation by generalized least squares. The resulting *GLS* standard errors could be used to construct asymptotically χ^2 hypothesis tests.

Phillips and Loretan (1991, p. 424) suggested that instead autocorrelation of the residuals of [19.3.19] could be handled by including lagged values of the residual of the cointegrating relation in the form of

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + \sum_{s=-p}^{p} \boldsymbol{\beta}_s'\Delta\mathbf{y}_{2,t-s} + \sum_{s=1}^{p} \phi_s(y_{1,t-s} - \boldsymbol{\gamma}'\mathbf{y}_{2,t-s}) + \varepsilon_{1t}. \qquad [19.3.35]$$

Their proposal was to estimate the parameters in [19.3.35] by numerical minimization of the sum of squared residuals.

Phillips and Hansen's Fully Modified OLS *Estimates*

A related approach was suggested by Phillips and Hansen (1990). Consider again a system with a single cointegrating relation written in the form

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + z_t^* \qquad [19.3.36]$$

$$\Delta\mathbf{y}_{2t} = \mathbf{u}_{2t} \qquad [19.3.37]$$

$$\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t} \end{bmatrix} = \mathbf{\Psi}^*(L)\boldsymbol{\varepsilon}_t$$

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{PP}',$$

where $\mathbf{y}_{2t}$ is a $(g \times 1)$ vector and $\boldsymbol{\varepsilon}_t$ is an $(n \times 1)$ i.i.d. zero-mean vector for $n \equiv (g + 1)$. Define

$$\mathbf{\Lambda}^* \equiv \mathbf{\Psi}^*(1)\cdot\mathbf{P}$$

$$\underset{(n\times n)}{\mathbf{\Sigma}^*} \equiv \mathbf{\Lambda}^*\cdot[\mathbf{\Lambda}^*]' \equiv \begin{bmatrix} \underset{(1\times 1)}{\Sigma_{11}^*} & \underset{(1\times g)}{\mathbf{\Sigma}_{21}^{*\prime}} \\ \underset{(g\times 1)}{\mathbf{\Sigma}_{21}^*} & \underset{(g\times g)}{\mathbf{\Sigma}_{22}^*} \end{bmatrix}, \qquad [19.3.38]$$

with $\mathbf{\Lambda}^*$ as always assumed to be a nonsingular matrix.

Recall from equation [10.3.4] that the autocovariance-generating function for $(z_t^*, \mathbf{u}_{2t}')'$ is given by

$$\begin{aligned}\mathbf{G}(z) &\equiv \sum_{v=-\infty}^{\infty} z^v \begin{bmatrix} E(z_t^* z_{t-v}^*) & E(z_t^*\mathbf{u}_{2,t-v}') \\ E(\mathbf{u}_{2t}z_{t-v}^*) & E(\mathbf{u}_{2t}\mathbf{u}_{2,t-v}') \end{bmatrix} \\ &= [\mathbf{\Psi}^*(z)]\cdot\mathbf{PP}'[\mathbf{\Psi}^*(z^{-1})]'.\end{aligned}$$

Thus, $\mathbf{\Sigma}^*$ could alternatively be described as the autocovariance-generating function $\mathbf{G}(z)$ evaluated at $z = 1$:

$$\begin{bmatrix} \Sigma_{11}^* & \mathbf{\Sigma}_{21}^{*\prime} \\ \mathbf{\Sigma}_{21}^* & \mathbf{\Sigma}_{22}^* \end{bmatrix} = \sum_{v=-\infty}^{\infty} \begin{bmatrix} E(z_t^* z_{t-v}^*) & E(z_t^*\mathbf{u}_{2,t-v}') \\ E(\mathbf{u}_{2t}z_{t-v}^*) & E(\mathbf{u}_{2t}\mathbf{u}_{2,t-v}') \end{bmatrix}. \qquad [19.3.39]$$

The difference between the general distribution for the estimated cointegrating vector described in Proposition 19.2 and the convenient special case investigated in [19.3.15] is due to two factors. The first is the possibility of a nonzero value for $\mathbf{\Sigma}_{21}^*$, and the second is the constant term that might appear in the variable $\mathbf{h}_2$ described in Proposition 19.2 arising from a nonzero value for

$$\aleph \equiv \sum_{v=0}^{\infty} E(\mathbf{u}_{2t}z_{t+v}^*). \qquad [19.3.40]$$

The first issue can be addressed by subtracting $\mathbf{\Sigma}_{21}^{*\prime}(\mathbf{\Sigma}_{22}^*)^{-1}\Delta\mathbf{y}_{2t}$ from both sides of [19.3.36], arriving at

$$y_{1t}^\dagger = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + z_t^\dagger,$$

where

$$\begin{aligned} y_{1t}^\dagger &\equiv y_{1t} - \mathbf{\Sigma}_{21}^{*\prime}(\mathbf{\Sigma}_{22}^*)^{-1}\Delta\mathbf{y}_{2t} \\ z_t^\dagger &\equiv z_t^* - \mathbf{\Sigma}_{21}^{*\prime}(\mathbf{\Sigma}_{22}^*)^{-1}\Delta\mathbf{y}_{2t}. \end{aligned} \qquad [19.3.41]$$

Notice that since $\Delta\mathbf{y}_{2t} = \mathbf{u}_{2t}$, the vector $(z_t^\dagger, \mathbf{u}_{2t}')'$ can be written as

$$\begin{bmatrix} z_t^\dagger \\ \mathbf{u}_{2t} \end{bmatrix} = \mathbf{L}'\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t} \end{bmatrix} \qquad [19.3.42]$$

for

$$\mathbf{L}' \equiv \begin{bmatrix} 1 & -\mathbf{\Sigma}_{21}^{*\prime}(\mathbf{\Sigma}_{22}^*)^{-1} \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix} \equiv \begin{bmatrix} \underset{(1\times n)}{\boldsymbol{\ell}_1'} \\ \underset{(g\times n)}{\mathbf{L}_2'} \end{bmatrix}. \qquad [19.3.43]$$

Suppose we were to estimate α and $\boldsymbol{\gamma}$ with an *OLS* regression of $y_{1t}^\dagger$ on a constant and $\mathbf{y}_{2t}$:

$$\begin{bmatrix} \hat{\alpha}_T^\dagger \\ \hat{\boldsymbol{\gamma}}_T^\dagger \end{bmatrix} = \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_{1t}^\dagger \\ \Sigma\mathbf{y}_{2t}y_{1t}^\dagger \end{bmatrix}. \qquad [19.3.44]$$

The distribution of the resulting estimates is readily found from Proposition 19.2. Note that the vector $\boldsymbol{\lambda}_1^{*\prime}$ used in Proposition 19.2 can be written as $\mathbf{e}_1'\boldsymbol{\Lambda}^*$ for $\mathbf{e}_1'$ the first row of $\mathbf{I}_n$, while the matrix $\boldsymbol{\Lambda}_2^*$ in Proposition 19.2 can be written as $\mathbf{L}_2'\boldsymbol{\Lambda}^*$ for $\mathbf{L}_2'$ the last g rows of $\mathbf{L}'$. The asymptotic distribution of the estimates in [19.3.44] is found by writing $\boldsymbol{\Lambda}_2^*$ in [19.2.13] as $\mathbf{L}_2'\boldsymbol{\Lambda}^*$, replacing $\boldsymbol{\lambda}_1^{*\prime} = \mathbf{e}_1'\boldsymbol{\Lambda}^*$ in [19.2.13] with $\boldsymbol{\ell}_1'\boldsymbol{\Lambda}^*$, and replacing $E(\mathbf{u}_{2t}z_{t+v}^*)$ with $E(\mathbf{u}_{2t}z_{t+v}^\dagger)$:

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T^\dagger - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T^\dagger - \boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma z_t^\dagger \\ T^{-1}\Sigma\mathbf{y}_{2t}z_t^\dagger \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} 1 & \left\{\int [\mathbf{W}(r)]'\,dr\right\}\boldsymbol{\Lambda}^{*\prime}\mathbf{L}_2 \\ \mathbf{L}_2'\boldsymbol{\Lambda}^*\int \mathbf{W}(r)\,dr & \mathbf{L}_2'\boldsymbol{\Lambda}^*\left\{\int [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\boldsymbol{\Lambda}^{*\prime}\mathbf{L}_2 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \boldsymbol{\ell}_1'\boldsymbol{\Lambda}^*\mathbf{W}(1) \\ \mathbf{L}_2'\boldsymbol{\Lambda}^*\left\{\int [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\boldsymbol{\Lambda}^{*\prime}\boldsymbol{\ell}_1 + \boldsymbol{\aleph}^\dagger \end{bmatrix}, \qquad [19.3.45]$$

where $\mathbf{W}(r)$ denotes n-dimensional standard Brownian motion and

$$\begin{aligned} \boldsymbol{\aleph}^\dagger &\equiv \sum_{v=0}^{\infty} E(\mathbf{u}_{2t}z_{t+v}^\dagger) \\ &= \sum_{v=0}^{\infty} E\{\mathbf{u}_{2t}[z_{t+v}^* - \boldsymbol{\Sigma}_{21}^{*\prime}(\boldsymbol{\Sigma}_{22}^*)^{-1}\mathbf{u}_{2,t+v}]\} \qquad [19.3.46] \\ &= \sum_{v=0}^{\infty} E\{\mathbf{u}_{2t}[z_{t+v}^* \quad \mathbf{u}_{2,t+v}']\}\begin{bmatrix} 1 \\ -(\boldsymbol{\Sigma}_{22}^*)^{-1}\boldsymbol{\Sigma}_{21}^* \end{bmatrix}. \end{aligned}$$

Now, consider the $(n \times 1)$ vector process defined by

$$\mathbf{B}(r) \equiv \begin{bmatrix} \boldsymbol{\ell}_1' \\ \mathbf{L}_2' \end{bmatrix}\boldsymbol{\Lambda}^*\cdot\mathbf{W}(r). \qquad [19.3.47]$$

From [19.3.43] and [19.3.38], this is Brownian motion with variance matrix

$$\begin{aligned} E\{[\mathbf{B}(1)]\cdot[\mathbf{B}(1)]'\} &= \begin{bmatrix} \boldsymbol{\ell}_1' \\ \mathbf{L}_2' \end{bmatrix}\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*\prime}[\boldsymbol{\ell}_1 \quad \mathbf{L}_2] \\ &= \begin{bmatrix} 1 & -\boldsymbol{\Sigma}_{21}^{*\prime}(\boldsymbol{\Sigma}_{22}^*)^{-1} \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix}\begin{bmatrix} \Sigma_{11}^* & \boldsymbol{\Sigma}_{21}^{*\prime} \\ \boldsymbol{\Sigma}_{21}^* & \boldsymbol{\Sigma}_{22}^* \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0}' \\ -(\boldsymbol{\Sigma}_{22}^*)^{-1}\boldsymbol{\Sigma}_{21}^* & \mathbf{I}_g \end{bmatrix} \\ &= \begin{bmatrix} (\sigma_1^\dagger)^2 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^* \end{bmatrix}, \end{aligned} \qquad [19.3.48]$$

where

$$(\sigma_1^\dagger)^2 \equiv \Sigma_{11}^* - \boldsymbol{\Sigma}_{21}^{*\prime}(\boldsymbol{\Sigma}_{22}^*)^{-1}\boldsymbol{\Sigma}_{21}^*. \qquad [19.3.49]$$

Partition $\mathbf{B}(r)$ as

$$\underset{(n\times 1)}{\mathbf{B}(r)} = \begin{bmatrix} \underset{(1\times 1)}{B_1(r)} \\ \underset{(g\times 1)}{\mathbf{B}_2(r)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\ell}_1'\boldsymbol{\Lambda}^*\mathbf{W}(r) \\ \mathbf{L}_2'\boldsymbol{\Lambda}^*\mathbf{W}(r) \end{bmatrix}.$$

Then [19.3.48] implies that $B_1(r)$ is scalar Brownian motion with variance $(\sigma_1^\dagger)^2$ while $\mathbf{B}_2(r)$ is g-dimensional Brownian motion with variance matrix $\mathbf{\Sigma}_{22}^*$, with $B_1(\cdot)$ independent of $\mathbf{B}_2(\cdot)$. The process $\mathbf{B}(r)$ in turn can be viewed as generated by a different standard Brownian motion $\mathbf{W}^\dagger(r)$, where

$$\begin{bmatrix} B_1(r) \\ \mathbf{B}_2(r) \end{bmatrix} = \begin{bmatrix} \sigma_1^\dagger & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{22}^* \end{bmatrix} \begin{bmatrix} W_1^\dagger(r) \\ \mathbf{W}_2^\dagger(r) \end{bmatrix}$$

for $\mathbf{P}_{22}^*\mathbf{P}_{22}^{*\prime} = \mathbf{\Sigma}_{22}^*$ the Cholesky factorization of $\mathbf{\Sigma}_{22}^*$. The result [19.3.45] can then equivalently be expressed as

$$\begin{aligned} &\begin{bmatrix} T^{1/2}(\hat{\alpha}_T^\dagger - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T^\dagger - \boldsymbol{\gamma}) \end{bmatrix} \\ &\xrightarrow{L} \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2^\dagger(r)]' \, dr\right\}\mathbf{P}_{22}^{*\prime} \\ \mathbf{P}_{22}^* \int \mathbf{W}_2^\dagger(r)\, dr & \mathbf{P}_{22}^*\left\{\int [\mathbf{W}_2^\dagger(r)]\cdot[\mathbf{W}_2^\dagger(r)]' \, dr\right\}\mathbf{P}_{22}^{*\prime} \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} \sigma_1^\dagger \cdot W_1^\dagger(1) \\ \mathbf{P}_{22}^*\left\{\int \mathbf{W}_2^\dagger(r)\, dW_1^\dagger(r)\right\}\sigma_1^\dagger + \boldsymbol{\aleph}^\dagger \end{bmatrix}. \end{aligned} \qquad [19.3.50]$$

If it were not for the presence of the constant $\boldsymbol{\aleph}^\dagger$, the distribution in [19.3.50] would be of the form of [19.3.11], from which it would follow that conditional on $\mathbf{W}_2^\dagger(\cdot)$, the variable in [19.3.50] would be Gaussian and test statistics that are asymptotically χ^2 could be generated as before.

Recalling [19.3.39], one might propose to estimate $\mathbf{\Sigma}^*$ by

$$\begin{bmatrix} \hat{\Sigma}_{11}^* & \hat{\mathbf{\Sigma}}_{21}^{*\prime} \\ \hat{\mathbf{\Sigma}}_{21}^* & \hat{\mathbf{\Sigma}}_{22}^* \end{bmatrix} = \hat{\mathbf{\Gamma}}_0 + \sum_{v=1}^{q} \{1 - [v/(q+1)]\}(\hat{\mathbf{\Gamma}}_v + \hat{\mathbf{\Gamma}}_v'), \qquad [19.3.51]$$

where

$$\begin{aligned} \hat{\mathbf{\Gamma}}_v &= T^{-1} \sum_{t=v+1}^{T} \begin{bmatrix} (\hat{z}_t^* \hat{z}_{t-v}^*) & (\hat{z}_t^* \hat{\mathbf{u}}_{2,t-v}') \\ (\hat{\mathbf{u}}_{2t} \hat{z}_{t-v}^*) & (\hat{\mathbf{u}}_{2t}\hat{\mathbf{u}}_{2,t-v}') \end{bmatrix} \\ &\equiv \begin{bmatrix} \hat{\Gamma}_{11}^{(v)} & \hat{\mathbf{\Gamma}}_{12}^{(v)} \\ \hat{\mathbf{\Gamma}}_{21}^{(v)} & \hat{\mathbf{\Gamma}}_{22}^{(v)} \end{bmatrix} \end{aligned} \qquad [19.3.52]$$

for $\hat{z}_t^*$ the sample residual resulting from estimation of [19.3.36] by *OLS* and $\hat{\mathbf{u}}_{2t} = \Delta\mathbf{y}_{2t}$. To arrive at a similar estimate of $\boldsymbol{\aleph}^\dagger$, note that [19.3.46] can be written

$$\begin{aligned} \boldsymbol{\aleph}^\dagger &= \sum_{v=0}^{\infty} E\{\mathbf{u}_{2,t-v}[z_t^* \quad \mathbf{u}_{2t}']\} \begin{bmatrix} 1 \\ -(\mathbf{\Sigma}_{22}^*)^{-1}\mathbf{\Sigma}_{21}^* \end{bmatrix} \\ &= \sum_{v=0}^{\infty} E\left\{\begin{bmatrix} z_t^* \mathbf{u}_{2,t-v}' \\ \mathbf{u}_{2t}\mathbf{u}_{2,t-v}' \end{bmatrix}'\right\} \begin{bmatrix} 1 \\ -(\mathbf{\Sigma}_{22}^*)^{-1}\mathbf{\Sigma}_{21}^* \end{bmatrix} \\ &= \sum_{v=0}^{\infty} \begin{bmatrix} \mathbf{\Gamma}_{12}^{(v)} \\ \mathbf{\Gamma}_{22}^{(v)} \end{bmatrix}' \begin{bmatrix} 1 \\ -(\mathbf{\Sigma}_{22}^*)^{-1}\mathbf{\Sigma}_{21}^* \end{bmatrix}. \end{aligned}$$

This suggests the estimator

$$\hat{\boldsymbol{\aleph}}_T^\dagger = \sum_{v=0}^{q} \left\{1 - [v/(q+1)]\right\}\left\{\begin{bmatrix} [\hat{\mathbf{\Gamma}}_{12}^{(v)}]' & [\hat{\mathbf{\Gamma}}_{22}^{(v)}]' \end{bmatrix}\right\} \begin{bmatrix} 1 \\ -(\hat{\mathbf{\Sigma}}_{22}^*)^{-1}\hat{\mathbf{\Sigma}}_{21}^* \end{bmatrix}. \qquad [19.3.53]$$

The fully modified *OLS* estimator proposed by Phillips and Hansen (1990) is then

$$\begin{bmatrix} \hat{\alpha}_T^{\dagger\dagger} \\ \hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} \end{bmatrix} = \begin{bmatrix} T & \Sigma \mathbf{y}_{2t}' \\ \Sigma \mathbf{y}_{2t} & \Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \Sigma \hat{y}_{1t}^{\dagger} \\ \{\Sigma \mathbf{y}_{2t}\hat{y}_{1t}^{\dagger} - T\hat{\aleph}_T^{\dagger}\} \end{bmatrix}$$

for $\hat{y}_{1t}^{\dagger} \equiv y_{1t} - \hat{\Sigma}_{21}^{*\prime}(\hat{\Sigma}_{22}^{*})^{-1}\Delta \mathbf{y}_{2t}$. This analysis implies that

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T^{\dagger\dagger} - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma \mathbf{y}_{2t}' \\ T^{-3/2}\Sigma \mathbf{y}_{2t} & T^{-2}\Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma \hat{z}_t^{\dagger} \\ T^{-1}\Sigma \mathbf{y}_{2t}\hat{z}_t^{\dagger} - \hat{\aleph}_T \end{bmatrix}$$

$$\xrightarrow{L} \sigma_1^{\dagger}\begin{bmatrix} \nu_1 \\ \boldsymbol{\nu}_2 \end{bmatrix},$$

where

$$\begin{bmatrix} \nu_1 \\ \boldsymbol{\nu}_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2^{\dagger}(r)]'\,dr\right\}\mathbf{P}_{22}^{*\prime} \\ \mathbf{P}_{22}^{*}\int \mathbf{W}_2^{\dagger}(r)\,dr & \mathbf{P}_{22}^{*}\left\{\int [\mathbf{W}_2^{\dagger}(r)]\cdot[\mathbf{W}_2^{\dagger}(r)]'\,dr\right\}\mathbf{P}_{22}^{*\prime} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} W_1^{\dagger}(1) \\ \mathbf{P}_{22}^{*}\left\{\int \mathbf{W}_2^{\dagger}(r)\,dW_1^{\dagger}(r)\right\} \end{bmatrix}.$$

It follows as in [19.3.12] that

$$\left[\begin{matrix} \nu_1 \\ \boldsymbol{\nu}_2 \end{matrix}\middle| \mathbf{W}_2^{\dagger}(\cdot)\right] \sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \mathbf{H}^{-1}\right)$$

for

$$\mathbf{H} \equiv \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2^{\dagger}(r)]'\,dr\right\}\mathbf{P}_{22}^{*\prime} \\ \mathbf{P}_{22}^{*}\int \mathbf{W}_2^{\dagger}(r)\,dr & \mathbf{P}_{22}^{*}\left\{\int [\mathbf{W}_2^{\dagger}(r)]\cdot[\mathbf{W}_2^{\dagger}(r)]'\,dr\right\}\mathbf{P}_{22}^{*\prime} \end{bmatrix}.$$

Furthermore, [19.3.49] suggests that a consistent estimate of $(\sigma_1^{\dagger})^2$ is provided by

$$(\hat{\sigma}_1^{\dagger})^2 = \hat{\Sigma}_{11}^{*} - \hat{\Sigma}_{21}^{*\prime}(\hat{\Sigma}_{22}^{*})^{-1}\hat{\Sigma}_{21}^{*},$$

with $\hat{\Sigma}_{ij}^{*}$ given by [19.3.51]. Thus, if we multiply the usual Wald form of the χ^2 test of m restrictions of the form $\mathbf{R}\boldsymbol{\gamma} = \mathbf{r}$ by $(s_T/\hat{\sigma}_1^{\dagger})^2$, the result is an asymptotically $\chi^2(m)$ statistic under the null hypothesis:

$$(s_T/\hat{\sigma}_1^{\dagger})^2\cdot\chi_T^2 = \{\mathbf{R}\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \mathbf{r}\}'\left\{(\hat{\sigma}_1^{\dagger})^2[\mathbf{0} \quad \mathbf{R}]\begin{bmatrix} T & \Sigma \mathbf{y}_{2t}' \\ \Sigma \mathbf{y}_{2t} & \Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}' \end{bmatrix}\right\}^{-1}\{\mathbf{R}\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \mathbf{r}\}$$

$$= \{\mathbf{R}\cdot T(\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \boldsymbol{\gamma})\}'\left\{(\hat{\sigma}_1^{\dagger})^2[\mathbf{0} \quad \mathbf{R}]\right.$$

$$\left.\times \begin{bmatrix} 1 & T^{-3/2}\Sigma \mathbf{y}_{2t}' \\ T^{-3/2}\Sigma \mathbf{y}_{2t} & T^{-2}\Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}' \end{bmatrix}\right\}^{-1}\{\mathbf{R}\cdot T(\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \boldsymbol{\gamma})\}$$

$$\xrightarrow{L} (\sigma_1^{\dagger})^2(\mathbf{R}\boldsymbol{\nu}_2)'\left\{(\sigma_1^{\dagger})^2[\mathbf{0} \quad \mathbf{R}]\mathbf{H}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}' \end{bmatrix}\right\}^{-1}(\mathbf{R}\boldsymbol{\nu}_2)$$

$$\sim \chi^2(m).$$

This description has assumed that there was no drift in any elements of the system. Hansen (1992) showed that the procedure is easily modified if $E(\Delta\mathbf{y}_{2t}) = \boldsymbol{\delta}_2 \neq \mathbf{0}$, simply by replacing $\hat{\mathbf{u}}_{2t} = \Delta\mathbf{y}_{2t}$ in [19.3.52] with

$$\hat{\mathbf{u}}_{2t} = \Delta\mathbf{y}_{2t} - \hat{\boldsymbol{\delta}}_2,$$

where

$$\hat{\boldsymbol{\delta}}_2 = T^{-1}\sum_{t=1}^{T}\Delta\mathbf{y}_{2t}.$$

Hansen also showed that a time trend could be added to the cointegrating relation, as in

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + \delta t + z_t^*,$$

for which the fully modified estimator is

$$\begin{bmatrix} \hat{\alpha}_T^{\dagger\dagger} \\ \hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} \\ \hat{\delta}_T^{\dagger\dagger} \end{bmatrix} = \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' & \Sigma t \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' & \Sigma\mathbf{y}_{2t}t \\ \Sigma t & \Sigma t\mathbf{y}_{2t}' & \Sigma t^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma\hat{y}_{1t}^{\dagger} \\ \Sigma\mathbf{y}_{2t}\hat{y}_{1t}^{\dagger} - T\hat{\boldsymbol{\aleph}}_T^{\dagger} \\ \Sigma t\hat{y}_{1t}^{\dagger} \end{bmatrix}.$$

Collecting these estimates in a vector $\mathbf{b}_T^{\dagger\dagger} = (\hat{\alpha}_T^{\dagger\dagger}, [\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger}]', \hat{\delta}_T^{\dagger\dagger})'$, a hypothesis involving m restrictions on $\boldsymbol{\beta}$ of the form $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ can be tested by

$$\{\mathbf{R}\mathbf{b}_T^{\dagger\dagger} - \mathbf{r}\}'\left\{(\hat{\sigma}_1^{\dagger})^2\mathbf{R}\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' & \Sigma t \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' & \Sigma\mathbf{y}_{2t}t \\ \Sigma t & \Sigma t\mathbf{y}_{2t}' & \Sigma t^2 \end{bmatrix}^{-1}\mathbf{R}'\right\}^{-1}\{\mathbf{R}\mathbf{b}_T^{\dagger\dagger} - \mathbf{r}\} \xrightarrow{L} \chi^2(m).$$

Park's Canonical Cointegrating Regressions

A closely related idea has been suggested by Park (1992). In Park's procedure, both the dependent and explanatory variables in [19.3.36] are transformed, and the resulting transformed regression can then be estimated by *OLS* and tested using standard procedures. Park and Ogaki (1991) explored the use of the *VAR* prewhitening technique of Andrews and Monahan (1992) to replace the Bartlett estimate in expressions such as [19.3.51].

APPENDIX 19.A. *Proofs of Chapter 19 Propositions*

■ **Proof of Proposition 19.2.** Define $\bar{y}_{1t} \equiv z_1^* + z_2^* + \cdots + z_t^*$ for $t = 1, 2, \ldots, T$ and $\bar{y}_{1,0} \equiv 0$. Then

$$\begin{bmatrix} \bar{y}_{1t} \\ \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y}_{2,0} \end{bmatrix} + \boldsymbol{\xi}_t^*,$$

where

$$\boldsymbol{\xi}_t^* \equiv \sum_{s=1}^{t}\begin{bmatrix} z_s^* \\ \mathbf{u}_{2s} \end{bmatrix}.$$

Hence, result (e) of Proposition 18.1 establishes that

$$T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \bar{y}_{1,t-1} \\ \mathbf{y}_{2,t-1} \end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}'] \xrightarrow{L} \boldsymbol{\Lambda}^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\boldsymbol{\Lambda}^{*\prime} + \sum_{v=1}^{\infty}\boldsymbol{\Gamma}_v^{*\prime} \qquad [19.A.1]$$

for

$$\Lambda^* \equiv \Psi^*(1)\cdot \mathbf{P}$$

$$\Gamma_v^{*\prime} \equiv E\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t} \end{bmatrix}[z_{t+v}^* \quad \mathbf{u}_{2,t+v}'].$$

It follows from [19.A.1] that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \bar{y}_{1t} \\ \mathbf{y}_{2t}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}'] &= T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \bar{y}_{1,t-1} \\ \mathbf{y}_{2,t-1}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}'] + T^{-1}\sum_{t=1}^{T}\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}'] \\ &\xrightarrow{L} \Lambda^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\Lambda^{*\prime} + \sum_{v=0}^{\infty}\Gamma_v^{*\prime}. \end{aligned} \qquad [19.A.2]$$

Similarly, results (a), (g), and (i) of Proposition 18.1 imply

$$T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t}\end{bmatrix} \xrightarrow{L} \Lambda^*\cdot\mathbf{W}(1) \qquad [19.A.3]$$

$$T^{-3/2}\sum_{t=1}^{T}\begin{bmatrix} \bar{y}_{1t} \\ \mathbf{y}_{2t}\end{bmatrix} \xrightarrow{L} \Lambda^*\cdot\int_0^1 \mathbf{W}(r)\,dr \qquad [19.A.4]$$

$$T^{-2}\sum_{t=1}^{T}\begin{bmatrix} \bar{y}_{1t} \\ \mathbf{y}_{2t}\end{bmatrix}[\bar{y}_{1t} \quad \mathbf{y}_{2t}'] \xrightarrow{L} \Lambda^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\cdot\Lambda^{*\prime}. \qquad [19.A.5]$$

Observe that the deviations of the *OLS* estimates in [19.2.12] from the population values α and γ that describe the cointegrating relation [19.2.9] are given by

$$\begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\gamma}_T - \gamma \end{bmatrix} = \begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} \Sigma z_t^* \\ \Sigma\mathbf{y}_{2t}z_t^* \end{bmatrix},$$

from which

$$\begin{aligned} \begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\gamma}_T - \gamma) \end{bmatrix} &= \left\{\begin{bmatrix} T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & T^{-1}\cdot\mathbf{I}_g \end{bmatrix}\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}\right. \\ &\quad \times \left.\begin{bmatrix} T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & T^{-1}\cdot\mathbf{I}_g \end{bmatrix}\right\}^{-1}\left\{\begin{bmatrix} T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & T^{-1}\cdot\mathbf{I}_g \end{bmatrix}\begin{bmatrix} \Sigma z_t^* \\ \Sigma\mathbf{y}_{2t}z_t^* \end{bmatrix}\right\} \qquad [19.A.6] \\ &= \begin{bmatrix} 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}\begin{bmatrix} T^{-1/2}\Sigma z_t^* \\ T^{-1}\Sigma\mathbf{y}_{2t}z_t^* \end{bmatrix}. \end{aligned}$$

But from [19.A.2],

$$\begin{aligned} T^{-1}\Sigma\mathbf{y}_{2t}z_t^* &= [\mathbf{0} \quad \mathbf{I}_g]T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \bar{y}_{1t} \\ \mathbf{y}_{2t}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}']\begin{bmatrix} 1 \\ \mathbf{0}\end{bmatrix} \\ &\xrightarrow{L} [\mathbf{0} \quad \mathbf{I}_g]\Lambda^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\Lambda^{*\prime}\begin{bmatrix} 1 \\ \mathbf{0}\end{bmatrix} \qquad [19.A.7] \\ &\quad + [\mathbf{0} \quad \mathbf{I}_g]\sum_{v=0}^{\infty}\Gamma_v^{*\prime}\begin{bmatrix} 1 \\ \mathbf{0}\end{bmatrix} \\ &= \Lambda_2^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\lambda_1^* + \sum_{v=0}^{\infty}E(\mathbf{u}_{2t}z_{t+v}^*). \end{aligned}$$

Similar use of [19.A.3] to [19.A.5] in [19.A.6] produces [19.2.13]. ■

■ **Proof of Proposition 19.3.** For simplicity of exposition, the discussion is restricted to the case when $E(\Delta\mathbf{y}_{2t}) = \mathbf{0}$, though it is straightforward to develop analogous results using a rescaling and rotation of variables similar to that in [18.2.43].

Consider first what the results would be from an OLS regression of z_{1t}^* on $\mathbf{z}_{2t}^* = (z_{2t}^*, z_{3t}^*, \ldots, z_{ht}^*)'$, a constant, and $\mathbf{y}_{2t}$:

$$z_{1t}^* = \boldsymbol{\beta}'\mathbf{z}_{2t}^* + \alpha^* + \boldsymbol{\aleph}^{*\prime}\mathbf{y}_{2t} + u_t. \qquad [19.A.8]$$

If this regression is evaluated at the true values $\alpha^* = 0$, $\boldsymbol{\aleph}^* = \mathbf{0}$, and $\boldsymbol{\beta} = (\beta_2, \beta_3, \ldots, \beta_h)'$ the vector of population projection coefficients in [19.2.18], then the disturbance u_t will be the residual defined in [19.2.18]. This residual had mean zero and was uncorrelated with $\mathbf{z}_{2t}^*$. The OLS estimates based on [19.A.8] would be

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_T \\ \hat{\alpha}_T^* \\ \hat{\boldsymbol{\aleph}}_T^* \end{bmatrix} = \begin{bmatrix} \Sigma\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime} & \Sigma\mathbf{z}_{2t}^* & \Sigma\mathbf{z}_{2t}^*\mathbf{y}_{2t}' \\ \Sigma\mathbf{z}_{2t}^{*\prime} & T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t}\mathbf{z}_{2t}^{*\prime} & \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \Sigma\mathbf{z}_{2t}^* z_{1t}^* \\ \Sigma z_{1t}^* \\ \Sigma\mathbf{y}_{2t} z_{1t}^* \end{bmatrix}. \qquad [19.A.9]$$

The deviations of these estimates from the corresponding population values satisfy

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} \\ \hat{\alpha}_T^* \\ T^{1/2}\hat{\boldsymbol{\aleph}}_T^* \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{h-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & T^{1/2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} \Sigma\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime} & \Sigma\mathbf{z}_{2t}^* & \Sigma\mathbf{z}_{2t}^*\mathbf{y}_{2t}' \\ \Sigma\mathbf{z}_{2t}^{*\prime} & T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t}\mathbf{z}_{2t}^{*\prime} & \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1}$$
$$\times \begin{bmatrix} T\cdot\mathbf{I}_{h-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & T & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & T^{3/2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} T\cdot\mathbf{I}_{h-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & T & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & T^{3/2}\mathbf{I}_g \end{bmatrix}^{-1} \begin{bmatrix} \Sigma\mathbf{z}_{2t}^* u_t \\ \Sigma u_t \\ \Sigma\mathbf{y}_{2t}u_t \end{bmatrix}$$
$$= \begin{bmatrix} T^{-1}\Sigma\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime} & T^{-1}\Sigma\mathbf{z}_{2t}^* & T^{-3/2}\Sigma\mathbf{z}_{2t}^*\mathbf{y}_{2t}' \\ T^{-1}\Sigma\mathbf{z}_{2t}^{*\prime} & 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t}\mathbf{z}_{2t}^{*\prime} & T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\Sigma\mathbf{z}_{2t}^* u_t \\ T^{-1}\Sigma u_t \\ T^{-3/2}\Sigma\mathbf{y}_{2t}u_t \end{bmatrix}.$$

[19.A.10]

Recalling that $E(\mathbf{z}_{2t}^* u_t) = \mathbf{0}$, one can show that $T^{-1}\Sigma\mathbf{z}_{2t}^* u_t \xrightarrow{p} \mathbf{0}$ and $T^{-1}\Sigma u_t \xrightarrow{p} 0$ by the law of large numbers. Also, $T^{-3/2}\Sigma\mathbf{y}_{2t}u_t \xrightarrow{p} \mathbf{0}$, from the argument given in [19.A.7]. Furthermore,

$$\begin{bmatrix} T^{-1}\Sigma\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime} & T^{-1}\Sigma\mathbf{z}_{2t}^* & T^{-3/2}\Sigma\mathbf{z}_{2t}^*\mathbf{y}_{2t}' \\ T^{-1}\Sigma\mathbf{z}_{2t}^{*\prime} & 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t}\mathbf{z}_{2t}^{*\prime} & T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}$$
$$\xrightarrow{L} \begin{bmatrix} E(\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \left\{\int [\mathbf{W}(r)]'\,dr\right\}\boldsymbol{\Lambda}_2^{*\prime} \\ \mathbf{0} & \boldsymbol{\Lambda}_2^*\int \mathbf{W}(r)\,dr & \boldsymbol{\Lambda}_2^*\left\{\int [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\boldsymbol{\Lambda}_2^{*\prime} \end{bmatrix}, \qquad [19.A.11]$$

where $\mathbf{W}(r)$ is n-dimensional standard Brownian motion and $\boldsymbol{\Lambda}_2^*$ is a $(g \times n)$ matrix constructed from the last g rows of $\boldsymbol{\Psi}^*(1)\cdot\mathbf{P}$. Notice that the matrix in [19.A.11] is almost surely nonsingular. Substituting these results into [19.A.10] establishes that

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} \\ \hat{\alpha}_T^* \\ T^{1/2}\hat{\boldsymbol{\aleph}}_T^* \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mathbf{0} \\ 0 \\ \mathbf{0} \end{bmatrix},$$

so that OLS estimation of [19.A.8] would produce consistent estimates of the parameters of the population linear projection [19.2.18].

An OLS regression of y_{1t} on a constant and the other elements of $\mathbf{y}_t$ is a simple transformation of the regression in [19.A.8]. To see this, notice that [19.A.8] can be written as

$$[1 \quad -\boldsymbol{\beta}']\mathbf{z}_t^* = \alpha^* + \boldsymbol{\aleph}^{*\prime}\mathbf{y}_{2t} + u_t. \qquad [19.A.12]$$

Solving [19.2.16] for $\mathbf{z}_t^*$ and substituting the result into [19.A.12] gives

$$[1 \quad -\boldsymbol{\beta}'](\mathbf{y}_{1t} - \boldsymbol{\mu}_1^* - \boldsymbol{\Gamma}'\mathbf{y}_{2t}) = \alpha^* + \boldsymbol{\aleph}^{*\prime}\mathbf{y}_{2t} + u_t,$$

or, since $\mathbf{y}_{1t} = (y_{1t}, y_{2t}, \ldots, y_{ht})'$, we have

$$y_{1t} = \beta_2 y_{2t} + \beta_3 y_{3t} + \cdots + \beta_h y_{ht} + \alpha + \boldsymbol{\aleph}'\mathbf{y}_{2t} + u_t, \qquad [19.A.13]$$

where $\alpha \equiv \alpha^* + [1 \quad -\boldsymbol{\beta}']\boldsymbol{\mu}_1^*$ and $\boldsymbol{\aleph}' \equiv \boldsymbol{\aleph}^{*\prime} + [1 \quad -\boldsymbol{\beta}']\boldsymbol{\Gamma}'$.

OLS estimation of [19.A.8] will produce identical fitted values to those resulting from *OLS* estimation of [19.A.13], with the relations between the estimated coefficients as just given. Since *OLS* estimation of [19.A.8] yields consistent estimates of [19.2.18], *OLS* estimation of [19.A.13] yields consistent estimates of the corresponding transformed parameters, as claimed by the proposition. ■

■ **Proof of Proposition 19.4.** As in Proposition 18.2, partition $\boldsymbol{\Lambda\Lambda}'$ as

$$\underset{(n\times n)}{\boldsymbol{\Lambda\Lambda}'} = \begin{bmatrix} \underset{(1\times 1)}{\Sigma_{11}} & \underset{(1\times g)}{\boldsymbol{\Sigma}_{21}'} \\ \underset{(g\times 1)}{\boldsymbol{\Sigma}_{21}} & \underset{(g\times g)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}, \qquad [19.A.14]$$

and define

$$\mathbf{L}' \equiv \begin{bmatrix} (1/\sigma_1^*) & (-1/\sigma_1^*)\cdot\boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{L}_{22}' \end{bmatrix}, \qquad [19.A.15]$$

where

$$(\sigma_1^*)^2 \equiv (\Sigma_{11} - \boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) \qquad [19.A.16]$$

and $\mathbf{L}_{22}$ is the Cholesky factor of $\boldsymbol{\Sigma}_{22}^{-1}$:

$$\boldsymbol{\Sigma}_{22}^{-1} = \mathbf{L}_{22}\mathbf{L}_{22}'. \qquad [19.A.17]$$

Recall from expression [18.A.16] that

$$\mathbf{L}'\boldsymbol{\Lambda\Lambda}'\mathbf{L} = \mathbf{I}_n, \qquad [19.A.18]$$

implying that $\boldsymbol{\Lambda\Lambda}' = (\mathbf{L}')^{-1}(\mathbf{L})^{-1}$ and $(\boldsymbol{\Lambda\Lambda}')^{-1} = \mathbf{LL}'$; thus, $\mathbf{L}$ is the Cholesky factor of $(\boldsymbol{\Lambda\Lambda}')^{-1}$ referred to in Proposition 19.4.

Note further that the residuals from *OLS* estimation of [19.2.24] are identical to the residuals from *OLS* estimation of

$$y_{1t}^* = \alpha^* + \boldsymbol{\gamma}^{*\prime}\mathbf{y}_{2t}^* + u_t^* \qquad [19.A.19]$$

for $y_{1t}^* \equiv y_{1t} - \boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1}\mathbf{y}_{2t}$ and $\mathbf{y}_{2t}^* \equiv \mathbf{L}_{22}'\mathbf{y}_{2t}$. Recall from equation [18.A.21] that

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T^*/\sigma_1^* \\ \hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix} \xrightarrow{L} \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix}. \qquad [19.A.20]$$

Finally, for the derivations that are to follow,

$$T^* \equiv T - 1.$$

Proof of (a). Since the sample residuals $\hat{u}_t^*$ for *OLS* estimation of [19.A.19] are identical to those for *OLS* estimation of [19.2.24], we have that

$$\begin{aligned} T^*(\hat{\rho}_T - 1) &= T^*\left\{\frac{\sum_{t=2}^{T}\hat{u}_{t-1}^*\hat{u}_t^*}{\sum_{t=2}^{T}(\hat{u}_{t-1}^*)^2} - 1\right\} \\ &= \frac{(T^*)^{-1}\sum_{t=2}^{T}\hat{u}_{t-1}^*(\hat{u}_t^* - \hat{u}_{t-1}^*)}{(T^*)^{-2}\sum_{t=2}^{T}(\hat{u}_{t-1}^*)^2}. \end{aligned} \qquad [19.A.21]$$

But

$$\begin{aligned}\hat{u}_t^* &= \sigma_1^* \cdot \{(y_{1t}^*/\sigma_1^*) - (1/\sigma_1^*) \cdot \hat{\boldsymbol{\gamma}}_T^{*\prime} \mathbf{y}_{2t}^* - (\hat{\alpha}_T^*/\sigma_1^*)\} \\ &\equiv \sigma_1^* \cdot \{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}_t^* - (\hat{\alpha}_T^*/\sigma_1^*)\}\end{aligned} \qquad [19.A.22]$$

for

$$\boldsymbol{\xi}_t^* \equiv \begin{bmatrix} y_{1t}^*/\sigma_1^* \\ \mathbf{y}_{2t}^* \end{bmatrix} = \mathbf{L}'\mathbf{y}_t. \qquad [19.A.23]$$

Differencing [19.A.22] results in

$$(\hat{u}_t^* - \hat{u}_{t-1}^*) = \sigma_1^* \cdot [1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\Delta\boldsymbol{\xi}_t^*. \qquad [19.A.24]$$

Using [19.A.22] and [19.A.24], the numerator of [19.A.21] can be written

$$\begin{aligned}&(T^*)^{-1} \sum_{t=2}^{T} \hat{u}_{t-1}^*(\hat{u}_t^* - \hat{u}_{t-1}^*) \\ &= (\sigma_1^*)^2 \cdot (T^*)^{-1} \sum_{t=2}^{T} \left\{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}_{t-1}^* - (\hat{\alpha}_T^*/\sigma_1^*)\right\}\left\{(\Delta\boldsymbol{\xi}_t^{*\prime})\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix}\right\} \\ &= (\sigma_1^*)^2 \cdot [1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix} \\ &\quad - (\sigma_1^*)^2 \cdot (T^*)^{-1/2}(\hat{\alpha}_T^*/\sigma_1^*) \cdot \left\{(T^*)^{-1/2} \sum_{t=2}^{T} (\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix}.\end{aligned} \qquad [19.A.25]$$

Notice that the expression

$$[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix}$$

is a scalar and accordingly equals its own transpose:

$$\begin{aligned}&[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix} \\ &= (1/2)\left\{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix}\right. \\ &\qquad \left. + [1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} (\Delta\boldsymbol{\xi}_t^*)(\boldsymbol{\xi}_{t-1}^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix}\right\} \\ &= (1/2)\left\{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\left\{(T^*)^{-1} \sum_{t=2}^{T} \left(\boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime}) + (\Delta\boldsymbol{\xi}_t^*)(\boldsymbol{\xi}_{t-1}^{*\prime})\right)\right\}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^* \end{bmatrix}\right\}.\end{aligned} \qquad [19.A.26]$$

But from result (d) of Proposition 18.1,

$$\begin{aligned}&(T^*)^{-1} \sum_{t=2}^{T} \left(\boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime}) + (\Delta\boldsymbol{\xi}_t^*)(\boldsymbol{\xi}_{t-1}^{*\prime})\right) \\ &\qquad = \mathbf{L}' \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \left(\mathbf{y}_{t-1}(\Delta\mathbf{y}_t') + (\Delta\mathbf{y}_t)(\mathbf{y}_{t-1}')\right)\right\} \cdot \mathbf{L} \qquad [19.A.27] \\ &\qquad \xrightarrow{L} \mathbf{L}' \cdot \{\boldsymbol{\Lambda} \cdot [\mathbf{W}(1)] \cdot [\mathbf{W}(1)]' \cdot \boldsymbol{\Lambda}' - E[(\Delta\mathbf{y}_t)(\Delta\mathbf{y}_t')]\} \cdot \mathbf{L} \\ &\qquad = [\mathbf{W}^*(1)] \cdot [\mathbf{W}^*(1)]' - E[(\Delta\boldsymbol{\xi}_t^*)(\Delta\boldsymbol{\xi}_t^{*\prime})]\end{aligned}$$

for $\mathbf{W}^*(r) \equiv \mathbf{L}'\mathbf{\Lambda}\cdot\mathbf{W}(r)$ the n-dimensional standard Brownian motion discussed in equation [18.A.17]. Substituting [19.A.27] and [19.A.20] into [19.A.26] produces

$$\begin{aligned}&[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\cdot\left\{(T^*)^{-1}\sum_{t=2}^{T}\boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\\ &\qquad\xrightarrow{L}(1/2)[1 \quad -\mathbf{h}_2']\{[\mathbf{W}^*(1)]\cdot[\mathbf{W}^*(1)]' - E[(\Delta\boldsymbol{\xi}_t^*)(\Delta\boldsymbol{\xi}_t^{*\prime})]\}\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}.\end{aligned} \qquad [19.A.28]$$

Similar analysis of the second term in [19.A.25] using result (a) of Proposition 18.1 reveals that

$$(T^*)^{-1/2}(\hat{\alpha}_T^*/\sigma_1^*)\cdot\left\{(T^*)^{-1/2}\sum_{t=2}^{T}(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\xrightarrow{L}h_1\cdot[\mathbf{W}^*(1)]'\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}. \qquad [19.A.29]$$

Substituting [19.A.28] and [19.A.29] into [19.A.25], we conclude that

$$\begin{aligned}&(T^*)^{-1}\sum_{t=2}^{T}\hat{u}_{t-1}^*(\hat{u}_t^* - \hat{u}_{t-1}^*)\\ &\quad\xrightarrow{L}(\sigma_1^*)^2\cdot\left\{\frac{1}{2}\left\{[1 \quad -\mathbf{h}_2']\cdot[\mathbf{W}^*(1)]\cdot[\mathbf{W}^*(1)]'\cdot\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}\right\} - h_1\cdot[\mathbf{W}^*(1)]'\cdot\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}\right.\\ &\qquad\left. - (1/2)\cdot[1 \quad -\mathbf{h}_2']\cdot\{E[(\Delta\boldsymbol{\xi}_t^*)(\Delta\boldsymbol{\xi}_t^{*\prime})]\}\cdot\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}\right\}.\end{aligned} \qquad [19.A.30]$$

The limiting distribution for the denominator of [19.A.21] was obtained in result (b) of Proposition 18.2:

$$(T^*)^{-2}\sum_{t=2}^{T}\hat{u}_{t-1}^2\xrightarrow{L}(\sigma_1^*)^2\cdot H_n. \qquad [19.A.31]$$

Substituting [19.A.30] and [19.A.31] into [19.A.21] produces [19.2.36].

Proof of (b). Notice that

$$\begin{aligned}\hat{c}_{j,T} &= (T^*)^{-1}\sum_{t=j+2}^{T}\hat{e}_t\hat{e}_{t-j}\\ &= (T^*)^{-1}\sum_{t=j+2}^{T}(\hat{u}_t^* - \hat{\rho}_T\hat{u}_{t-1}^*)(\hat{u}_{t-j}^* - \hat{\rho}_T\hat{u}_{t-j-1}^*)\\ &= (T^*)^{-1}\sum_{t=j+2}^{T}\{\Delta\hat{u}_t^* - (\hat{\rho}_T - 1)\hat{u}_{t-1}^*\}\cdot\{\Delta\hat{u}_{t-j}^* - (\hat{\rho}_T - 1)\hat{u}_{t-j-1}^*\}.\end{aligned} \qquad [19.A.32]$$

But [19.A.22] and [19.A.24] can be used to write

$$\begin{aligned}&(T^*)^{-1}\sum_{t=j+2}^{T}(\hat{\rho}_T - 1)\hat{u}_{t-1}^*\Delta\hat{u}_{t-j}^*\\ &\quad= (\sigma_1^*)^2\cdot(\hat{\rho}_T - 1)\cdot(T^*)^{-1}\sum_{t=j+2}^{T}\left\{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}_{t-1}^* - (\hat{\alpha}_T^*/\sigma_1^*)\right\}(\Delta\boldsymbol{\xi}_{t-j}^{*\prime})\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\\ &\quad= \left\{(\sigma_1^*)^2\cdot[(T^*)^{1/2}(\hat{\rho}_T - 1)]\cdot[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\cdot(T^*)^{-3/2}\sum_{t=j+2}^{T}\boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_{t-j}^{*\prime})\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\right\}\\ &\qquad- \left\{(\sigma_1^*)^2\cdot[(T^*)^{1/2}(\hat{\rho}_T - 1)]\cdot[(T^*)^{-1/2}(\hat{\alpha}_T^*/\sigma_1^*)](T^*)^{-1}\sum_{t=j+2}^{T}(\Delta\boldsymbol{\xi}_{t-j}^{*\prime})\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\right\}.\end{aligned} \qquad [19.A.33]$$

But result (a) implies that $(T^*)^{1/2}(\hat{\rho}_T - 1)\xrightarrow{p}0$, while the other terms in [19.A.33] have convergent distributions in the light of [19.A.20] and results (a) and (e) of Proposition 18.1.

Hence,

$$(T^*)^{-1}\sum_{t=j+2}^{T}(\hat\rho_T - 1)\hat u^*_{t-1}\Delta\hat u^*_{t-j} \xrightarrow{p} 0. \qquad [19.A.34]$$

Similarly,

$$\begin{aligned}
&(T^*)^{-1}\sum_{t=j+2}^{T}(\hat\rho_T - 1)^2\hat u^*_{t-1}\hat u^*_{t-j-1} \\
&\quad = (\sigma_1^*)^2\cdot(T^*)^{-1}\sum_{t=j+2}^{T}(\hat\rho_T - 1)^2\left\{[1 \quad -\hat{\boldsymbol\gamma}_T^{*\prime}/\sigma_1^*]\boldsymbol\xi^*_{t-1} - (\hat\alpha_T^*/\sigma_1^*)\right\} \\
&\qquad \times\left\{[1 \quad -\hat{\boldsymbol\gamma}_T^{*\prime}/\sigma_1^*]\boldsymbol\xi^*_{t-j-1} - (\hat\alpha_T^*/\sigma_1^*)\right\} \\
&\quad = (\sigma_1^*)^2\cdot(T^*)^{-1}\sum_{t=j+2}^{T}(\hat\rho_T - 1)^2\begin{bmatrix}1 & -\hat{\boldsymbol\gamma}_T^{*\prime}/\sigma_1^* & -(T^*)^{-1/2}\hat\alpha_T^*/\sigma_1^*\end{bmatrix}\begin{bmatrix}\boldsymbol\xi^*_{t-1} \\ (T^*)^{1/2}\end{bmatrix} \\
&\qquad \times[\boldsymbol\xi^{*\prime}_{t-j-1} \quad (T^*)^{1/2}]\,[1 \quad -\hat{\boldsymbol\gamma}_T^{*\prime}/\sigma_1^* \quad -(T^*)^{-1/2}\hat\alpha_T^*/\sigma_1^*]' \qquad [19.A.35] \\
&\quad = (\sigma_1^*)^2\cdot[(T^*)^{1/2}(\hat\rho_T - 1)]^2\cdot[1 \quad -\hat{\boldsymbol\gamma}_T^{*\prime}/\sigma_1^* \quad -(T^*)^{-1/2}\hat\alpha_T^*/\sigma_1^*] \\
&\qquad \times\left\{(T^*)^{-2}\sum_{t=j+2}^{T}\begin{bmatrix}\boldsymbol\xi^*_{t-1}\boldsymbol\xi^{*\prime}_{t-j-1} & (T^*)^{1/2}\boldsymbol\xi^*_{t-1} \\ (T^*)^{1/2}\boldsymbol\xi^{*\prime}_{t-j-1} & T^*\end{bmatrix}\right\} \\
&\qquad \times[1 \quad -\hat{\boldsymbol\gamma}_T^{*\prime}/\sigma_1^* \quad -(T^*)^{-1/2}\hat\alpha_T^*/\sigma_1^*]' \\
&\quad \xrightarrow{p} 0,
\end{aligned}$$

given that $(T^*)^{-2}\Sigma_{t=j+2}^{T}\boldsymbol\xi^*_{t-1}\cdot\boldsymbol\xi^{*\prime}_{t-j-1}$ and $(T^*)^{-3/2}\Sigma\boldsymbol\xi^*_{t-s}$ are $O_p(1)$ by results (i) and (g) of Proposition 18.1. Substituting [19.A.34], [19.A.35], and then [19.A.24] into [19.A.32] gives

$$\begin{aligned}
\hat c_{j,T} &\xrightarrow{p} (T^*)^{-1}\sum_{t=j+2}^{T}(\Delta\hat u_t^*)\cdot(\Delta\hat u^*_{t-j}) \\
&= (\sigma_1^*)^2\cdot[1 \quad -\hat{\boldsymbol\gamma}_T^{*\prime}/\sigma_1^*](T^*)^{-1}\sum_{t=j+2}^{T}(\Delta\boldsymbol\xi_t^*)\cdot(\Delta\boldsymbol\xi^{*\prime}_{t-j})\begin{bmatrix}1 \\ -\hat{\boldsymbol\gamma}_T^*/\sigma_1^*\end{bmatrix} \qquad [19.A.36] \\
&\xrightarrow{L} (\sigma_1^*)^2\cdot[1 \quad -\mathbf h_2']\cdot E\{(\Delta\boldsymbol\xi_t^*)\cdot(\Delta\boldsymbol\xi^{*\prime}_{t-j})\}\begin{bmatrix}1 \\ -\mathbf h_2\end{bmatrix} \\
&= (\sigma_1^*)^2\cdot[1 \quad -\mathbf h_2']\cdot\mathbf L'\cdot E\{(\Delta\mathbf y_t)\cdot(\Delta\mathbf y'_{t-j})\}\cdot\mathbf L\begin{bmatrix}1 \\ -\mathbf h_2\end{bmatrix}.
\end{aligned}$$

It follows that for given q,

$$\begin{aligned}
\hat\lambda_T^2 &= \hat c_{0,T} + 2\cdot\sum_{j=1}^{q}[1 - j/(q+1)]\hat c_{j,T} \\
&\xrightarrow{L}(\sigma_1^*)^2\cdot[1 \quad -\mathbf h_2']\cdot\mathbf L'\left\{\sum_{j=-q}^{q}[1 - |j|/(q+1)]\cdot E[(\Delta\mathbf y_t)\cdot(\Delta\mathbf y'_{t-j})]\right\}\cdot\mathbf L\cdot\begin{bmatrix}1 \\ -\mathbf h_2\end{bmatrix}.
\end{aligned}$$

Thus, if $q \to \infty$ with $q/T \to 0$,

$$\begin{aligned}
\hat\lambda_T^2 &\xrightarrow{L} (\sigma_1^*)^2\cdot[1 \quad -\mathbf h_2']\cdot\mathbf L'\cdot\left\{\sum_{j=-\infty}^{\infty}E[(\Delta\mathbf y_t)\cdot(\Delta\mathbf y'_{t-j})]\right\}\cdot\mathbf L\cdot\begin{bmatrix}1 \\ -\mathbf h_2\end{bmatrix} \\
&= (\sigma_1^*)^2\cdot[1 \quad -\mathbf h_2']\cdot\mathbf L'\boldsymbol\Psi(1)\mathbf P\mathbf P'[\boldsymbol\Psi(1)]'\mathbf L\cdot\begin{bmatrix}1 \\ -\mathbf h_2\end{bmatrix} \qquad [19.A.37] \\
&= (\sigma_1^*)^2\cdot[1 \quad -\mathbf h_2']\cdot\mathbf I_n\cdot\begin{bmatrix}1 \\ -\mathbf h_2\end{bmatrix},
\end{aligned}$$

by virtue of [19.A.18].

But from [19.2.29] and [19.A.31],

$$(T^*)^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2 = \frac{1}{(T^*)^{-2}\sum_{t=2}^{T}\hat{u}_{t-1}^2} \xrightarrow{L} \frac{1}{(\sigma_1^*)^2\cdot H_n}. \qquad [19.A.38]$$

It then follows from [19.A.36] and [19.A.37] that

$$\{(T^*)^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2\}\cdot\{\hat{\lambda}_T^2 - \hat{c}_{0,T}\} \xrightarrow{L} [1 \quad -\mathbf{h}_2']\cdot\{\mathbf{I}_n - (\mathbf{L}'\cdot E[(\Delta\mathbf{y}_t)\cdot(\Delta\mathbf{y}_t')]\cdot\mathbf{L})\}\cdot\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix} \div H_n. \qquad [19.A.39]$$

Subtracting $\frac{1}{2}$ times [19.A.39] from [19.2.36] yields [19.2.37].

Proof of (c). Notice from [19.2.33] that

$$\begin{aligned} Z_{t,T} &= (1/\hat{\lambda}_T)\cdot\left\{(\hat{c}_{0,T}/s_T^2)^{1/2}\frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}\div s_T} - (1/2)\cdot\{T^*\cdot\hat{\sigma}_{\hat{\rho}_T} \div s_T\}\cdot\{\hat{\lambda}_T^2 - \hat{c}_{0,T}\}\right\} \\ &= (1/\hat{\lambda}_T)\frac{1}{T^*\cdot\hat{\sigma}_{\hat{\rho}_T}\div s_T}\left\{(\hat{c}_{0,T}/s_T^2)^{1/2}T^*(\hat{\rho}_T - 1) - (1/2)\cdot\{(T^*)^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2\}\cdot\{\hat{\lambda}_T^2 - \hat{c}_{0,T}\}\right\}. \end{aligned} \qquad [19.A.40]$$

But since

$$(\hat{c}_{0,T}/s_T^2) = (T-2)/(T-1) \rightarrow 1,$$

it follows that

$$\begin{aligned} Z_{t,T} &\xrightarrow{P} (1/\hat{\lambda}_T)\frac{1}{T^*\cdot\hat{\sigma}_{\hat{\rho}_T}\div s_T}Z_{\rho,T} \\ &\xrightarrow{L} \frac{1}{\sigma_1^*\cdot(1 + \mathbf{h}_2'\mathbf{h}_2)^{1/2}}(\sigma_1^*\cdot\sqrt{H_n})Z_n, \end{aligned} \qquad [19.A.41]$$

with the last line following from [19.A.37], [19.A.38], and [19.2.37].

Proof of (d). See Phillips and Ouliaris (1990). ■

Chapter 19 Exercises

19.1. Let

$$\begin{bmatrix}\Delta y_{1t}\\ \Delta y_{2t}\end{bmatrix} = \begin{bmatrix}\delta_1\\ \delta_2\end{bmatrix} + \begin{bmatrix}u_{1t}\\ u_{2t}\end{bmatrix},$$

where $\delta_2 \neq 0$ and δ_1 may or may not be zero. Let $\mathbf{u}_t \equiv (u_{1t}, u_{2t})'$, and suppose that $\mathbf{u}_t = \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t$ for $\boldsymbol{\varepsilon}_t$ an i.i.d. (2×1) vector with mean zero, variance $\mathbf{PP}'$, and finite fourth moments. Assume further that $\{s\cdot\mathbf{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable and that $\mathbf{\Psi}(1)\cdot\mathbf{P}$ is nonsingular. Define $\xi_{1t} \equiv \Sigma_{s=1}^{t}u_{1s}$, $\xi_{2t} \equiv \Sigma_{s=1}^{t}u_{2s}$, and $\gamma_0 \equiv \delta_1/\delta_2$.

(a) Show that the *OLS* estimates of

$$y_{1t} = \alpha + \gamma y_{2t} + u_t$$

satisfy

$$\begin{bmatrix} T^{-1/2}\hat{\alpha}_T \\ T^{1/2}(\hat{\gamma}_T - \gamma_0) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} 1 & \delta_2/2 \\ \delta_2/2 & \delta_2^2/3 \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2}\Sigma(\xi_{1t} - \gamma_0\xi_{2t}) \\ T^{-5/2}\Sigma\delta_2 t(\xi_{1t} - \gamma_0\xi_{2t}) \end{bmatrix}.$$

Conclude that $\hat{\alpha}_T$ and $\hat{\gamma}_T$ have the same asymptotic distribution as the coefficients from a regression of $(\xi_{1t} - \gamma_0\xi_{2t})$ on a constant and δ_2 times a time trend.:

$$(\xi_{1t} - \gamma_0\xi_{2t}) = \alpha + \gamma\cdot\delta_2 t + u_t.$$

(b) Show that first differences of the OLS residuals converge to

$$\Delta\hat{u}_t \xrightarrow{p} u_{1t} - \gamma_0 u_{2t}.$$

19.2. Verify [19.3.23].

19.3. Verify [19.3.25].

19.4. Consider the regression model

$$y_{1t} = \boldsymbol{\beta}'\mathbf{w}_t + \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + \delta t + u_t,$$

where

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

Let $\Delta\mathbf{y}_{2t} = \mathbf{u}_{2t}$, where

$$\begin{bmatrix} u_t \\ \mathbf{u}_{2t} \end{bmatrix} = \bar{\boldsymbol{\Psi}}(L)\boldsymbol{\varepsilon}_t = \begin{bmatrix} \bar{\psi}_{11}(L) & \mathbf{0}' \\ \mathbf{0} & \bar{\boldsymbol{\Psi}}_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix}$$

and where $\boldsymbol{\varepsilon}_t$ is i.i.d. with mean zero, finite fourth moments, and variance

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \begin{bmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix} \begin{bmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{22}' \end{bmatrix}.$$

Suppose that $\{s\cdot\bar{\boldsymbol{\Psi}}_s\}_{s=0}^{\infty}$ is absolutely summable, $\bar{\lambda}_{11} \equiv \sigma_1\cdot\bar{\psi}_{11}(1) \neq 0$, and $\bar{\boldsymbol{\Lambda}}_{22} \equiv \bar{\boldsymbol{\Psi}}_{22}(1)\cdot\mathbf{P}_{22}$ is nonsingular. Show that the *OLS* estimates satisfy

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \\ T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{Q}^{-1}\mathbf{h}_1 \\ \bar{\lambda}_{11}\cdot\nu_1 \\ \bar{\lambda}_{11}\cdot\boldsymbol{\nu}_2 \\ \bar{\lambda}_{11}\cdot\nu_3 \end{bmatrix},$$

where $\mathbf{Q} = \text{plim } T^{-1}\Sigma\mathbf{w}_t\mathbf{w}_t'$, $T^{-1/2}\Sigma\mathbf{w}_t u_t \xrightarrow{L} \mathbf{h}_1$, and

$$\begin{bmatrix} \nu_1 \\ \boldsymbol{\nu}_2 \\ \nu_3 \end{bmatrix} \equiv \mathbf{H}^{-1} \begin{bmatrix} W_1(1) \\ \bar{\boldsymbol{\Lambda}}_{22}\cdot\left\{\int [\mathbf{W}_2(r)]\, dW_1(r)\right\} \\ \left\{W_1(1) - \int W_1(r)\, dr\right\} \end{bmatrix}$$

$$\mathbf{H} \equiv \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\, dr\right\}\bar{\boldsymbol{\Lambda}}_{22}' & 1/2 \\ \bar{\boldsymbol{\Lambda}}_{22}\int \mathbf{W}_2(r)\, dr & \bar{\boldsymbol{\Lambda}}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\, dr\right\}\bar{\boldsymbol{\Lambda}}_{22}' & \bar{\boldsymbol{\Lambda}}_{22}\int r\mathbf{W}_2(r)\, dr \\ 1/2 & \left\{\int r[\mathbf{W}_2(r)]'\, dr\right\}\bar{\boldsymbol{\Lambda}}_{22}' & 1/3 \end{bmatrix}.$$

Reason as in [19.3.12] that conditional on $\mathbf{W}_2(\cdot)$, the vector $(\nu_1, \boldsymbol{\nu}_2', \nu_3)'$ is Gaussian with mean zero and variance $\mathbf{H}^{-1}$. Use this to show that the Wald form of the *OLS* χ^2 test of any m restrictions involving α, $\boldsymbol{\gamma}$, or δ converges to $(\bar{\lambda}_{11}^2/s_T^2)$ times a $\chi^2(m)$ variable.

19.5. Consider the regression model

$$y_{1t} = \boldsymbol{\beta}'\mathbf{w}_t + \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + u_t,$$

where

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

Suppose that

$$\Delta\mathbf{y}_{2t} = \boldsymbol{\delta}_2 + \mathbf{u}_{2t},$$

where at least one of the elements of $\boldsymbol{\delta}_2$ is nonzero. Let u_t and $\mathbf{u}_{2t}$ satisfy the same conditions as in Exercise 19.4.

Let $\mathbf{y}_{2t} = (y_{2t}, y_{3t}, \ldots, y_{nt})'$ and $\boldsymbol{\delta}_2 = (\delta_2, \delta_3, \ldots, \delta_n)'$, and suppose that the elements of $\mathbf{y}_{2t}$ are ordered so that $E(\Delta y_{nt}) = \delta_n \neq 0$. Notice that the fitted values for the regression are identical to those of

$$y_{1t} = \boldsymbol{\beta}'\mathbf{w}_t^* + \alpha^* + \boldsymbol{\gamma}^{*\prime}\mathbf{y}_{2t}^* + \delta^* y_{nt} + u_t,$$

where

$$\mathbf{w}_t^* = [(\Delta\mathbf{y}_{2,t-p} - \boldsymbol{\delta}_2)', (\Delta\mathbf{y}_{2,t-p+1} - \boldsymbol{\delta}_2)', \ldots, (\Delta\mathbf{y}_{2,t+p} = \boldsymbol{\delta}_2)']'$$

$$\underset{[(g-1)\times 1]}{\mathbf{y}_{2t}^*} \equiv \begin{bmatrix} y_{2t} - (\delta_2/\delta_n)y_{nt} \\ y_{3t} - (\delta_3/\delta_n)y_{nt} \\ \vdots \\ y_{n-1,t} - (\delta_{n-1}/\delta_n)y_{nt} \end{bmatrix}$$

$$\boldsymbol{\gamma}^* \equiv \begin{bmatrix} \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_{n-1} \end{bmatrix}$$

$$\delta^* = \gamma_n + \gamma_2(\delta_2/\delta_n) + \gamma_3(\delta_3/\delta_n) + \cdots + \gamma_{n-1}(\delta_{n-1}/\delta_n)$$

$$\alpha^* = \alpha + \boldsymbol{\beta}'(\mathbf{1} \otimes \boldsymbol{\delta}_2),$$

with $\mathbf{1}$ a $[(2p + 1) \times 1]$ column of 1s.

Show that the asymptotic properties of the transformed regression are identical to those of the time trend regression in Exercise 19.4. Conclude that any F test involving $\boldsymbol{\gamma}$ in the original regression can be multiplied by $(s_T^2/\hat{\lambda}_{11}^2)$ and compared with the usual F tables for an asymptotically valid test.

Chapter 19 References

Ahn, S. K., and G. C. Reinsel. 1990. "Estimation for Partially Nonstationary Multivariate Autoregressive Models." *Journal of the American Statistical Association* 85:813–23.

Anderson, T. W. 1958. *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Andrews, Donald W. K., and J. Christopher Monahan. 1992. "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica* 60:953–66.

Baillie, Richard T., and David D. Selover. 1987. "Cointegration and Models of Exchange Rate Determination." *International Journal of Forecasting* 3:43–51.

Campbell, John Y., and Robert J. Shiller. 1988a. "Interpreting Cointegrated Models." *Journal of Economic Dynamics and Control* 12:505–22.

——— and ———. 1988b. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1:195–228.

Clarida, Richard. 1991. "Co-Integration, Aggregate Consumption, and the Demand for Imports: A Structural Econometric Investigation." Columbia University. Mimeo.

Corbae, Dean, and Sam Ouliaris. 1988. "Cointegration and Tests of Purchasing Power Parity." *Review of Economics and Statistics* 70:508–11.

Davidson, James E. H., David F. Hendry, Frank Srba, and Stephen Yeo. 1978. "Econometric Modelling of the Aggregate Time-Series Relationship between Consumers' Expenditure and Income in the United Kingdom." *Economic Journal* 88:661–92.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55:251–76.

——— and Byung Sam Yoo. 1987. "Forecasting and Testing in Co-Integrated Systems." *Journal of Econometrics* 35:143–59.

Granger, C. W. J. 1983. "Co-Integrated Variables and Error-Correcting Models." Unpublished University of California, San Diego, Discussion Paper 83-13.

——— and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2:111–20.

Hansen, Bruce E. 1990. "A Powerful, Simple Test for Cointegration Using Cochrane-Orcutt." University of Rochester. Mimeo.

———. 1992. "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends." *Journal of Econometrics* 53:87–121.

Haug, Alfred A. 1992. "Critical Values for the $\hat{Z}_\alpha$-Phillips-Ouliaris Test for Cointegration." *Oxford Bulletin of Economics and Statistics* 54:473–80.

Johansen, Søren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12:231–54.

———. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59:1551–80.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. 1991. "Stochastic Trends and Economic Fluctuations." *American Economic Review* 81:819–40.

Kremers, Jeroen J. M. 1989. "U.S. Federal Indebtedness and the Conduct of Fiscal Policy." *Journal of Monetary Economics* 23:219–38.

Mosconi, Rocco, and Carlo Giannini. 1992. "Non-Causality in Cointegrated Systems: Representation, Estimation and Testing." *Oxford Bulletin of Economics and Statistics* 54:399–417.

Ogaki, Masao. 1992. "Engel's Law and Cointegration." *Journal of Political Economy* 100:1027–46.

——— and Joon Y. Park. 1992. "A Cointegration Approach to Estimating Preference Parameters." Department of Economics, University of Rochester. Mimeo.

Park, Joon Y. 1992. "Canonical Cointegrating Regressions." *Econometrica* 60:119–43.

——— and Masao Ogaki. 1991. "Inference in Cointegrated Models Using VAR Prewhitening to Estimate Shortrun Dynamics." University of Rochester. Mimeo.

———, S. Ouliaris, and B. Choi. 1988. "Spurious Regressions and Tests for Cointegration." Cornell University. Mimeo.

Phillips, Peter C. B. 1987. "Time Series Regression with a Unit Root." *Econometrica* 55:277–301.

———. 1991. "Optimal Inference in Cointegrated Systems." *Econometrica* 59:283–306.

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

——— and Bruce E. Hansen. 1990. "Statistical Inference in Instrumental Variables Regression with I(1) Processes." *Review of Economic Studies* 57:99–125.

——— and Mico Loretan. 1991. "Estimating Long-Run Economic Equilibria." *Review of Economic Studies* 58:407–36.

——— and S. Ouliaris. 1990. "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica* 58:165–93.

Saikkonen, Pentti. 1991. "Asymptotically Efficient Estimation of Cointegration Regressions." *Econometric Theory* 7:1–21.

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.

Stock, James H. 1987. "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors." *Econometrica* 55:1035–56.

———. 1990. "A Class of Tests for Integration and Cointegration." Harvard University. Mimeo.

Stock, James H., and Mark W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* 83:1097–1107.

——— and ———. 1993. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica* 61:783–820.

Wooldridge, Jeffrey M. 1991. "Notes on Regression with Difference-Stationary Data." Michigan State University. Mimeo.

20 Full-Information Maximum Likelihood Analysis of Cointegrated Systems

An $(n \times 1)$ vector $\mathbf{y}_t$ was said to exhibit h cointegrating relations if there exist h linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ such that $\mathbf{a}_i'\mathbf{y}_t$ is stationary. If such vectors exist, their values are not uniquely defined, since any linear combinations of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ would also be described as cointegrating vectors. The approaches described in the previous chapter sidestepped this problem by imposing normalization conditions such as $a_{11} = 1$. For this normalization we would put y_{1t} on the left side of a regression and the other elements of $\mathbf{y}_t$ on the right side. We might equally well have normalized $a_{12} = 1$ instead, in which case y_{2t} would be the variable that belongs on the left side of the regression. The results obtained in practice can thus depend on an essentially arbitrary assumption. Furthermore, if the first variable does not appear in the cointegrating relation at all ($a_{11} = 0$), then setting $a_{11} = 1$ is not a harmless normalization but instead results in a fundamentally misspecified model.

For these reasons there is some value in using full-information maximum likelihood (*FIML*) to estimate the linear space spanned by the cointegrating vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$. This chapter describes the solution to this problem developed by Johansen (1988, 1991), whose work is closely related to that of Ahn and Reinsel (1990), and more distantly to that of Stock and Watson (1988). Another advantage of *FIML* is that it allows us to test for the number of cointegrating relations. The approach of Phillips and Ouliaris (1990) described in Chapter 19 tested the null hypothesis that there are no cointegrating relations. This chapter presents more general tests of the null hypothesis that there are h_0 cointegrating relations, where h_0 could be $0, 1, \ldots,$ or $n - 1$.

To develop these ideas, Section 20.1 begins with a discussion of canonical correlation analysis. Section 20.2 then develops the *FIML* estimates, while Section 20.3 describes hypothesis testing in cointegrated systems. Section 20.4 offers a brief overview of unit roots in time series analysis.

20.1. Canonical Correlation

Population Canonical Correlations

Let the $(n_1 \times 1)$ vector $\mathbf{y}_t$ and the $(n_2 \times 1)$ vector $\mathbf{x}_t$ denote stationary random variables. Typically, $\mathbf{y}_t$ and $\mathbf{x}_t$ are measured as deviations from their population means, so that $E(\mathbf{y}_t\mathbf{y}_t')$ represents the variance-covariance matrix of $\mathbf{y}_t$. In general, there might be complicated correlations among the elements of $\mathbf{y}_t$ and $\mathbf{x}_t$, summarized by the joint variance-covariance matrix

$$\begin{bmatrix} \underset{(n_1\times n_1)}{E(\mathbf{y}_t\mathbf{y}_t')} & \underset{(n_1\times n_2)}{E(\mathbf{y}_t\mathbf{x}_t')} \\ \underset{(n_2\times n_1)}{E(\mathbf{x}_t\mathbf{y}_t')} & \underset{(n_2\times n_2)}{E(\mathbf{x}_t\mathbf{x}_t')} \end{bmatrix} = \begin{bmatrix} \underset{(n_1\times n_1)}{\Sigma_{\mathbf{YY}}} & \underset{(n_1\times n_2)}{\Sigma_{\mathbf{YX}}} \\ \underset{(n_2\times n_1)}{\Sigma_{\mathbf{XY}}} & \underset{(n_2\times n_2)}{\Sigma_{\mathbf{XX}}} \end{bmatrix}.$$

We can often gain some insight into the nature of these correlations by defining two new $(n \times 1)$ random vectors, $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$, where n is the smaller of n_1 and n_2. These vectors are linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$, respectively:

$$\boldsymbol{\eta}_t \equiv \mathcal{K}'\mathbf{y}_t \tag{20.1.1}$$

$$\boldsymbol{\xi}_t \equiv \mathcal{A}'\mathbf{x}_t. \tag{20.1.2}$$

Here $\mathcal{K}'$and $\mathcal{A}'$ are $(n \times n_1)$ and $(n \times n_2)$ matrices, respectively. The matrices $\mathcal{K}'$ and $\mathcal{A}'$ are chosen so that the following conditions hold.

(1) The individual elements of $\boldsymbol{\eta}_t$ have unit variance and are uncorrelated with one another:

$$E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \mathcal{K}'\Sigma_{\mathbf{YY}}\mathcal{K} = \mathbf{I}_n. \tag{20.1.3}$$

(2) The individual elements of $\boldsymbol{\xi}_t$ have unit variance and are uncorrelated with one another:

$$E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \mathcal{A}'\Sigma_{\mathbf{XX}}\mathcal{A} = \mathbf{I}_n. \tag{20.1.4}$$

(3) The ith element of $\boldsymbol{\eta}_t$ is uncorrelated with the jth element of $\boldsymbol{\xi}_t$ for $i \neq j$; for $i = j$, the correlation is positive and is given by r_i:

$$E(\boldsymbol{\xi}_t\boldsymbol{\eta}_t') = \mathcal{A}'\Sigma_{\mathbf{XY}}\mathcal{K} = \mathbf{R}, \tag{20.1.5}$$

where

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix}. \tag{20.1.6}$$

(4) The elements of $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ are ordered in such a way that

$$(1 \geq r_1 \geq r_2 \geq \cdots \geq r_n \geq 0). \tag{20.1.7}$$

The correlation r_i is known as the ith *population canonical correlation* between $\mathbf{y}_t$ and $\mathbf{x}_t$.

The population canonical correlations and the values of $\mathcal{K}$ and $\mathcal{A}$ can be calculated from $\Sigma_{\mathbf{YY}}$, $\Sigma_{\mathbf{XX}}$, and $\Sigma_{\mathbf{XY}}$ using any computer program that generates eigenvalues and eigenvectors, as we now describe.

Let $(\lambda_1, \lambda_2, \ldots, \lambda_{n_1})$ denote the eigenvalues of the $(n_1 \times n_1)$ matrix

$$\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}, \tag{20.1.8}$$

ordered as

$$(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_1}), \tag{20.1.9}$$

with associated eigenvectors $(\tilde{\mathbf{k}}_1, \tilde{\mathbf{k}}_2, \ldots, \tilde{\mathbf{k}}_{n_1})$. Recall that the eigenvalue-eigenvector pair $(\lambda_i, \tilde{\mathbf{k}}_i)$ satisfies

$$\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\tilde{\mathbf{k}}_i = \lambda_i\tilde{\mathbf{k}}_i. \tag{20.1.10}$$

Notice that if $\tilde{\mathbf{k}}_i$ satisfies [20.1.10], then so does $c\tilde{\mathbf{k}}_i$ for any value of c. The usual normalization convention for choosing c and thus for determining "the" eigenvector $\tilde{\mathbf{k}}_i$ to associate with λ_i is to set $\tilde{\mathbf{k}}_i'\tilde{\mathbf{k}}_i = 1$. For canonical correlation analysis, however, it is more convenient to choose c so as to ensure that

$$\mathbf{k}_i'\Sigma_{\mathbf{YY}}\mathbf{k}_i = 1 \qquad \text{for } i = 1, 2, \ldots, n_1. \tag{20.1.11}$$

If a computer program has calculated eigenvectors $(\tilde{\mathbf{k}}_1, \tilde{\mathbf{k}}_2, \ldots, \tilde{\mathbf{k}}_{n_1})$ of the matrix in [20.1.8] normalized by $(\tilde{\mathbf{k}}_i'\tilde{\mathbf{k}}_i) = 1$, it is trivial to change these to eigenvectors $(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{n_1})$ normalized by the condition [20.1.11] by setting

$$\mathbf{k}_i = \tilde{\mathbf{k}}_i \div \sqrt{\tilde{\mathbf{k}}_i'\Sigma_{\mathbf{YY}}\tilde{\mathbf{k}}_i}.$$

We further may multiply $\mathbf{k}_i$ by -1 so as to satisfy a certain sign convention to be detailed in the paragraphs following the next proposition.

The canonical correlations $(r_1, r_2, \ldots, r_n)$ turn out to be given by the square roots of the corresponding first n eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ of [20.1.8]. The associated $(n_1 \times 1)$ eigenvectors $\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_n$, when normalized by [20.1.11] and a sign convention, turn out to make up the rows of the $(n \times n_1)$ matrix $\mathcal{K}'$ appearing in [20.1.1]. The matrix $\mathcal{A}'$ in [20.1.2] can be obtained from the normalized eigenvectors of a matrix closely related to [20.1.8]. These results are developed in the following proposition, proved in Appendix 20.A at the end of this chapter.

Proposition 20.1: *Let*

$$\underset{(n_1+n_2)\times(n_1+n_2)}{\Sigma} \equiv \begin{bmatrix} \underset{(n_1\times n_1)}{\Sigma_{\mathbf{YY}}} & \underset{(n_1\times n_2)}{\Sigma_{\mathbf{YX}}} \\ \underset{(n_2\times n_1)}{\Sigma_{\mathbf{XY}}} & \underset{(n_2\times n_2)}{\Sigma_{\mathbf{XX}}} \end{bmatrix}$$

be a positive definite symmetric matrix and let $(\lambda_1, \lambda_2, \ldots, \lambda_{n_1})$ be the eigenvalues of the matrix in [20.1.8], ordered $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_1}$. Let $(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{n_1})$ be the associated $(n_1 \times 1)$ eigenvectors as normalized by [20.1.11]. Let $(\mu_1, \mu_2, \ldots, \mu_{n_2})$ be the eigenvalues of the $(n_2 \times n_2)$ matrix

$$\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}, \tag{20.1.12}$$

ordered $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{n_2}$. Let $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n_2})$ be the eigenvectors of [20.1.12]:

$$\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\mathbf{a}_i = \mu_i\mathbf{a}_i, \tag{20.1.13}$$

normalized by

$$\mathbf{a}_i'\Sigma_{\mathbf{XX}}\mathbf{a}_i = 1 \qquad \textit{for } i = 1, 2, \ldots, n_2. \tag{20.1.14}$$

Let n be the smaller of n_1 and n_2, and collect the first n vectors $\mathbf{k}_i$ and the first n vectors $\mathbf{a}_i$ in matrices

$$\underset{(n_1\times n)}{\mathcal{K}} = [\mathbf{k}_1 \quad \mathbf{k}_2 \quad \cdots \quad \mathbf{k}_n]$$
$$\underset{(n_2\times n)}{\mathcal{A}} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n].$$

Assuming that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct, then

(a) $0 \leq \lambda_i < 1$ *for* $i = 1, 2, \ldots, n_1$ *and* $0 \leq \mu_j < 1$ *for* $j = 1, 2, \ldots, n_2$;
(b) $\lambda_i = \mu_i$ *for* $i = 1, 2, \ldots, n$;
(c) $\mathcal{K}'\Sigma_{\mathbf{YY}}\mathcal{K} = \mathbf{I}_n$ *and* $\mathcal{A}'\Sigma_{\mathbf{XX}}\mathcal{A} = \mathbf{I}_n$:
(d) $\mathcal{A}'\Sigma_{\mathbf{XY}}\mathcal{K} = \mathbf{R}$,

where $\mathbf{R}$ is a diagonal matrix whose squared diagonal elements correspond to the *eigenvalues of* [20.1.8]:

$$\mathbf{R}^2 = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

If Σ denotes the variance-covariance matrix of the vector $(\mathbf{y}_t', \mathbf{x}_t')'$, then results (c) and (d) are the characterization of the canonical correlations given in [20.1.3] through [20.1.5]. Thus, the proposition establishes that the squares of the canonical correlations $(r_1^2, r_2^2, \ldots, r_n^2)$ can be found from the first n eigenvalues of the matrix in [20.1.8]. Result (b) states that these are the same as the first n eigenvalues of the matrix in [20.1.12]. The matrices $\mathcal{K}$ and $\mathcal{A}$ that characterize the canonical variates in [20.1.1] and [20.1.2] can be found from the normalized eigenvectors of these matrices.

The magnitude $\mathbf{a}_i'\Sigma_{\mathbf{XY}}\mathbf{k}_i$ calculated by the algorithm described in Proposition 20.1 need not be positive—the proposition only ensures that its square is equal to the square of the corresponding canonical correlation. If $\mathbf{a}_i'\Sigma_{\mathbf{XY}}\mathbf{k}_i < 0$ for some i, one can replace $\mathbf{k}_i$ as calculated with $-\mathbf{k}_i$, so that the ith diagonal element of $\mathbf{R}$ will correspond to the positive square root of λ_i.

As an illustration, suppose that $\mathbf{y}_t$ consists of a single variable ($n_1 = n = 1$). In this case, the matrix [20.1.8] is just a scalar, a (1×1) "matrix" that is equal to its own eigenvalue. Thus, the squared population canonical correlation between a scalar y_t and a set of n_2 explanatory variables $\mathbf{x}_t$ is given by

$$r_1^2 = \frac{\Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{X}Y}}{\Sigma_{YY}}.$$

To interpret this expression, recall from equation [4.1.15] that the mean squared error of a linear projection of y_t on $\mathbf{x}_t$ is given by

$$MSE = \Sigma_{YY} - \Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{X}Y},$$

and so

$$1 - r_1^2 = \frac{\Sigma_{YY}}{\Sigma_{YY}} - \frac{\Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{X}Y}}{\Sigma_{YY}} = \frac{MSE}{\Sigma_{YY}}. \qquad [20.1.15]$$

Thus, for this simple case, r_1^2 is the fraction of the population variance that is explained by the linear projection; that is, r_1^2 is the population squared multiple correlation coefficient, commonly denoted R^2.

Another interpretation of canonical correlations is also sometimes helpful. The first canonical variates η_{1t} and ξ_{1t} can be interpreted as those linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$, respectively, such that the correlation between η_{1t} and ξ_{1t} is as large as possible (see Exercise 20.1). The variates η_{2t} and ξ_{2t} give those linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$ that are uncorrelated with η_{1t} and ξ_{1t} and yet yield the largest remaining correlation between η_{2t} and ξ_{2t}, and so on.

Sample Canonical Correlations

The canonical correlations r_i calculated by the procedure just described are population parameters—they are functions of the population moments $\Sigma_{\mathbf{YY}}$, $\Sigma_{\mathbf{YX}}$, and $\Sigma_{\mathbf{XX}}$. Here we describe their sample analogs, to be denoted $\hat{r}_i$.

Suppose we have a sample of T observations on the $(n_1 \times 1)$ vector $\mathbf{y}_t$ and the $(n_2 \times 1)$ vector $\mathbf{x}_t$, whose sample moments are given by

$$\hat{\Sigma}_{YY} = (1/T) \sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t' \qquad [20.1.16]$$

$$\hat{\Sigma}_{YX} = (1/T) \sum_{t=1}^{T} \mathbf{y}_t\mathbf{x}_t' \qquad [20.1.17]$$

$$\hat{\Sigma}_{XX} = (1/T) \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'. \qquad [20.1.18]$$

Again, in many applications, $\mathbf{y}_t$ and $\mathbf{x}_t$ would be measured in deviations from their sample means.

To calculate sample canonical correlations, the objective is to generate a set of T observations on a new $(n \times 1)$ vector $\hat{\boldsymbol{\eta}}_t$, where n is the smaller of n_1 and n_2. The vector $\hat{\boldsymbol{\eta}}_t$ is a linear combination of the observed value of $\mathbf{y}_t$:

$$\hat{\boldsymbol{\eta}}_t = \hat{\mathcal{K}}'\mathbf{y}_t, \qquad [20.1.19]$$

for $\hat{\mathcal{K}}$ an $(n_1 \times n)$ matrix to be estimated from the data. The task will be to choose $\hat{\mathcal{K}}$ so that the ith generated series $(\hat{\eta}_{it})$ has unit sample variance and is orthogonal to the jth generated series:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t' = \mathbf{I}_n. \qquad [20.1.20]$$

Similarly, we will generate an $(n \times 1)$ vector $\hat{\boldsymbol{\xi}}_t$ from the elements of $\mathbf{x}_t$:

$$\hat{\boldsymbol{\xi}}_t = \hat{\mathcal{A}}'\mathbf{x}_t. \qquad [20.1.21]$$

Each of the variables $\hat{\xi}_{it}$ has unit sample variance and is orthogonal to $\hat{\xi}_{jt}$ for $i \neq j$:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t' = \mathbf{I}_n. \qquad [20.1.22]$$

Finally, $\hat{\eta}_{it}$ is orthogonal to $\hat{\xi}_{jt}$ for $i \neq j$, while the sample correlation between $\hat{\eta}_{it}$ and $\hat{\xi}_{it}$ is called the *sample canonical correlation coefficient*:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\eta}}_t' = \hat{\mathbf{R}} \qquad [20.1.23]$$

for

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}_1 & 0 & \cdots & 0 \\ 0 & \hat{r}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \hat{r}_n \end{bmatrix}. \qquad [20.1.24]$$

Finding matrices $\hat{\mathcal{K}}$, $\hat{\mathcal{A}}$, and $\hat{\mathbf{R}}$ satisfying [20.1.20], [20.1.22], and [20.1.23] involves exactly the same calculations as did finding matrices $\mathcal{K}$, $\mathcal{A}$, and $\mathbf{R}$ satisfying [20.1.3] through [20.1.5]. For example, [20.1.19] allows us to write [20.1.20] as

$$\mathbf{I}_n = (1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t' = \hat{\mathcal{K}}'(1/T) \sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t'\,\hat{\mathcal{K}} = \hat{\mathcal{K}}'\hat{\Sigma}_{YY}\hat{\mathcal{K}}, \qquad [20.1.25]$$

where the last line follows from [20.1.16]. Expression [20.1.25] is identical to

[20.1.3] with hats placed over the variables. Similarly, substituting [20.1.21] into [20.1.22] gives $\hat{\mathcal{A}}'\hat{\Sigma}_{XX}\hat{\mathcal{A}} = \mathbf{I}_n$, which corresponds to [20.1.4]. Equation [20.1.23] becomes $\hat{\mathcal{A}}'\hat{\Sigma}_{XY}\hat{\mathcal{K}} = \hat{\mathbf{R}}$, as in [20.1.5]. Again, we can replace $\hat{\mathbf{k}}_i$ with $-\hat{\mathbf{k}}_i$ if any of the elements of $\hat{\mathbf{R}}$ should turn out negative.

Thus, to calculate the sample canonical correlations, the procedure described in Proposition 20.1 is simply applied to the sample moments ($\hat{\Sigma}_{YY}$, $\hat{\Sigma}_{YX}$, and $\hat{\Sigma}_{XX}$) rather than to the population moments. In particular, the square of the ith sample canonical correlation ($\hat{r}_i^2$) is given by the ith largest eigenvalue of the matrix

$$\begin{aligned} \hat{\Sigma}_{YY}^{-1}\hat{\Sigma}_{YX}\hat{\Sigma}_{XX}^{-1}\hat{\Sigma}_{XY} &= \left\{(1/T)\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_t'\right\} \\ &\quad\times\left\{(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{y}_t'\right\}. \end{aligned} \qquad [20.1.26]$$

The ith column of $\hat{\mathcal{K}}$ is given by the eigenvector associated with this ith eigenvalue, normalized so that

$$\hat{\mathbf{k}}_i'\left\{(1/T)\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right\}\hat{\mathbf{k}}_i = 1.$$

The ith column of $\hat{\mathcal{A}}$ is given by the eigenvector associated with the eigenvalue $\hat{\lambda}_i$ for the matrix $\hat{\Sigma}_{XX}^{-1}\hat{\Sigma}_{XY}\hat{\Sigma}_{YY}^{-1}\hat{\Sigma}_{YX}$ normalized by the condition that $\hat{\mathbf{a}}_i'\hat{\Sigma}_{XX}\hat{\mathbf{a}}_i = 1$.

For example, suppose that y_t is a scalar ($n = n_1 = 1$). Then [20.1.26] is a scalar equal to its own eigenvalue. Hence, the sample squared canonical correlation between the scalar y_t and a set of n_2 explanatory variables $\mathbf{x}_t$ is given by

$$\begin{aligned} \hat{r}_1^2 &= \frac{\{T^{-1}\Sigma y_t\mathbf{x}_t'\}\{T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t'\}^{-1}\{T^{-1}\Sigma\mathbf{x}_t y_t\}}{\{T^{-1}\Sigma y_t^2\}} \\ &= \frac{\{\Sigma y_t\mathbf{x}_t'\}\{\Sigma\mathbf{x}_t\mathbf{x}_t'\}^{-1}\{\Sigma\mathbf{x}_t y_t\}}{\{\Sigma y_t^2\}}, \end{aligned}$$

which is just the squared sample multiple correlation coefficient R^2.

20.2. *Maximum Likelihood Estimation*

We are now in a position to describe Johansen's approach (1988, 1991) to full-information maximum likelihood estimation of a system characterized by exactly h cointegrating relations.

Let $\mathbf{y}_t$ denote an $(n \times 1)$ vector. The maintained hypothesis is that $\mathbf{y}_t$ follows a $VAR(p)$ in levels. Recall from equation [19.1.39] that any pth-order VAR can be written in the form

$$\begin{aligned} \Delta\mathbf{y}_t &= \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} \\ &\quad + \boldsymbol{\alpha} + \boldsymbol{\zeta}_0\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \end{aligned} \qquad [20.2.1]$$

with

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$$

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_\tau') = \begin{cases} \boldsymbol{\Omega} & \text{for } t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Suppose that each individual variable y_{it} is $I(1)$, although h linear combinations of $\mathbf{y}_t$ are stationary. We saw in equations [19.1.35] and [19.1.40] that this implies that $\boldsymbol{\zeta}_0$ can be written in the form

$$\boldsymbol{\zeta}_0 = -\mathbf{BA}' \qquad [20.2.2]$$

for $\mathbf{B}$ an $(n \times h)$ matrix and $\mathbf{A}'$ an $(h \times n)$ matrix. That is, under the hypothesis of h cointegrating relations, only h separate linear combinations of the level of $\mathbf{y}_{t-1}$ (the h elements of $\mathbf{z}_{t-1} = \mathbf{A}'\mathbf{y}_{t-1}$) appear in [20.2.1].

Consider a sample of $T + p$ observations on $\mathbf{y}$, denoted $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \ldots, \mathbf{y}_T)$. If the disturbances $\boldsymbol{\varepsilon}_t$ are Gaussian, then the log likelihood of $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$ conditional on $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \ldots, \mathbf{y}_0)$ is given by

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\Omega}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1}, \boldsymbol{\alpha}, \boldsymbol{\zeta}_0) \\
&\quad = (-Tn/2)\log(2\pi) - (T/2)\log|\boldsymbol{\Omega}| \\
&\qquad - (1/2)\sum_{t=1}^{T}\Big[(\Delta\mathbf{y}_t - \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} - \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} - \cdots - \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} - \boldsymbol{\alpha} - \boldsymbol{\zeta}_0\mathbf{y}_{t-1})' \\
&\qquad \times \boldsymbol{\Omega}^{-1}(\Delta\mathbf{y}_t - \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} - \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} - \cdots - \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} - \boldsymbol{\alpha} - \boldsymbol{\zeta}_0\mathbf{y}_{t-1})\Big].
\end{aligned} \qquad [20.2.3]$$

The goal is to chose $(\boldsymbol{\Omega}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1}, \boldsymbol{\alpha}, \boldsymbol{\zeta}_0)$ so as to maximize [20.2.3] subject to the constraint that $\boldsymbol{\zeta}_0$ can be written in the form of [20.2.2].

We will first summarize Johansen's algorithm, and then verify that it indeed calculates the maximum likelihood estimates.

Step 1: Calculate Auxiliary Regressions

The first step is to estimate a $(p-1)$th-order *VAR* for $\Delta\mathbf{y}_t$; that is, regress the scalar Δy_{it} on a constant and all the elements of the vectors $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$ by *OLS*. Collect the $i = 1, 2, \ldots, n$ *OLS* regressions in vector form as

$$\Delta\mathbf{y}_t = \hat{\boldsymbol{\pi}}_0 + \hat{\boldsymbol{\Pi}}_1\Delta\mathbf{y}_{t-1} + \hat{\boldsymbol{\Pi}}_2\Delta\mathbf{y}_{t-2} + \cdots + \hat{\boldsymbol{\Pi}}_{p-1}\Delta\mathbf{y}_{t-p+1} + \hat{\mathbf{u}}_t, \qquad [20.2.4]$$

where $\hat{\boldsymbol{\Pi}}_i$ denotes an $(n \times n)$ matrix of *OLS* coefficient estimates and $\hat{\mathbf{u}}_t$ denotes the $(n \times 1)$ vector of *OLS* residuals. We also estimate a second battery of regressions, regressing the scalar $y_{i,t-1}$ on a constant and $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$ for $i = 1, 2, \ldots, n$. Write this second set of *OLS* regressions as[1]

$$\mathbf{y}_{t-1} = \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\aleph}}_1\Delta\mathbf{y}_{t-1} + \hat{\boldsymbol{\aleph}}_2\Delta\mathbf{y}_{t-2} + \cdots + \hat{\boldsymbol{\aleph}}_{p-1}\Delta\mathbf{y}_{t-p+1} + \hat{\mathbf{v}}_t, \qquad [20.2.5]$$

with $\hat{\mathbf{v}}_t$ the $(n \times 1)$ vector of residuals from this second battery of regressions.

[1]Johansen (1991) described his procedure as calculating $\tilde{\mathbf{v}}_t$ in place of $\hat{\mathbf{v}}_t$, where $\tilde{\mathbf{v}}_t$ is the *OLS* residual from a regression of $\mathbf{y}_{t-p}$ on a constant and $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$. Since $\mathbf{y}_{t-p} = \mathbf{y}_{t-1} - \Delta\mathbf{y}_{t-1} - \Delta\mathbf{y}_{t-2} - \cdots - \Delta\mathbf{y}_{t-p+1}$, the residual $\tilde{\mathbf{v}}_t$ is numerically identical to $\hat{\mathbf{v}}_t$ described in the text.

Step 2: Calculate Canonical Correlations

Next calculate the sample variance-covariance matrices of the *OLS* residuals $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$:

$$\hat{\Sigma}_{VV} \equiv (1/T) \sum_{t=1}^{T} \hat{\mathbf{v}}_t\hat{\mathbf{v}}_t' \qquad [20.2.6]$$

$$\hat{\Sigma}_{UU} \equiv (1/T) \sum_{t=1}^{T} \hat{\mathbf{u}}_t\hat{\mathbf{u}}_t' \qquad [20.2.7]$$

$$\hat{\Sigma}_{UV} \equiv (1/T) \sum_{t=1}^{T} \hat{\mathbf{u}}_t\hat{\mathbf{v}}_t' \qquad [20.2.8]$$

$$\hat{\Sigma}_{VU} \equiv \hat{\Sigma}_{UV}'.$$

From these, find the eigenvalues of the matrix

$$\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}\hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV} \qquad [20.2.9]$$

with the eigenvalues ordered $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$. The maximum value attained by the log likelihood function subject to the constraint that there are h cointegrating relations is given by

$$\begin{aligned}\mathcal{L}^* = &-(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\Sigma}_{UU}| \\ &- (T/2)\sum_{i=1}^{h}\log(1 - \hat{\lambda}_i).\end{aligned} \qquad [20.2.10]$$

Step 3: Calculate Maximum Likelihood Estimates of Parameters

If we are interested only in a likelihood ratio test of the number of cointegrating relations, step 2 provides all the information needed. If maximum likelihood estimates of parameters are also desired, these can be calculated as follows. Let $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_h$ denote the $(n \times 1)$ eigenvectors of [20.2.9] associated with the h largest eigenvalues. These provide a basis for the space of cointegrating relations; that is, the maximum likelihood estimate is that any cointegrating vector can be written in the form

$$\mathbf{a} = b_1\hat{\mathbf{a}}_1 + b_2\hat{\mathbf{a}}_2 + \cdots + b_h\hat{\mathbf{a}}_h$$

for some choice of scalars $(b_1, b_2, \ldots ; b_h)$. Johansen suggested normalizing these vectors $\hat{\mathbf{a}}_i$ so that $\hat{\mathbf{a}}_i'\hat{\Sigma}_{VV}\hat{\mathbf{a}}_i = 1$. For example, if the eigenvectors $\tilde{\mathbf{a}}_i$ of [20.2.9] are calculated from a standard computer program that normalizes $\tilde{\mathbf{a}}_i'\tilde{\mathbf{a}}_i = 1$, Johansen's estimate is $\hat{\mathbf{a}}_i = \tilde{\mathbf{a}}_i \div \sqrt{\tilde{\mathbf{a}}_i'\hat{\Sigma}_{VV}\tilde{\mathbf{a}}_i}$. Collect the first h normalized vectors in an $(n \times h)$ matrix $\hat{\mathbf{A}}$:

$$\hat{\mathbf{A}} \equiv [\hat{\mathbf{a}}_1 \quad \hat{\mathbf{a}}_2 \quad \cdots \quad \hat{\mathbf{a}}_h]. \qquad [20.2.11]$$

Then the *MLE* of $\boldsymbol{\zeta}_0$ is given by

$$\hat{\boldsymbol{\zeta}}_0 = \hat{\Sigma}_{UV}\hat{\mathbf{A}}\hat{\mathbf{A}}'. \qquad [20.2.12]$$

The *MLE* of $\boldsymbol{\zeta}_i$ for $i = 1, 2, \ldots, p - 1$ is

$$\hat{\boldsymbol{\zeta}}_i = \hat{\boldsymbol{\Pi}}_i - \hat{\boldsymbol{\zeta}}_0\hat{\boldsymbol{\aleph}}_i, \qquad [20.2.13]$$

and the *MLE* of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\pi}}_0 - \hat{\boldsymbol{\zeta}}_0\hat{\boldsymbol{\theta}}. \qquad [20.2.14]$$

The *MLE* of $\boldsymbol{\Omega}$ is

$$\hat{\boldsymbol{\Omega}} = (1/T) \sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \hat{\boldsymbol{\zeta}}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \hat{\boldsymbol{\zeta}}_0\hat{\mathbf{v}}_t)']. \qquad [20.2.15]$$

We now review the logic behind each of these steps in turn.

Motivation for Auxiliary Regressions

The first step involves *concentrating* the likelihood function.[2] This means taking $\boldsymbol{\Omega}$ and $\boldsymbol{\zeta}_0$ as given and maximizing [20.2.3] with respect to $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$. This restricted maximization problem takes the form of seemingly unrelated regressions of the elements of the $(n \times 1)$ vector $\Delta\mathbf{y}_t - \boldsymbol{\zeta}_0\mathbf{y}_{t-1}$ on a constant and the explanatory variables $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$. Since each of the n regressions in this system has the identical explanatory variables, the estimates of $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$ would come from *OLS* regressions of each of the elements of $\Delta\mathbf{y}_t - \boldsymbol{\zeta}_0\mathbf{y}_{t-1}$ on a constant and $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$. Denote the values of $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$ that maximize [20.2.3] for a given value of $\boldsymbol{\zeta}_0$ by

$$[\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0), \ldots, \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0)].$$

These values are characterized by the condition that the following residual vector must have sample mean zero and be orthogonal to $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$:

$$\begin{aligned}[\Delta\mathbf{y}_t - \boldsymbol{\zeta}_0\mathbf{y}_{t-1}] - \{\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0) + \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0)\Delta\mathbf{y}_{t-1} + \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0)\Delta\mathbf{y}_{t-2} \\ + \cdots + \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0)\Delta\mathbf{y}_{t-p+1}\}.\end{aligned} \qquad [20.2.16]$$

But notice that the *OLS* residuals $\hat{\mathbf{u}}_t$ in [20.2.4] and $\hat{\mathbf{v}}_t$ in [20.2.5] each satisfy this orthogonality requirement, and therefore the vector $\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t$ also has sample mean zero and is orthogonal to $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$. Moreover, $\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t$ is of the form of expression [20.2.16],

$$\begin{aligned}\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t = (\Delta\mathbf{y}_t - \hat{\boldsymbol{\pi}}_0 - \hat{\boldsymbol{\Pi}}_1\Delta\mathbf{y}_{t-1} - \hat{\boldsymbol{\Pi}}_2\Delta\mathbf{y}_{t-2} - \cdots - \hat{\boldsymbol{\Pi}}_{p-1}\Delta\mathbf{y}_{t-p+1}) \\ - \boldsymbol{\zeta}_0(\mathbf{y}_{t-1} - \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\aleph}}_1\Delta\mathbf{y}_{t-1} - \hat{\boldsymbol{\aleph}}_2\Delta\mathbf{y}_{t-2} - \cdots - \hat{\boldsymbol{\aleph}}_{p-1}\Delta\mathbf{y}_{t-p+1}),\end{aligned}$$

with

$$\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0) = \hat{\boldsymbol{\pi}}_0 - \boldsymbol{\zeta}_0\hat{\boldsymbol{\theta}} \qquad [20.2.17]$$

$$\hat{\boldsymbol{\zeta}}_i^*(\boldsymbol{\zeta}_0) = \hat{\boldsymbol{\Pi}}_i - \boldsymbol{\zeta}_0\hat{\boldsymbol{\aleph}}_i \qquad \text{for } i = 1, 2, \ldots, p-1. \qquad [20.2.18]$$

Thus, the vector in [20.2.16] is given by $\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t$.

The concentrated log likelihood function (to be denoted $\mathcal{M}$) is found by replacing $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$ in [20.2.3] with $[\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0), \ldots, \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0)]$:

$$\begin{aligned}\mathcal{M}(\boldsymbol{\Omega}, \boldsymbol{\zeta}_0) &\equiv \mathcal{L}\{\boldsymbol{\Omega}, \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0), \ldots, \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0), \boldsymbol{\zeta}_0\} \\ &= -(Tn/2)\log(2\pi) - (T/2)\log|\boldsymbol{\Omega}| \\ &\quad - (1/2)\sum_{t=1}^{T}[(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)'\boldsymbol{\Omega}^{-1}(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)].\end{aligned} \qquad [20.2.19]$$

The idea behind concentrating the likelihood function in this way is that if we can find the values of $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\zeta}}_0$ for which $\mathcal{M}$ is maximized, then these same values (along with $\hat{\boldsymbol{\alpha}}^*(\hat{\boldsymbol{\zeta}}_0)$ and $\hat{\boldsymbol{\zeta}}_i^*(\hat{\boldsymbol{\zeta}}_0)$) will maximize [20.2.3].

[2]See Koopmans and Hood (1953, pp. 156–58) for more background on concentration of likelihood functions.

Continuing the concentration one step further, recall from the analysis of [11.1.25] that the value of $\mathbf{\Omega}$ that maximizes [20.2.19] (still regarding $\boldsymbol{\zeta}_0$ as fixed) is given by

$$\hat{\mathbf{\Omega}}^*(\boldsymbol{\zeta}_0) = (1/T)\sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)']. \qquad [20.2.20]$$

As in expression [11.1.32], the value obtained for [20.2.19] when evaluated at [20.2.20] is then

$$\begin{aligned} \mathcal{N}(\boldsymbol{\zeta}_0) &\equiv \mathcal{M}\{\hat{\mathbf{\Omega}}^*(\boldsymbol{\zeta}_0), \boldsymbol{\zeta}_0\} \\ &= -(Tn/2)\log(2\pi) - (T/2)\log|\hat{\mathbf{\Omega}}^*(\boldsymbol{\zeta}_0)| - (Tn/2) \\ &= -(Tn/2)\log(2\pi) - (Tn/2) \qquad [20.2.21] \\ &\quad - (T/2)\log\left|(1/T)\sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)']\right|. \end{aligned}$$

Expression [20.2.21] represents the biggest value one can achieve for the log likelihood for any given value of $\boldsymbol{\zeta}_0$. Maximizing the likelihood function thus comes down to choosing $\boldsymbol{\zeta}_0$ so as to minimize

$$\left|(1/T)\sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)']\right| \qquad [20.2.22]$$

subject to the constraint of [20.2.2].

Motivation for Canonical Correlation Analysis

To see the motivation for calculating canonical correlations, consider first a simpler problem. Suppose that by an astounding coincidence, $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$ were already in canonical form,

$$\hat{\mathbf{u}}_t = \hat{\boldsymbol{\eta}}_t$$
$$\hat{\mathbf{v}}_t = \hat{\boldsymbol{\xi}}_t,$$

with

$$(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t' = \mathbf{I}_n \qquad [20.2.23]$$

$$(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t' = \mathbf{I}_n \qquad [20.2.24]$$

$$(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\eta}}_t' = \hat{\mathbf{R}} \qquad [20.2.25]$$

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}_1 & 0 & \cdots & 0 \\ 0 & \hat{r}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \hat{r}_n \end{bmatrix}. \qquad [20.2.26]$$

Suppose that for these canonical data we were asked to choose $\boldsymbol{\zeta}_0$ so as to minimize

$$\left|(1/T)\sum_{t=1}^{T} \left[(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0\hat{\boldsymbol{\xi}}_t)(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0\hat{\boldsymbol{\xi}}_t)'\right]\right| \qquad [20.2.27]$$

subject to the constraint that $\boldsymbol{\zeta}_0 \hat{\boldsymbol{\xi}}_t$ could make use of only h linear combinations of $\hat{\boldsymbol{\xi}}_t$. If there were no restrictions on $\boldsymbol{\zeta}_0$ (so that $h = n$), then expression [20.2.27] would be minimized by OLS regressions of $\hat{\eta}_{it}$ on $\hat{\boldsymbol{\xi}}_t$ for $i = 1, 2, \ldots, n$. Conditions [20.2.24] and [20.2.25] establish that the ith regression would have an estimated coefficient vector of

$$\left\{(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t \hat{\boldsymbol{\xi}}_t'\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t \hat{\eta}_{it}\right\} = \hat{r}_i \cdot \mathbf{e}_i,$$

where $\mathbf{e}_i$ denotes the ith column of $\mathbf{I}_n$. Thus, even if all n elements of $\hat{\boldsymbol{\xi}}_t$ appeared in the regression, only the ith element $\hat{\xi}_{it}$ would have a nonzero coefficient in the regression used to explain $\hat{\eta}_{it}$. The average squared residual for this regression would be

$$\begin{aligned}
&\left\{(1/T)\sum_{t=1}^{T} (\hat{\eta}_{it})^2\right\} - \left\{(1/T)\sum_{t=1}^{T} (\hat{\eta}_{it}\hat{\boldsymbol{\xi}}_t')\right\}\left\{(1/T)\sum_{t=1}^{T} (\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t')\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T} (\hat{\boldsymbol{\xi}}_t\hat{\eta}_{it})\right\} \\
&\qquad = 1 - \hat{r}_i \cdot \mathbf{e}_i' \cdot \mathbf{I}_n \cdot \mathbf{e}_i \cdot \hat{r}_i \\
&\qquad = 1 - \hat{r}_i^2.
\end{aligned}$$

Moreover, conditions [20.2.23] through [20.2.25] imply that the residual for the ith regression, $\hat{\eta}_{it} - \hat{r}_i \hat{\xi}_{it}$, would be orthogonal to the residual from the jth regression, $\hat{\eta}_{jt} - \hat{r}_j \hat{\xi}_{jt}$, for $i \neq j$. Thus, if $\boldsymbol{\zeta}_0$ were unrestricted, the optimal value for the matrix in [20.2.27] would be a diagonal matrix with $(1 - \hat{r}_i^2)$ in the row i, column i position and zero elsewhere.

Now suppose that we are restricted to use only h linear combinations of $\hat{\boldsymbol{\xi}}_t$ as regressors. From the preceding analysis, we might guess that the best we can do is use the h elements of $\hat{\boldsymbol{\xi}}_t$ that have the highest correlations with elements of $\hat{\boldsymbol{\eta}}_t$, that is, choose $(\hat{\xi}_{1t}, \hat{\xi}_{2t}, \ldots, \hat{\xi}_{ht})$ as regressors.[3] When this set of regressors is used to explain $\hat{\eta}_{it}$ for $i \leq h$, the average squared residual will be $(1 - \hat{r}_i^2)$, as before. When this set of regressors is used to explain $\hat{\eta}_{it}$ for $i > h$, all of the regressors are orthogonal to $\hat{\eta}_{it}$ and would receive regression coefficients of zero. The average squared residual for the latter regression is simply $(1/T)\Sigma_{t=1}^{T} \hat{\eta}_{it}^2 = 1$ for $i = h + 1, h + 2, \ldots, n$. Thus, if we are restricted to using only h linear combinations of $\hat{\boldsymbol{\xi}}_t$, the optimized value of [20.2.27] will be

$$\begin{aligned}
&\left|(1/T)\sum_{t=1}^{T} [(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0^* \hat{\boldsymbol{\xi}}_t)(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0^* \hat{\boldsymbol{\xi}}_t)']\right| \\
&\qquad = \begin{vmatrix}
1 - \hat{r}_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 - \hat{r}_2^2 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
0 & 0 & \cdots & 1 - \hat{r}_h^2 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{vmatrix} \qquad \text{[20.2.28]} \\
&\qquad = \prod_{i=1}^{h} (1 - \hat{r}_i^2).
\end{aligned}$$

[3]See Johansen (1988) for a more formal demonstration of this claim.

Of course, the actual data $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$ will not be in exact canonical form. However, the previous section described how to find $(n \times n)$ matrices $\hat{\mathcal{K}}$ and $\hat{\mathcal{A}}$ such that

$$\hat{\boldsymbol{\eta}}_t = \hat{\mathcal{K}}'\hat{\mathbf{u}}_t \qquad [20.2.29]$$

$$\hat{\boldsymbol{\xi}}_t = \hat{\mathcal{A}}'\hat{\mathbf{v}}_t. \qquad [20.2.30]$$

The columns of $\hat{\mathcal{A}}$ are given by the eigenvectors of the matrix in [20.2.9], normalized by the condition $\hat{\mathcal{A}}'\hat{\boldsymbol{\Sigma}}_{\mathrm{VV}}\hat{\mathcal{A}} = \mathbf{I}_n$. The eigenvalues of [20.2.9] give the squares of the canonical correlations:

$$\hat{\lambda}_i = \hat{r}_i^2. \qquad [20.2.31]$$

The columns of $\hat{\mathcal{K}}$ correspond to the normalized eigenvectors of the matrix $\hat{\boldsymbol{\Sigma}}_{\mathrm{UU}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathrm{UV}}\hat{\boldsymbol{\Sigma}}_{\mathrm{VV}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathrm{VU}}$, though it turns out that $\hat{\mathcal{K}}$ does not actually have to be calculated in order to use the following results. Assuming that $\hat{\mathcal{K}}$ and $\hat{\mathcal{A}}$ are nonsingular, [20.2.29] and [20.2.30] allow [20.2.22] to be written

$$\begin{aligned}
&\left|(1/T)\sum_{t=1}^{T}\left[(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)'\right]\right| \\
&= \left|(1/T)\sum_{t=1}^{T}\left[[(\hat{\mathcal{K}}')^{-1}\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t][(\hat{\mathcal{K}}')^{-1}\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t]'\right]\right| \\
&= \left|(\hat{\mathcal{K}}')^{-1}(1/T)\sum_{t=1}^{T}\left[[\hat{\boldsymbol{\eta}}_t - \hat{\mathcal{K}}'\boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t][\hat{\boldsymbol{\eta}}_t - \hat{\mathcal{K}}'\boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t]'\right](\hat{\mathcal{K}})^{-1}\right| \\
&= |(\hat{\mathcal{K}}')^{-1}|\left|(1/T)\sum_{t=1}^{T}\left[[\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t][\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t]'\right]\right||(\hat{\mathcal{K}})^{-1}| \\
&= \left|(1/T)\sum_{t=1}^{T}\left[[\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t][\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t]'\right]\right| \div |\hat{\mathcal{K}}|^2,
\end{aligned}$$

[20.2.32]

where

$$\hat{\boldsymbol{\Pi}} \equiv \hat{\mathcal{K}}'\boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}. \qquad [20.2.33]$$

Recall that maximizing the concentrated log likelihood function for the actual data [20.2.21] is equivalent to choosing $\boldsymbol{\zeta}_0$ so as to minimize the expression in [20.2.32] subject to the requirement that $\boldsymbol{\zeta}_0$ can be written as $\mathbf{BA}'$ for some $(n \times h)$ matrices $\mathbf{B}$ and $\mathbf{A}$. But $\boldsymbol{\zeta}_0$ can be written in this form if and only if $\hat{\boldsymbol{\Pi}}$ in [20.2.33] can be written in the form $\boldsymbol{\beta}\boldsymbol{\gamma}'$ for some $(n \times h)$ matrices $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Thus, the task can be described as choosing $\hat{\boldsymbol{\Pi}}$ so as to minimize [20.2.32] subject to this condition. But this is precisely the problem solved in [20.2.28]—the solution is to use as regressors the first h elements of $\hat{\boldsymbol{\xi}}_t$. The value of [20.2.32] at the optimum is given by

$$\prod_{i=1}^{h}(1 - \hat{r}_i^2) \div |\hat{\mathcal{K}}|^2. \qquad [20.2.34]$$

Moreover, the matrix $\hat{\mathcal{K}}$ satisfies

$$\mathbf{I}_n = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t' = (1/T)\sum_{t=1}^{T}\hat{\mathcal{K}}'\hat{\mathbf{u}}_t\hat{\mathbf{u}}_t'\hat{\mathcal{K}} = \hat{\mathcal{K}}'\hat{\boldsymbol{\Sigma}}_{\mathrm{UU}}\hat{\mathcal{K}}. \qquad [20.2.35]$$

Taking determinants of both sides of [20.2.35] establishes

$$1 = |\hat{\mathcal{K}}'|\,|\hat{\Sigma}_{UU}|\,|\hat{\mathcal{K}}|$$

or

$$1/|\hat{\mathcal{K}}|^2 = |\hat{\Sigma}_{UU}|.$$

Substituting this back into [20.2.34], it appears that the optimized value of [20.2.32] is equal to

$$|\hat{\Sigma}_{UU}| \times \prod_{i=1}^{h} (1 - \hat{r}_i^2).$$

Comparing [20.2.32] with [20.2.21], it follows that the maximum value achieved for the log likelihood function is given by

$$\mathcal{L}^* = \mathcal{N}(\hat{\zeta}_0) = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log\left\{|\hat{\Sigma}_{UU}| \times \prod_{i=1}^{h} (1 - \hat{r}_i^2)\right\},$$

as claimed in [20.2.10].

Motivation for Maximum Likelihood Estimates of Parameters

We have seen that the concentrated log likelihood function [20.2.21] is maximized by selecting as regressors the first h elements of $\hat{\xi}_t$. Since $\hat{\xi}_t = \hat{\mathcal{A}}'\hat{\mathbf{v}}_t$, this means using $\hat{\mathbf{A}}'\hat{\mathbf{v}}_t$ as regressors, where the $(n \times h)$ matrix $\hat{\mathbf{A}}$ denotes the first h columns of the $(n \times n)$ matrix $\hat{\mathcal{A}}$. Thus,

$$\zeta_0\hat{\mathbf{v}}_t = -\mathbf{B}\hat{\mathbf{A}}'\hat{\mathbf{v}}_t \qquad [20.2.36]$$

for some $(n \times h)$ matrix $\mathbf{B}$. This verifies the claim that $\hat{\mathbf{A}}$ is the maximum likelihood estimate of a basis for the space of cointegrating vectors.

Given that we want to choose $\hat{\mathbf{w}}_t \equiv \hat{\mathbf{A}}'\hat{\mathbf{v}}_t$ as regressors, the value of $\mathbf{B}$ for which the concentrated likelihood function will be maximized is obtained from *OLS* regressions of $\hat{\mathbf{u}}_t$ on $\hat{\mathbf{w}}_t$:

$$\hat{\mathbf{B}} = -\left[(1/T)\sum_{t=1}^{T} \hat{\mathbf{u}}_t\hat{\mathbf{w}}_t'\right]\left[(1/T)\sum_{t=1}^{T} \hat{\mathbf{w}}_t\hat{\mathbf{w}}_t'\right]^{-1}. \qquad [20.2.37]$$

But $\hat{\mathbf{w}}_t$ is composed of h canonical variates, meaning that

$$\left[(1/T)\sum_{t=1}^{T} \hat{\mathbf{w}}_t\hat{\mathbf{w}}_t'\right] = \mathbf{I}_h. \qquad [20.2.38]$$

Moreover,

$$\begin{aligned}\left[(1/T)\sum_{t=1}^{T} \hat{\mathbf{u}}_t\hat{\mathbf{w}}_t'\right] &= \left[(1/T)\sum_{t=1}^{T} \hat{\mathbf{u}}_t\hat{\mathbf{v}}_t'\hat{\mathbf{A}}\right] \\ &= \hat{\Sigma}_{UV}\hat{\mathbf{A}}.\end{aligned} \qquad [20.2.39]$$

Substituting [20.2.39] and [20.2.38] into [20.2.37],

$$\hat{\mathbf{B}} = -\hat{\Sigma}_{UV}\hat{\mathbf{A}},$$

and so, from [20.2.2], the maximum likelihood estimate of ζ_0 is given by

$$\hat{\zeta}_0 = \hat{\Sigma}_{UV}\hat{\mathbf{A}}\hat{\mathbf{A}}'$$

as claimed in [20.2.12].

Expressions [20.2.17] and [20.2.18] gave the values of α and ζ_i that maximized the likelihood function for any given value of ζ_0. Since the likelihood function is maximized with respect to ζ_0 by choosing $\hat{\zeta}_0$ according to [20.2.12], it is maximized with respect to α and ζ_i by substituting $\hat{\zeta}_0$ into [20.2.17] and [20.2.18], as claimed in [20.2.14] and [20.2.13]. Finally, substituting $\hat{\zeta}_0$ into [20.2.20] verifies [20.2.15].

Maximum Likelihood Estimation in the Absence of Deterministic Time Trends

The preceding analysis assumed that α, the $(n \times 1)$ vector of constant terms in the *VAR*, was unrestricted. The value of α contributes h constant terms for the h cointegrating relations, along with $g \equiv n - h$ deterministic time trends that are common to each of the n elements of $\mathbf{y}_t$. In some applications it might be of interest to allow constant terms in the cointegrating relations but to rule out deterministic time trends for any of the variables. We saw in equation [19.1.45] that this would require

$$\alpha = \mathbf{B}\mu_1^*, \qquad [20.2.40]$$

where $\mathbf{B}$ is the $(n \times h)$ matrix appearing in [20.2.2] while μ_1^* is an $(h \times 1)$ vector corresponding to the unconditional mean of $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$. Thus, for this restricted case, we want to estimate only the h elements of μ_1^* rather than all n elements of α.

To maximize the likelihood function subject to the restrictions that there are h cointegrating relations and no deterministic time trends in any of the series, Johansen's (1991) first step was to concentrate out $\zeta_1, \zeta_2, \ldots,$ and ζ_{p-1} (but not α). For given α and ζ_0, this is achieved by *OLS* regression of $(\Delta\mathbf{y}_t - \alpha - \zeta_0\mathbf{y}_{t-1})$ on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$. The residuals from this regression are related to the residuals from three separate regressions:

(1) A regression of $\Delta\mathbf{y}_t$ on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$ with no constant term,

$$\Delta\mathbf{y}_t = \bar{\Pi}_1\Delta\mathbf{y}_{t-1} + \bar{\Pi}_2\Delta\mathbf{y}_{t-2} + \cdots + \bar{\Pi}_{p-1}\Delta\mathbf{y}_{t-p+1} + \bar{\mathbf{u}}_t; \qquad [20.2.41]$$

(2) A regression of a constant term on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$,

$$1 = \bar{\omega}_1'\Delta\mathbf{y}_{t-1} + \bar{\omega}_2'\Delta\mathbf{y}_{t-2} + \cdots + \bar{\omega}_{p-1}'\Delta\mathbf{y}_{t-p+1} + \bar{w}_t; \qquad [20.2.42]$$

(3) A regression of $\mathbf{y}_{t-1}$ on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$ with no constant term,

$$\mathbf{y}_{t-1} = \bar{\aleph}_1\Delta\mathbf{y}_{t-1} + \bar{\aleph}_2\Delta\mathbf{y}_{t-2} + \cdots + \bar{\aleph}_{p-1}\Delta\mathbf{y}_{t-p+1} + \bar{\mathbf{v}}_t. \qquad [20.2.43]$$

The concentrated log likelihood function is then

$$\bar{\mathcal{M}}(\Omega, \alpha, \zeta_0) = -(Tn/2)\log(2\pi) - (T/2)\log|\Omega| - (1/2)\sum_{t=1}^{T}[(\bar{\mathbf{u}}_t - \alpha\bar{w}_t - \zeta_0\bar{\mathbf{v}}_t)'\Omega^{-1}(\bar{\mathbf{u}}_t - \alpha\bar{w}_t - \zeta_0\bar{\mathbf{v}}_t)].$$

Further concentrating out Ω results in

$$\begin{aligned}\bar{\mathcal{N}}(\alpha, \zeta_0) &= -(Tn/2)\log(2\pi) - (Tn/2) \\ &\quad - (T/2)\log\left|\sum_{t=1}^{T}(1/T)\left\{(\bar{\mathbf{u}}_t - \alpha\bar{w}_t - \zeta_0\bar{\mathbf{v}}_t)(\bar{\mathbf{u}}_t - \alpha\bar{w}_t - \zeta_0\bar{\mathbf{v}}_t)'\right\}\right|.\end{aligned} \qquad [20.2.44]$$

Imposing the constraints $\alpha = \mathbf{B}\mu_1^*$ and $\zeta_0 = -\mathbf{BA}'$, the magnitude in [20.2.44] can be written

$$\bar{\mathcal{N}}(\boldsymbol{\alpha}, \boldsymbol{\zeta}_0) = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log\left|\sum_{t=1}^{T}(1/T)\{(\hat{\mathbf{u}}_t + \mathbf{B}\bar{\mathbf{A}}'\bar{\mathbf{w}}_t)(\hat{\mathbf{u}}_t + \mathbf{B}\bar{\mathbf{A}}'\bar{\mathbf{w}}_t)'\}\right|, \qquad [20.2.45]$$

where

$$\underset{(n+1)\times 1}{\bar{\mathbf{w}}_t} \equiv \begin{bmatrix} \hat{w}_t \\ \hat{\mathbf{v}}_t \end{bmatrix}$$

$$\underset{h\times(n+1)}{\bar{\mathbf{A}}'} \equiv [-\boldsymbol{\mu}_1^* \quad \mathbf{A}']. \qquad [20.2.46]$$

But setting $\boldsymbol{\zeta}_0 = -\mathbf{BA}'$ in [20.2.21] produces an expression of exactly the same form as [20.2.45], with $\mathbf{A}$ in [20.2.21] replaced by $\bar{\mathbf{A}}$ and $\hat{\mathbf{v}}_t$ replaced by $\bar{\mathbf{w}}_t$. Thus, the restricted log likelihood is maximized simply by replacing $\hat{\mathbf{v}}_t$ in the analysis of [20.2.21] with $\bar{\mathbf{w}}_t$.

To summarize, construct

$$\bar{\Sigma}_{\mathrm{WW}} = (1/T)\sum_{t=1}^{T}\bar{\mathbf{w}}_t\bar{\mathbf{w}}_t'$$

$$\hat{\Sigma}_{\mathrm{UU}} = (1/T)\sum_{t=1}^{T}\hat{\mathbf{u}}_t\hat{\mathbf{u}}_t'$$

$$\bar{\Sigma}_{\mathrm{UW}} = (1/T)\sum_{t=1}^{T}\hat{\mathbf{u}}_t\bar{\mathbf{w}}_t'$$

and find the eigenvalues of the $(n + 1) \times (n + 1)$ matrix

$$\bar{\Sigma}_{\mathrm{WW}}^{-1}\bar{\Sigma}_{\mathrm{WU}}\hat{\Sigma}_{\mathrm{UU}}^{-1}\bar{\Sigma}_{\mathrm{UW}}, \qquad [20.2.47]$$

ordered $\bar{\lambda}_1 > \bar{\lambda}_2 > \cdots > \bar{\lambda}_{n+1}$. The maximum value achieved for the log likelihood function subject to the constraint that there are h cointegrating relations and no deterministic time trends is

$$\bar{\mathcal{L}}_h = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\Sigma}_{\mathrm{UU}}| - (T/2)\sum_{i=1}^{h}\log(1 - \bar{\lambda}_i). \qquad [20.2.48]$$

Let $\bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \ldots, \bar{\mathbf{a}}_{n+1}$ denote the eigenvectors of [20.2.47] normalized by $\bar{\mathbf{a}}_i'\bar{\Sigma}_{\mathrm{WW}}\bar{\mathbf{a}}_i = 1$. Then the maximum likelihood estimate of $\bar{\mathbf{A}}$ is given by the matrix $[\bar{\mathbf{a}}_1 \quad \bar{\mathbf{a}}_2 \quad \cdots \quad \bar{\mathbf{a}}_h]$. The maximum likelihood estimate of $\mathbf{B}\bar{\mathbf{A}}'$ is

$$\bar{\mathbf{B}}\bar{\mathbf{A}}' = -\bar{\Sigma}_{\mathrm{UW}}\bar{\mathbf{A}}\bar{\mathbf{A}}'. \qquad [20.2.49]$$

Recall from [20.2.46] that

$$\mathbf{B}\bar{\mathbf{A}}' = [-\mathbf{B}\boldsymbol{\mu}_1^* \quad \mathbf{BA}'] = [-\boldsymbol{\alpha} \quad -\boldsymbol{\zeta}_0]. \qquad [20.2.50]$$

Thus, [20.2.49] implies that the maximum likelihood estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}_0$ are given by

$$[\tilde{\boldsymbol{\alpha}} \quad \tilde{\boldsymbol{\zeta}}_0] = \bar{\Sigma}_{\mathrm{UW}}\bar{\mathbf{A}}\bar{\mathbf{A}}'.$$

The *MLE* of $\boldsymbol{\zeta}_i$ is

$$\tilde{\boldsymbol{\zeta}}_i = \hat{\boldsymbol{\Pi}}_i - \tilde{\boldsymbol{\alpha}}\hat{\boldsymbol{\omega}}_i' - \tilde{\boldsymbol{\zeta}}_0\hat{\boldsymbol{\aleph}}_i \qquad \text{for } i = 1, 2, \ldots, p-1,$$

while the MLE of $\mathbf{\Omega}$ is

$$\hat{\mathbf{\Omega}} = (1/T) \sum_{t=1}^{T} [(\tilde{\mathbf{u}}_t - \hat{\alpha}\tilde{w}_t - \tilde{\zeta}_0\tilde{\mathbf{v}}_t)(\tilde{\mathbf{u}}_t - \hat{\alpha}\tilde{w}_t - \tilde{\zeta}_0\tilde{\mathbf{v}}_t)'].$$

20.3. *Hypothesis Testing*

We saw in the previous chapter that tests of the null hypothesis of no cointegration typically involve nonstandard asymptotic distributions, while tests about the value of the cointegrating vector under the maintained hypothesis that cointegration is present will have asymptotic χ^2 distributions, provided that suitable allowance is made for the serial correlation in the data. These results generalize to $FIML$ analysis. The asymptotic distribution of a test of the number of cointegrating relations is nonstandard, but tests about the cointegrating vector are often χ^2.

Testing the Null Hypothesis of h *Cointegrating Relations*

Suppose that an $(n \times 1)$ vector $\mathbf{y}_t$ can be characterized by a $VAR(p)$ in levels, which we write in the form of [20.2.1]:

$$\Delta\mathbf{y}_t = \zeta_1\Delta\mathbf{y}_{t-1} + \zeta_2\Delta\mathbf{y}_{t-2} + \cdots + \zeta_{p-1}\Delta\mathbf{y}_{t-p+1} + \alpha + \zeta_0\mathbf{y}_{t-1} + \varepsilon_t. \qquad [20.3.1]$$

Under the null hypothesis H_0 that there are exactly h cointegrating relations among the elements of $\mathbf{y}_t$, this VAR is restricted by the requirement that ζ_0 can be written in the form $\zeta_0 = -\mathbf{BA}'$, for $\mathbf{B}$ an $(n \times h)$ matrix and $\mathbf{A}'$ an $(h \times n)$ matrix. Another way of describing this restriction is that only h linear combinations of the levels of $\mathbf{y}_{t-1}$ can be used in the regressions in [20.3.1]. The largest value that can be achieved for the log likelihood function under this constraint was given by [20.2.10]:

$$\begin{aligned}\mathcal{L}_0^* = &-(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\Sigma}_{UU}| \\ &- (T/2)\sum_{i=1}^{h}\log(1 - \hat{\lambda}_i).\end{aligned} \qquad [20.3.2]$$

Consider the alternative hypothesis H_A that there are n cointegrating relations, where n is the number of elements of $\mathbf{y}_t$. This amounts to the claim that *every* linear combination of $\mathbf{y}_t$ is stationary, in which case $\mathbf{y}_{t-1}$ would appear in [20.3.1] without constraints and no restrictions are imposed on ζ_0. The value for the log likelihood function in the absence of constraints is given by

$$\begin{aligned}\mathcal{L}_A^* = &-(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\Sigma}_{UU}| \\ &- (T/2)\sum_{i=1}^{n}\log(1 - \hat{\lambda}_i).\end{aligned} \qquad [20.3.3]$$

A likelihood ratio test of H_0 against H_A can be based on

$$\mathcal{L}_A^* - \mathcal{L}_0^* = -(T/2)\sum_{i=h+1}^{n}\log(1 - \hat{\lambda}_i).$$

If the hypothesis involved just $I(0)$ variables, we would expect twice the log likelihood ratio,

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T\sum_{i=h+1}^{n}\log(1 - \hat{\lambda}_i), \qquad [20.3.4]$$

to be asymptotically distributed as χ^2. In the case of H_0, however, the hypothesis involves the coefficient on $\mathbf{y}_{t-1}$, which, from the Stock-Watson common trends representation, depends on the value of $g \equiv (n - h)$ separate random walks. Let $\mathbf{W}(r)$ be g-dimensional standard Brownian motion. Suppose that the true value of the constant term $\boldsymbol{\alpha}$ in [20.3.1] is zero, meaning that there is no intercept in any of the cointegrating relations and no deterministic time trend in any of the elements of $\mathbf{y}_t$. Suppose further that no constant term is included in the auxiliary regressions [20.2.4] and [20.2.5] that were used to construct $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$. Johansen (1988) showed that under these conditions the asymptotic distribution of the statistic in [20.3.4] is the same as that of the trace of the following matrix:

$$\mathbf{Q} \equiv \left[\int_0^1 \mathbf{W}(r)\, d\mathbf{W}(r)'\right]'\left[\int_0^1 \mathbf{W}(r)\mathbf{W}(r)'\, dr\right]^{-1}\left[\int_0^1 \mathbf{W}(r)\, d\mathbf{W}(r)'\right]. \qquad [20.3.5]$$

Percentiles for the trace of the matrix in [20.3.5] are reported in the case 1 portion of Table B.10. These are based on Monte Carlo simulations.

If the number of cointegrating relations (h) is 1 less than the number of variables (n), then $g = 1$ and [20.3.5] describes the following scalar:

$$Q = \frac{\left\{\int_0^1 W(r)\, dW(r)\right\}^2}{\left\{\int_0^1 [W(r)]^2\, dr\right\}} = \frac{(1/2)^2\left\{[W(1)]^2 - 1\right\}^2}{\left\{\int_0^1 [W(r)]^2\, dr\right\}}, \qquad [20.3.6]$$

where the second equality follows from [18.1.15]. Expression [20.3.6] will be recognized as the square of the statistic [17.4.12] that described the asymptotic distribution of the Dickey-Fuller test based on the *OLS* t statistic. For example, if we are considering an autoregression involving a single variable ($n = 1$), the null hypothesis of no cointegrating relations ($h = 0$) amounts to the claim that $\zeta_0 = 0$ in [20.3.1] or that Δy_t follows an $AR(p - 1)$ process. Thus, Johansen's procedure provides an alternative approach to testing for unit roots in univariate series, an idea explored further in Exercise 20.4.

Another approach would be to test the null hypothesis of h cointegrating relations against the alternative of $h + 1$ cointegrating relations. Twice the log likelihood ratio for this case is given by

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T \log(1 - \hat{\lambda}_{h+1}). \qquad [20.3.7]$$

Again, under the assumption that the true value of $\boldsymbol{\alpha} = \mathbf{0}$ and that no constant term is included in [20.2.4] or [20.2.5], the asymptotic distribution of the statistic in [20.3.7] is the same as that of the largest eigenvalue of the matrix $\mathbf{Q}$ defined in [20.3.5]. Monte Carlo estimates of this distribution are reported in the case 1 section of Table B.11.

Note that if $g = 1$, then $n = h + 1$. In this case the statistics [20.3.4] and [20.3.7] are identical. For this reason, the first row in Table B.10 is the same as the first row of Table B.11.

Typically, the cointegrating relations could include nonzero intercepts, in which case we would want to include constants in the auxiliary regressions [20.2.4] and [20.2.5]. As one might guess from the analysis in Chapter 18, the asymptotic distribution in this case depends on whether or not any of the series exhibit deterministic time trends. Suppose that the true value of $\boldsymbol{\alpha}$ is such that there are no deterministic trends in any of the series, so that the true $\boldsymbol{\alpha}$ satisfies $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\mu}_1^*$ as in [20.2.40]. Assuming that no restrictions are imposed on the constant term in the estimation of the auxiliary regressions [20.2.4] and [20.2.5], then the asymptotic distribution of [20.3.4] is given in the case 2 section of Table B.10, while the asymptotic distribution of [20.3.7] is given in the case 2 panel of Table B.11. By contrast, if any of the variables exhibit deterministic time trends (one or more elements of $\boldsymbol{\alpha} - \mathbf{B}\boldsymbol{\mu}_1^*$ are nonzero), then the asymptotic distribution of [20.3.4] is that of the variable in the case 3 section of Table B.10, while the asymptotic distribution of [20.3.7] is given in the case 3 section of Table B.11.

When $g = 1$ and $\boldsymbol{\alpha} \neq \mathbf{B}\boldsymbol{\mu}_1^*$, the single random walk that is common to $\mathbf{y}_t$ is dominated by a deterministic time trend. In this situation, Johansen and Juselius (1990, p. 180) noted that the case 3 analog of [20.3.6] has a $\chi^2(1)$ distribution, for reasons similar to those noted by West (1988) and discussed in Chapter 18. The modest differences between the first row of the case 3 part of Table B.10 or B.11 and the first row of Table B.2 are presumably due to sampling error implicit in the Monte Carlo procedure used to generate the values in Tables B.10 and B.11.

Application to Exchange Rate Data

Consider for illustration the monthly data for Italy and the United States plotted in Figure 19.2. The systems of equations in [20.2.4] and [20.2.5] were estimated by *OLS* for $\mathbf{y}_t = (p_t, s_t, p_t^*)'$, where p_t is 100 times the log of the U.S. price level, s_t is 100 times the log of the dollar-lira exchange rate, and p_t^* is 100 times the log of the Italian price level. The regressions were estimated over $t =$ 1974:2 through 1989:10 (so that the number of observations used for estimation was $T = 189$); $p = 12$ lags were assumed for the *VAR* in levels.

The sample variance-covariance matrices for the residuals $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$ were calculated from [20.2.6] through [20.2.8] to be

$$\hat{\boldsymbol{\Sigma}}_{UU} = \begin{bmatrix} 0.0435114 & -0.0316283 & 0.0154297 \\ -0.0316283 & 4.68650 & 0.0319877 \\ 0.0154297 & 0.0319877 & 0.179927 \end{bmatrix}$$

$$\hat{\boldsymbol{\Sigma}}_{VV} = \begin{bmatrix} 427.366 & -370.699 & 805.812 \\ -370.699 & 424.083 & -709.036 \\ 805.812 & -709.036 & 1525.45 \end{bmatrix}$$

$$\hat{\boldsymbol{\Sigma}}_{UV} = \begin{bmatrix} -0.484857 & 0.498758 & -0.837701 \\ -1.81401 & -2.95927 & -2.46896 \\ -1.80836 & 1.46897 & -3.58991 \end{bmatrix}.$$

The eigenvalues of the matrix in [20.2.9] are then[4]

$$\hat{\lambda}_1 = 0.1105$$
$$\hat{\lambda}_2 = 0.05603$$
$$\hat{\lambda}_3 = 0.03039$$

with

$$T\log(1 - \hat{\lambda}_1) = -22.12$$
$$T\log(1 - \hat{\lambda}_2) = -10.90$$
$$T\log(1 - \hat{\lambda}_3) = -5.83.$$

[4]Calculations were based on more significant digits than reported, and so the reader may find slight discrepancies in trying to reproduce these results from the figures reported.

The likelihood ratio test of the null hypothesis of $h = 0$ cointegrating relations against the alternative of $h = 3$ cointegrating relations is then calculated from [20.3.4] to be

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = 22.12 + 10.90 + 5.83 = 38.85. \qquad [20.3.8]$$

Here the number of unit roots under the null hypothesis is $g = n - h = 3$. Given the evidence of deterministic time trends, the magnitude in [20.3.8] is to be compared with the case 3 section of Table B.10. Since $38.85 > 29.5$, the null hypothesis of no cointegration is rejected at the 5% level. Similarly, the likelihood ratio test [20.3.7] of the null hypothesis of no cointegrating relations ($h = 0$) against the alternative of a single cointegrating relation ($h = 1$) is given by 22.12. Comparing this with the case 3 section of Table B.11, we see that $22.12 > 20.8$, so that the null hypothesis of no cointegration is also rejected by this test.

This differs from the conclusion of the Phillips-Ouliaris test for no cointegration between these series, on the basis of which the null hypothesis of no cointegration for these variables was found to be accepted in Chapter 19.

Searching for evidence of a possible second cointegrating relation, consider the likelihood ratio test of the null hypothesis of $h = 1$ cointegrating relation against the alternative of $h = 3$ cointegrating relations:

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = 10.90 + 5.83 = 16.73.$$

For this test, $g = 2$. Since $16.73 > 15.2$, the null hypothesis of a single cointegrating relation is rejected at the 5% level. The likelihood ratio test of the null hypothesis of $h = 1$ cointegrating relation against the alternative of $h = 2$ relations is $10.90 < 14.0$; hence, the two tests offer conflicting evidence as to the presence of a second cointegrating relation.

The eigenvector $\hat{\mathbf{a}}_1$ of the matrix in [20.2.9] associated with $\hat{\lambda}_1$, normalized so that $\hat{\mathbf{a}}_1'\hat{\Sigma}_{VV}\hat{\mathbf{a}}_1 = 1$, is given by

$$\hat{\mathbf{a}}_1' = [-0.7579 \quad 0.02801 \quad 0.4220]. \qquad [20.3.9]$$

It is natural to renormalize this by taking the first element to be unity:

$$\hat{\mathbf{a}}_1' = [1.00 \quad -0.04 \quad -0.56].$$

This is virtually identical to the estimate of the cointegrating vector based on *OLS* from [19.2.49].

Likelihood Ratio Tests About the Cointegrating Vector

Consider a system of n variables that is assumed (under both the null and the alternative) to be characterized by h cointegrating relations. We might then want to test a restriction on these cointegrating vectors, such as that only q of the variables are involved in the cointegrating relations. For example, we might be interested in whether the middle coefficient in [20.3.9] is zero, that is, in whether the cointegrating relation involves solely the U.S. and Italian price levels. For this example $h = 1$, $q = 2$, and $n = 3$. In general it must be the case that $h \leq q \leq n$. Since h linear combinations of the q variables included in the cointegrating relations are stationary, if $q = h$, then all q of the included variables would have to be stationary in levels. If $q = n$, then the null hypothesis places no restrictions on the cointegrating relations.

Consider the general restriction that there is a known $(q \times n)$ matrix $\mathbf{D}'$ such that the cointegrating relations involve only $\mathbf{D}'\mathbf{y}_t$. For the preceding example,

$$\mathbf{D}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad [20.3.10]$$

Hence, the error-correction term in [20.3.1] will take the form

$$\zeta_0 \mathbf{y}_{t-1} = -\mathbf{B}\mathbf{A}'\mathbf{D}'\mathbf{y}_{t-1},$$

where $\mathbf{B}$ is now an $(n \times h)$ matrix and $\mathbf{A}'$ is an $(h \times q)$ matrix. Maximum likelihood estimation proceeds exactly as in the previous section, where $\hat{\mathbf{v}}_t$ in [20.2.5] is replaced by the *OLS* residuals from regressions of $\mathbf{D}'\mathbf{y}_{t-1}$ on a constant and $\Delta\mathbf{y}_{t-1}$, $\Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$. This is equivalent to replacing $\hat{\Sigma}_{VV}$ in [20.2.6] and $\hat{\Sigma}_{UV}$ in [20.2.8] with

$$\tilde{\Sigma}_{VV} \equiv \mathbf{D}'\hat{\Sigma}_{VV}\mathbf{D} \qquad [20.3.11]$$

$$\tilde{\Sigma}_{UV} \equiv \hat{\Sigma}_{UV}\mathbf{D}. \qquad [20.3.12]$$

Let $\tilde{\lambda}_i$ denote the ith largest eigenvalue of

$$\tilde{\Sigma}_{VV}^{-1}\tilde{\Sigma}_{VU}\hat{\Sigma}_{UU}^{-1}\tilde{\Sigma}_{UV}. \qquad [20.3.13]$$

The maximized value for the restricted log likelihood is then

$$\mathscr{L}_0^* = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\Sigma}_{UU}| - (T/2)\sum_{i=1}^{h}\log(1 - \tilde{\lambda}_i).$$

A likelihood ratio test of the null hypothesis that the h cointegrating relations only involve $\mathbf{D}'\mathbf{y}_t$ against the alternative hypothesis that the h cointegrating relations could involve any elements of $\mathbf{y}_t$ would then be

$$2(\mathscr{L}_A^* - \mathscr{L}_0^*) = -T\sum_{i=1}^{h}\log(1 - \hat{\lambda}_i) + T\sum_{i=1}^{h}\log(1 - \tilde{\lambda}_i). \qquad [20.3.14]$$

In this case, the null hypothesis involves only coefficients on $I(0)$ variables (the error-correction terms $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$), and standard asymptotic distribution theory turns out to apply. Johansen (1988, 1991) showed that the likelihood ratio statistic [20.3.14] has an asymptotic χ^2 distribution with $h\cdot(n - q)$ degrees of freedom.

For illustration, consider the restriction represented by [20.3.10] that the exchange rate has a coefficient of zero in the cointegrating vector [20.3.9]. From [20.3.11] and [20.3.12], we calculate

$$\tilde{\Sigma}_{VV} = \begin{bmatrix} 427.366 & 805.812 \\ 805.812 & 1525.45 \end{bmatrix}$$

$$\tilde{\Sigma}_{UV} = \begin{bmatrix} -0.484857 & -0.837701 \\ -1.81401 & -2.46896 \\ -1.80836 & -3.58991 \end{bmatrix}.$$

The eigenvalues for the matrix in [20.3.13] are then

$$\tilde{\lambda}_1 = 0.1059 \qquad \tilde{\lambda}_2 = 0.04681,$$

with

$$T\log(1 - \tilde{\lambda}_1) = -21.15 \qquad T\log(1 - \tilde{\lambda}_2) = -9.06.$$

The likelihood ratio statistic [20.3.14] is

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = 22.12 - 21.15$$
$$= 0.97.$$

The degrees of freedom for this statistic are

$$h \cdot (n - q) = 1 \cdot (3 - 2) = 1;$$

the null hypothesis imposes a single restriction on the cointegrating vector. The 5% critical value for a $\chi^2(1)$ variable is seen from Table B.2 to be 3.84. Since $0.97 < 3.84$, the null hypothesis that the exchange rate does not appear in the cointegrating relation is accepted. The restricted cointegrating vector (normalized with the coefficient on the U.S. price level to be unity) is

$$\tilde{\mathbf{a}}_1' = [1.00 \quad 0.00 \quad -0.54].$$

As a second example, consider the hypothesis that originally suggested interest in a possible cointegrating relation between these three variables. This was the hypothesis that the real exchange rate is stationary, or that the cointegrating vector is proportional to $(1, -1, -1)'$. For this hypothesis, $\mathbf{D}' = (1, -1, -1)$ and

$$\hat{\Sigma}_{VV} = 88.5977$$

$$\hat{\Sigma}_{UV} = \begin{bmatrix} -0.145914 \\ 3.61422 \\ 0.312582 \end{bmatrix}.$$

In this case, the matrix [20.3.13] is the scalar 0.0424498, and so $\tilde{\lambda}_1 = 0.0424498$ and $T \log(1 - \tilde{\lambda}_1) = -8.20$. Thus, the likelihood ratio test of the null hypothesis that the cointegrating vector is proportional to $(1, -1, -1)'$ is

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = 22.12 - 8.20$$
$$= 13.92.$$

In this case, the degrees of freedom are

$$h \cdot (n - q) = 1 \cdot (3 - 1) = 2.$$

The 5% critical value for a $\chi^2(2)$ variable is 5.99. Since $13.92 > 5.99$, the null hypothesis that the cointegrating vector is proportional to $(1, -1, -1)'$ is rejected.

Other Hypothesis Tests

A number of other hypotheses can be tested in this framework. For example, Johansen (1991) showed that the null hypothesis that there are no deterministic time trends in any of the series can be tested by taking twice the difference between [20.2.10] and [20.2.48]. Under the null hypothesis, this likelihood ratio statistic is asymptotically χ^2 with $g = n - h$ degrees of freedom. Johansen also discussed construction of Wald-type tests of hypotheses involving the cointegrating vectors.

Not all hypothesis tests about the coefficients in Johansen's framework are asymptotically χ^2. Consider an error-correction *VAR* of the form of [20.2.1] where $\zeta_0 = -\mathbf{BA}'$. Suppose we are interested in the null hypothesis that the last n_3 elements of $\mathbf{y}_t$ fail to Granger-cause the first n_1 elements of $\mathbf{y}_t$. Toda and Phillips (forthcoming) showed that a Wald test of this null hypothesis can have a nonstandard distribution. See Mosconi and Giannini (1992) for further discussion.

Comparison Between FIML *and Other Approaches*

Johansen's *FIML* estimation represents the short-run dynamics of a system in terms of a vector autoregression in differences with the error-correction vector $\mathbf{z}_{t-1}$ added. Short-run dynamics can also be modeled with what are sometimes called nonparametric methods, such as the Bartlett window used to construct the fully modified Phillips-Hansen (1990) estimator in equation [19.3.53]. Related nonparametric estimators have been proposed by Phillips (1990, 1991a), Park (1992), and Park and Ogaki (1991). Park (1990) established the asymptotic equivalence of the parametric and nonparametric approaches, and Phillips (1991a) discussed the sense in which any *FIML* estimator is asymptotically efficient. Johansen (1992) provided a further discussion of the relation between limited-information and full-information estimation strategies.

In practice, the parametric and nonparametric approaches differ not just in their treatment of short-run dynamics but also in the normalizations employed. The fact that Johansen's method seeks to estimate the *space* of cointegrating relations rather than a particular set of coefficients can be both an asset and a liability. It is an asset if the researcher has no prior information about which variables appear in the cointegrating relations and is concerned about inadvertently normalizing $a_{11} = 1$ when the true value of $a_{11} = 0$. On the other hand, Phillips (1991b) has stressed that if the researcher wants to make structural interpretations of the separate cointegrating relations, this logically requires imposing further restrictions on the matrix $\mathbf{A}'$.

For example, let r_t denote the nominal interest rate on 3-month corporate debt, i_t the nominal interest rate on 3-month government debt, and π_t the 3-month inflation rate. Suppose that these three variables appear to be $I(1)$ and exhibit two cointegrating relations. A natural view is that these cointegrating relations represent two stabilizing relations. The first reflects forces that keep the risk premium stationary, so that

$$r_t = \mu_{11}^* + \gamma_1 i_t + z_{1t}^*, \qquad [20.3.15]$$

with $z_{1t}^* \sim I(0)$. A second force is the Fisher effect, which tends to keep the real interest rate stationary:

$$\pi_t = \mu_{21}^* + \gamma_2 i_t + z_{2t}^*, \qquad [20.3.16]$$

with $z_{2t}^* \sim I(0)$. The system of [20.3.15] and [20.3.16] will be recognized as an example of Phillips's (1991a) triangular representation [19.1.20] for the vector $\mathbf{y}_t = (r_t, \pi_t, i_t)'$. Thus, in this example theoretical considerations suggest a natural ordering of variables for which the normalization used by Phillips would be of particular interest for structural inference—the coefficients μ_{11}^* and γ_1 tell us about the risk premium, and the coefficients μ_{21}^* and γ_2 tell us about the Fisher effect.

20.4. *Overview of Unit Roots—To Difference or Not to Difference?*

The preceding chapters have explored a number of issues in the statistical analysis of unit roots. This section attempts to summarize what all this means in practice.

Consider a vector of variables $\mathbf{y}_t$ whose dynamics we would like to describe and some of whose elements may be nonstationary. For concreteness, let us assume that the goal is to characterize these dynamics in terms of a vector autoregression.

One option is to ignore the nonstationarity altogether and simply estimate the *VAR* in levels, relying on standard t and F distributions for testing any hypotheses. This strategy has the following features to recommend it. (1) The parameters that describe the system's dynamics are estimated consistently. (2) Even if the true model is a *VAR* in differences, certain functions of the parameters and hypothesis tests based on a *VAR* in levels have the same asymptotic distribution as would estimates based on differenced data. (3) A Bayesian motivation can be given for the usual t or F distributions for test statistics even when the classical asymptotic theory for these statistics is nonstandard.

A second option is routinely to difference any apparently nonstationary variables before estimating the *VAR*. If the true process is a *VAR* in differences, then differencing should improve the small-sample performance of all of the estimates and eliminate altogether the nonstandad asymptotic distributions associated with certain hypothesis tests. The drawback to this approach is that the true process may not be a *VAR* in differences. Some of the series may in fact have been stationary, or perhaps some linear combinations of the series are stationary, as in a cointegrated *VAR*. In such circumstances a *VAR* in differenced form is misspecified.

Yet a third approach is to investigate carefully the nature of the nonstationarity, testing each series individually for unit roots and then testing for possible cointegration among the series. Once the nature of the nonstationarity is understood, a stationary representation for the system can be estimated. For example, suppose that in a four-variable system we determine that the first variable y_{1t} is stationary while the other variables (y_{2t}, y_{3t}, and y_{4t}) are each individually $I(1)$. Suppose we further conclude that y_{2t}, y_{3t}, and y_{4t} are characterized by a single cointegrating relation. For $\mathbf{y}_{2t} \equiv (y_{2t}, y_{3t}, y_{4t})'$, this implies a vector error-correction representation of the form

$$\begin{bmatrix} y_{1t} \\ \Delta\mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} + \begin{bmatrix} \zeta_{11}^{(1)} & \zeta_{12}^{(1)} \\ \zeta_{21}^{(1)} & \zeta_{22}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ \Delta\mathbf{y}_{2,t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{11}^{(2)} & \zeta_{12}^{(2)} \\ \zeta_{21}^{(2)} & \zeta_{22}^{(2)} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ \Delta\mathbf{y}_{2,t-2} \end{bmatrix} + \cdots$$

$$+ \begin{bmatrix} \zeta_{11}^{(p-1)} & \zeta_{12}^{(p-1)} \\ \zeta_{21}^{(p-1)} & \zeta_{22}^{(p-1)} \end{bmatrix} \begin{bmatrix} y_{1,t-p+1} \\ \Delta\mathbf{y}_{2,t-p+1} \end{bmatrix} + \begin{bmatrix} \zeta_{1}^{(0)} \\ \zeta_{2}^{(0)} \end{bmatrix} \mathbf{y}_{2,t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix},$$

where the (4×3) matrix $\begin{bmatrix} \zeta_{1}^{(0)} \\ \zeta_{2}^{(0)} \end{bmatrix}$ is restricted to be of the form $\mathbf{ba}'$ where $\mathbf{b}$ is (4×1) and $\mathbf{a}'$ is (1×3). Such a system can then be estimated by adapting the methods described in Section 20.2, and most hypothesis tests on this system should be asymptotically χ^2.

The disadvantage of the third approach is that, despite the care one exercises, the restrictions imposed may still be invalid—the investigator may have accepted a null hypothesis even though it is false, or rejected a null hypothesis that is actually true. Moreover, alternative tests for unit roots and cointegration can produce conflicting results, and the investigator may be unsure as to which should be followed.

Experts differ in the advice offered for applied work. One practical solution is to employ parts of all three approaches. This eclectic strategy would begin by estimating the *VAR* in levels without restrictions. The next step is to make a quick assessment as to which series are likely nonstationary. This assessment could be based on graphs of the data, prior information about the series and their likely cointegrating relations, or any of the more formal tests discussed in Chapter 17. Any nonstationary series can then be differenced or expressed in error-correction form and a stationary *VAR* could then be estimated. For example, to estimate a *VAR* that includes the log of income (y_t) and the log of consumption (c_t), these two variables might be included in a stationary *VAR* as Δy_t and $(c_t - y_t)$. If the *VAR* for the data in levels form yields similar inferences to those for the *VAR* in stationary form, then the researcher might be satisfied that the results were not governed by the assumptions made about unit roots. If the answers differ, then some attempt to reconcile the results should be made. Careful efforts along the lines of the third strategy described in this section might convince the investigator that the stationary formulation was misspecified, or alternatively that the levels results can be explained by the appropriate asymptotic theory. A nice example of how asymptotic theory could be used to reconcile conflicting findings was provided by Stock and Watson (1989). Alternatively, Christiano and Ljungqvist (1988) proposed simulating data from the estimated levels model, and seeing whether incorrectly fitting such simulated data with the stationary specification would spuriously produce the results found when the stationary specification was fitted to the actual data. Similarly, data could be simulated from the stationary model to see if it could account for the finding of the levels specification. If we find that a single specification can account for both the levels and the stationary results, then our confidence in that specification increases.

APPENDIX 20.A. *Proof of Chapter 20 Proposition*

■ **Proof of Proposition 20.1.**

(a) First we show that $\lambda_i < 1$ for $i = 1, 2, \ldots, n_1$. Any eigenvalue λ of [20.1.8] satisfies

$$|\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} - \lambda \mathbf{I}_{n_1}| = 0.$$

Since Σ_{YY} is positive definite, this will be true if and only if

$$|\lambda\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}| = 0. \qquad [20.A.1]$$

But from the triangular factorization of Σ in equation [4.5.26], the matrix

$$\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \qquad [20.A.2]$$

is positive definite. Hence, the determinant in [20.A.1] could not be zero at $\lambda = 1$. Note further that

$$\lambda\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} = (\lambda - 1)\Sigma_{YY} + [\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}]. \qquad [20.A.3]$$

If $\lambda > 1$, then the right side of expression [20.A.3] would be the sum of two positive definite matrices and so would be positive definite. The left side of [20.A.3] would then be positive definite, implying that the determinant in [20.A.1] could not be zero for $\lambda > 1$. Hence, $\lambda \geq 1$ is not consistent with [20.A.1].

To see that $\lambda_i \geq 0$, notice that if λ were less than zero, then $\lambda\Sigma_{YY}$ would be a negative number times a positive definite matrix so that $\lambda\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$ would also be a negative number times a positive definite matrix. Hence, the determinant in [20.A.1] could not be zero for any value of $\lambda < 0$.

Parallel arguments establish that $0 \leq \mu_j < 1$ for $j = 1, 2, \ldots, n_2$.

(b) Let $\mathbf{k}_i$ be an eigenvector associated with a nonzero eigenvalue λ_i of [20.1.8]:

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_i = \lambda_i\mathbf{k}_i. \qquad [20.A.4]$$

Premultiplying both sides of [20.A.4] by Σ_{XY} results in

$$[\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}][\Sigma_{XY}\mathbf{k}_i] = \lambda_i[\Sigma_{XY}\mathbf{k}_i]. \qquad [20.A.5]$$

But $[\Sigma_{XY}\mathbf{k}_i]$ cannot be zero, for if $[\Sigma_{XY}\mathbf{k}_i]$ did equal zero, then the left side of [20.A.4] would be zero, implying that $\lambda_i = 0$. Thus, [20.A.5] implies that λ_i is also an eigenvalue of the matrix $[\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}]$ associated with the eigenvector $[\Sigma_{XY}\mathbf{k}_i]$. Recall further that eigenvalues are unchanged by transposition of a matrix:

$$[\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}]' = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX},$$

which is the matrix [20.1.12]. This proves that if λ_i is a nonzero eigenvalue of [20.1.8], then it is also an eigenvalue of [20.1.12]. Exactly parallel calculations show that if μ_i is a nonzero eigenvalue of [20.1.12], then it is also an eigenvalue of [20.1.8].

(c) Premultiply [20.1.10] by $\mathbf{k}_j'\Sigma_{YY}$:

$$\mathbf{k}_j'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_i = \lambda_i\mathbf{k}_j'\Sigma_{YY}\mathbf{k}_i. \qquad [20.A.6]$$

Similarly, replace i with j in [20.1.10]:

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_j\mathbf{k}_j, \qquad [20.A.7]$$

and premultiply by $\mathbf{k}_i'\Sigma_{YY}$;

$$\mathbf{k}_i'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_j\mathbf{k}_i'\Sigma_{YY}\mathbf{k}_j. \qquad [20.A.8]$$

Subtracting [20.A.8] from [20.A.6], we see that

$$0 = (\lambda_i - \lambda_j)\mathbf{k}_j'\Sigma_{YY}\mathbf{k}_i. \qquad [20.A.9]$$

If $i \neq j$, then $\lambda_i \neq \lambda_j$ and [20.A.9] establishes that $\mathbf{k}_j'\Sigma_{YY}\mathbf{k}_i = 0$ for $i \neq j$. For $i = j$, we normalized $\mathbf{k}_i'\Sigma_{YY}\mathbf{k}_i = 1$ in [20.1.11]. Thus we have established condition [20.1.3] for the case of distinct eigenvalues.

Virtually identical calculations show that [20.1.13] and [20.1.14] imply [20.1.4].

(d) Transpose [20.1.13] and postmultiply by $\Sigma_{XY}\mathbf{k}_j$:

$$\mathbf{a}_i'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_i\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j. \qquad [20.A.10]$$

Similarly, premultiply [20.A.7] by $\mathbf{a}_i'\Sigma_{XY}$:

$$\mathbf{a}_i'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_j\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j. \qquad [20.A.11]$$

Subtracting [20.A.11] from [20.A.10] results in

$$0 = (\lambda_i - \lambda_j)\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j.$$

This shows that $\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j = 0$ for $\lambda_i \neq \lambda_j$, as required by [20.1.5].

To find the value of $\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i$ for $i = j$, premultiply [20.1.13] by $\mathbf{a}_i'\Sigma_{XX}$, making use of [20.1.14]:

$$\mathbf{a}_i'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a}_i = \lambda_i. \qquad [20.A.12]$$

Let us suppose for illustration that n_1 is the smaller of n_1 and n_2; that is, $n = n_1$.[5] Then the matrix of eigenvectors $\mathcal{K}$ is $(n \times n)$ and nonsingular. In this case, [20.1.3] implies that

$$\Sigma_{YY} = [\mathcal{K}']^{-1}\mathcal{K}^{-1},$$

or, taking inverses,

$$\Sigma_{YY}^{-1} = \mathcal{K}\mathcal{K}'. \qquad [20.A.13]$$

Substituting [20.A.13] into [20.A.12], we find that

$$\mathbf{a}_i'\Sigma_{XY}\mathcal{K}\mathcal{K}'\Sigma_{YX}\mathbf{a}_i = \lambda_i. \qquad [20.A.14]$$

Now,

$$\begin{aligned}\mathbf{a}_i'\Sigma_{XY}\mathcal{K} &= \mathbf{a}_i'\Sigma_{XY}[\mathbf{k}_1 \quad \mathbf{k}_2 \quad \cdots \quad \mathbf{k}_n] \\ &= [\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_1 \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_2 \quad \cdots \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i \quad \cdots \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_n] \qquad [20.A.15] \\ &= [0 \quad 0 \quad \cdots \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i \quad \cdots \quad 0].\end{aligned}$$

Substituting [20.A.15] into [20.A.14], it follows that

$$(\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i)^2 = \lambda_i.$$

Thus, the ith canonical correlation,

$$r_i \equiv \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i,$$

is given by the square root of the eigenvalue λ_i, as claimed:

$$r_i^2 = \lambda_i. \quad \blacksquare$$

[5]In the converse case when $n = n_2$, a parallel argument can be constructed using the fact that

$$\mathbf{k}_i'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_i = \lambda_i.$$

Chapter 20 Exercises

20.1. In this problem you are asked to verify the claim in the text that the first canonical variates η_{1t} and ξ_{1t} represent the linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$ with maximum possible correlation. Consider the following maximization problem:

$$\max_{\{\mathbf{k}_1, \mathbf{a}_1\}} E(\mathbf{k}_1'\mathbf{y}_t\mathbf{x}_t'\mathbf{a}_1)$$

subject to

$$E(\mathbf{k}_1'\mathbf{y}_t\mathbf{y}_t'\mathbf{k}_1) = 1$$

$$E(\mathbf{a}_1'\mathbf{x}_t\mathbf{x}_t'\mathbf{a}_1) = 1.$$

Show that the maximum value achieved for this problem is given by the square root of the largest eigenvalue of the matrix $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$, and that $\mathbf{a}_1$ is the associated eigenvector normalized as stated. Show that $\mathbf{k}_1$ is the normalized eigenvector of $\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$, associated with this same eigenvalue.

20.2. It was claimed in the text that the maximized log likelihood function under the null hypothesis of h cointegrating relations was given by [20.3.2]. What is the nature of the restriction on the VAR in [20.3.1] when $h = 0$? Show that the value of [20.3.2] for this case is the same as the log likelihood for a $VAR(p - 1)$ process fitted to the differenced data $\Delta\mathbf{y}_t$.

20.3. It was claimed in the text that the maximized log likelihood function under the alternative hypothesis of n cointegrating relations was given by [20.3.3]. This case involves regressing $\Delta\mathbf{y}_t$ on a constant, $\mathbf{y}_{t-1}$, and $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$ without restrictions. Let $\hat{\mathbf{g}}_t$ denote the residuals from this unrestricted regression, with $\hat{\Sigma}_{GG} = (1/T)\Sigma_{t=1}^T \hat{\mathbf{g}}_t\hat{\mathbf{g}}_t'$. Equation [11.1.32] would then assert that the maximized log likelihood function should be given by

$$\mathcal{L}_A^* = -(Tn/2)\log(2\pi) - (T/2)\log|\hat{\Sigma}_{GG}| - (Tn/2).$$

Show that this number is the same as that given by formula [20.3.3].

20.4. Consider applying Johansen's likelihood ratio test to univariate data ($n = 1$). Show that the test of the null hypothesis that y_t is nonstationary ($h = 0$) against the alternative that y_t is stationary ($h = 1$) can be written

$$T[\log(\hat{\sigma}_0^2) - \log(\hat{\sigma}_1^2)],$$

where $\hat{\sigma}_0^2$ is the average squared residual from a regression of Δy_t on a constant and Δy_{t-1}, $\Delta y_{t-2}, \ldots, \Delta y_{t-p+1}$ while $\hat{\sigma}_1^2$ is the average squared residual when y_{t-1} is added as an explanatory variable to this regression.

Chapter 20 References

Ahn, S. K., and G. C. Reinsel. 1990. "Estimation for Partially Nonstationary Multivariate Autoregressive Models." *Journal of the American Statistical Association* 85:813–23.

Christiano, Lawrence J., and Lars Ljungqvist. 1988. "Money Does Granger-Cause Output in the Bivariate Money-Output Relation." *Journal of Monetary Economics* 22:217–35.

Johansen, Søren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12:231–54.

———. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59:1551–80.

———. 1992. "Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis." *Journal of Econometrics* 52:389–402.

——— and Katarina Juselius. 1990. "Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money." *Oxford Bulletin of Economics and Statistics* 52:169–210.

Koopmans, Tjalling C., and William C. Hood. 1953. "The Estimation of Simultaneous Linear Economic Relationships," in William C. Hood and Tjalling C. Koopmans, eds., *Studies in Econometric Method*. New York: Wiley.

Mosconi, Rocco, and Carlo Giannini. 1992. "Non-Causality in Cointegrated Systems: Representation, Estimation and Testing," *Oxford Bulletin of Economics and Statistics*. 54:399–417.

Park, Joon Y. 1990. "Maximum Likelihood Estimation of Simultaneous Cointegrated Models." University of Aarhus. Mimeo.

———. 1992. "Canonical Cointegrating Regressions." *Econometrica* 60:119–43.

——— and Masao Ogaki. 1991. "Inference in Cointegrated Models Using VAR Prewhitening to Estimate Shortrun Dynamics." University of Rochester. Mimeo.

Phillips, Peter C. B. 1990. "Spectral Regression for Cointegrated Time Series," in William Barnett, James Powell, and George Tauchen, eds., *Nonparametric and Semiparametric Methods in Economics and Statistics*. New York: Cambridge University Press.

———. 1991a. "Optimal Inference in Cointegrated Systems." *Econometrica* 59:283–306.

———. 1991b. "Unidentified Components in Reduced Rank Regression Estimation of ECM's." Yale University. Mimeo.

——— and Bruce E. Hansen. 1990. "Statistical Inference in Instrumental Variables Regression with I(1) Processes." *Review of Economic Studies* 57:99–125.

——— and S. Ouliaris. 1990. "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica* 58:165–93.

Stock, James H., and Mark W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* 83:1097–1107.

——— and ———. 1989. "Interpreting the Evidence on Money-Income Causality." *Journal of Econometrics* 40:161–81.

Toda, H. Y., and Peter C. B. Phillips. Forthcoming. "Vector Autoregression and Causality." *Econometrica*.

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

21 Time Series Models of Heteroskedasticity

21.1. *Autoregressive Conditional Heteroskedasticity* (ARCH)

An autoregressive process of order p (denoted $AR(p)$) for an observed variable y_t takes the form

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t, \qquad [21.1.1]$$

where u_t is white noise:

$$E(u_t) = 0 \qquad [21.1.2]$$

$$E(u_t u_\tau) = \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases} \qquad [21.1.3]$$

The process is covariance-stationary provided that the roots of

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

are outside the unit circle. The optimal linear forecast of the level of y_t for an $AR(p)$ process is

$$\hat{E}(y_t | y_{t-1}, y_{t-2}, \ldots) = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}, \qquad [21.1.4]$$

where $\hat{E}(y_t | y_{t-1}, y_{t-2}, \ldots)$ denotes the linear projection of y_t on a constant and $(y_{t-1}, y_{t-2}, \ldots)$. While the conditional mean of y_t changes over time according to [21.1.4], provided that the process is covariance-stationary, the unconditional mean of y_t is constant:

$$E(y_t) = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p).$$

Sometimes we might be interested in forecasting not only the level of the series y_t but also its variance. For example, Figure 21.1 plots the federal funds rate, which is an interest rate charged on overnight loans from one bank to another. This interest rate has been much more volatile at some times than at others. Changes in the variance are quite important for understanding financial markets, since investors require higher expected returns as compensation for holding riskier assets. A variance that changes over time also has implications for the validity and efficiency of statistical inference about the parameters $(c, \phi_1, \phi_2, \ldots, \phi_p)$ that describe the dynamics of the level of y_t.

Although [21.1.3] implies that the unconditional variance of u_t is the constant σ^2, the conditional variance of u_t could change over time. One approach is to

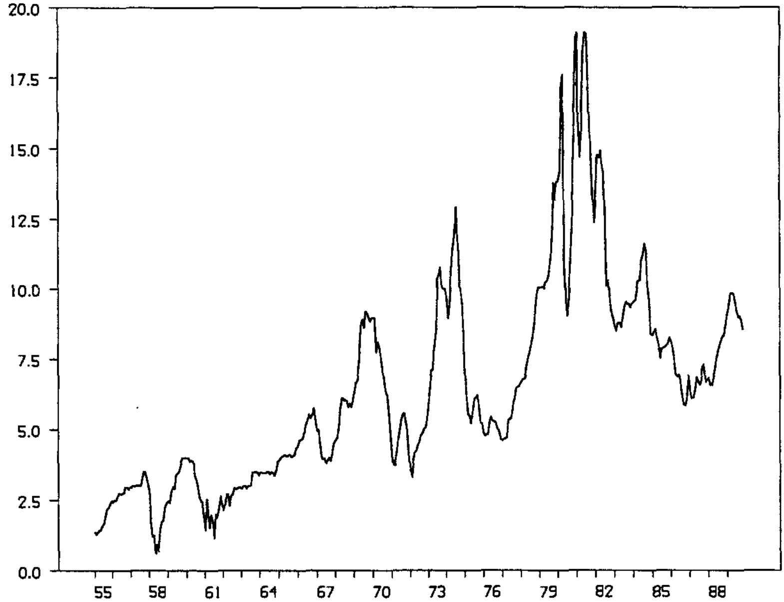


FIGURE 21.1 U.S. federal funds rate (monthly averages quoted at an annual rate), 1955–89.

describe the square of u_t as itself following an $AR(m)$ process:

$$u_t^2 = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 + w_t, \qquad [21.1.5]$$

where w_t is a new white noise process:

$$E(w_t) = 0$$

$$E(w_t w_\tau) = \begin{cases} \lambda^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Since u_t is the error in forecasting y_t, expression [21.1.5] implies that the linear projection of the squared error of a forecast of y_t on the previous m squared forecast errors is given by

$$\hat{E}(u_t^2 | u_{t-1}^2, u_{t-2}^2, \ldots) = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2. \qquad [21.1.6]$$

A white noise process u_t satisfying [21.1.5] is described as an *autoregressive conditional heteroskedastic* process of order m, denoted $u_t \sim ARCH(m)$. This class of processes was introduced by Engle (1982).[1]

Since u_t is random and u_t^2 cannot be negative, this can be a sensible representation only if [21.1.6] is positive and [21.1.5] is nonnegative for all realizations of $\{u_t\}$. This can be ensured if w_t is bounded from below by $-\zeta$ with $\zeta > 0$ and if $\alpha_j \geq 0$ for $j = 1, 2, \ldots, m$. In order for u_t^2 to be covariance-stationary, we further require that the roots of

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_m z^m = 0$$

[1]A nice survey of *ARCH*-related models was provided by Bollerslev, Chou, and Kroner (1992).

lie outside the unit circle. If the α_j are all nonnegative, this is equivalent to the requirement that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1. \qquad [21.1.7]$$

When these conditions are satisfied, the unconditional variance of u_t is given by

$$\sigma^2 = E(u_t^2) = \zeta/(1 - \alpha_1 - \alpha_2 - \cdots - \alpha_m). \qquad [21.1.8]$$

Let $\hat{u}^2_{t+s|t}$ denote an s-period-ahead linear forecast:

$$\hat{u}^2_{t+s|t} = \hat{E}(u^2_{t+s}|u_t^2, u^2_{t-1}, \ldots).$$

This can be calculated as in [4.2.27] by iterating on

$$(\hat{u}^2_{t+j|t} - \sigma^2) = \alpha_1(\hat{u}^2_{t+j-1|t} - \sigma^2) + \alpha_2(\hat{u}^2_{t+j-2|t} - \sigma^2) + \cdots + \alpha_m(\hat{u}^2_{t+j-m|t} - \sigma^2)$$

for $j = 1, 2, \ldots, s$ where

$$\hat{u}^2_{\tau|t} = u_\tau^2 \qquad \text{for } \tau \leq t.$$

The s-period-ahead forecast $\hat{u}^2_{t+s|t}$ converges in probability to σ^2 as $s \to \infty$, assuming that w_t has finite variance and that [21.1.7] is satisfied.

It is often convenient to use an alternative representation for an $ARCH(m)$ process that imposes slightly stronger assumptions about the serial dependence of u_t. Suppose that

$$u_t = \sqrt{h_t} \cdot v_t, \qquad [21.1.9]$$

where $\{v_t\}$ is an i.i.d. sequence with zero mean and unit variance:

$$E(v_t) = 0 \qquad E(v_t^2) = 1.$$

If h_t evolves according to

$$h_t = \zeta + \alpha_1 u^2_{t-1} + \alpha_2 u^2_{t-2} + \cdots + \alpha_m u^2_{t-m}, \qquad [21.1.10]$$

then [21.1.9] implies that

$$E(u_t^2|u_{t-1}, u_{t-2}, \ldots) = \zeta + \alpha_1 u^2_{t-1} + \alpha_2 u^2_{t-2} + \cdots + \alpha_m u^2_{t-m}. \qquad [21.1.11]$$

Hence, if u_t is generated by [21.1.9] and [21.1.10], then u_t follows an $ARCH(m)$ process in which the linear projection [21.1.6] is also the conditional expectation.

Notice further that when [21.1.9] and [21.1.10] are substituted into [21.1.5], the result is

$$h_t \cdot v_t^2 = h_t + w_t.$$

Hence, under the specification in [21.1.9], the innovation w_t in the $AR(m)$ representation for u_t^2 in [21.1.5] can be expressed as

$$w_t = h_t \cdot (v_t^2 - 1). \qquad [21.1.12]$$

Note from [21.1.12] that although the unconditional variance of w_t was assumed to be constant,

$$E(w_t^2) = \lambda^2, \qquad [21.1.13]$$

the conditional variance of w_t changes over time.

The unconditional variance of w_t reflects the fourth moment of u_t, and this fourth moment does not exist for all stationary $ARCH$ models. One can see this by squaring [21.1.12] and calculating the unconditional expectation of both sides:

$$E(w_t^2) = E(h_t^2) \cdot E(v_t^2 - 1)^2. \qquad [21.1.14]$$

Taking the $ARCH(1)$ specification for illustration, we find with a little manipulation of the formulas for the mean and variance of an $AR(1)$ process that

$$\begin{aligned} E(h_t^2) &= E(\zeta + \alpha_1 u_{t-1}^2)^2 \\ &= E\{(\alpha_1^2 \cdot u_{t-1}^4) + (2\alpha_1\zeta \cdot u_{t-1}^2) + \zeta^2\} \\ &= \alpha_1^2 \cdot [\text{Var}(u_{t-1}^2) + [E(u_{t-1}^2)]^2] + 2\alpha_1\zeta \cdot E(u_{t-1}^2) + \zeta^2 \qquad [21.1.15] \\ &= \alpha_1^2 \cdot \left[\frac{\lambda^2}{1-\alpha_1^2} + \frac{\zeta^2}{(1-\alpha_1)^2}\right] + \frac{2\alpha_1\zeta^2}{1-\alpha_1} + \zeta^2 \\ &= \frac{\alpha_1^2\lambda^2}{1-\alpha_1^2} + \frac{\zeta^2}{(1-\alpha_1)^2}. \end{aligned}$$

Substituting [21.1.15] and [21.1.13] into [21.1.14], we conclude that λ^2 (the unconditional variance of w_t) must satisfy

$$\lambda^2 = \left[\frac{\alpha_1^2\lambda^2}{1-\alpha_1^2} + \frac{\zeta^2}{(1-\alpha_1)^2}\right] \times E(v_t^2 - 1)^2. \qquad [21.1.16]$$

Even when $|\alpha_1| < 1$, equation [21.1.16] may not have any real solution for λ. For example, if $v_t \sim N(0, 1)$, then $E(v_t^2 - 1)^2 = 2$ and [21.1.16] requires that

$$\frac{(1 - 3\alpha_1^2)\lambda^2}{1 - \alpha_1^2} = \frac{2\zeta^2}{(1 - \alpha_1)^2}.$$

This equation has no real solution for λ whenever $\alpha_1^2 \geq \frac{1}{3}$. Thus, if $u_t \sim ARCH(1)$ with the innovations v_t in [21.1.9] coming from a Gaussian distribution, then the second moment of w_t (or the fourth moment of u_t) does not exist unless $\alpha_1^2 < \frac{1}{3}$.

Maximum Likelihood Estimation with Gaussian v_t

Suppose that we are interested in estimating the parameters of a regression model with *ARCH* disturbances. Let the regression equation be

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t. \qquad [21.1.17]$$

Here $\mathbf{x}_t$ denotes a vector of predetermined explanatory variables, which could include lagged values of y. The disturbance term u_t is assumed to satisfy [21.1.9] and [21.1.10]. It is convenient to condition on the first m observations ($t = -m + 1, -m + 2, \ldots, 0$) and to use observations $t = 1, 2, \ldots, T$ for estimation. Let $\mathcal{Y}_t$ denote the vector of observations obtained through date t:

$$\mathcal{Y}_t = (y_t, y_{t-1}, \ldots, y_1, y_0, \ldots, y_{-m+1}, \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_1', \mathbf{x}_0', \ldots, \mathbf{x}_{-m+1}')'.$$

If $v_t \sim$ i.i.d. $N(0, 1)$ with v_t independent of both $\mathbf{x}_t$ and $\mathcal{Y}_{t-1}$, then the conditional distribution of y_t is Gaussian with mean $\mathbf{x}_t'\boldsymbol{\beta}$ and variance h_t:

$$f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{2h_t}\right), \qquad [21.1.18]$$

where

$$\begin{aligned} h_t &= \zeta + \alpha_1(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2 + \alpha_2(y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta})^2 + \cdots \\ &\quad + \alpha_m(y_{t-m} - \mathbf{x}_{t-m}'\boldsymbol{\beta})^2 \qquad [21.1.19] \\ &= [\mathbf{z}_t(\boldsymbol{\beta})]'\boldsymbol{\delta} \end{aligned}$$

for

$$\boldsymbol{\delta} \equiv (\zeta, \alpha_1, \alpha_2, \ldots, \alpha_m)'$$

$$[\mathbf{z}_t(\boldsymbol{\beta})]' \equiv [1, (y_{t-1} - \mathbf{x}'_{t-1}\boldsymbol{\beta})^2, (y_{t-2} - \mathbf{x}'_{t-2}\boldsymbol{\beta})^2, \ldots, (y_{t-m} - \mathbf{x}'_{t-m}\boldsymbol{\beta})^2].$$

Collect the unknown parameters to be estimated in an $(a \times 1)$ vector $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} \equiv (\boldsymbol{\beta}', \boldsymbol{\delta}')'.$$

The sample log likelihood conditional on the first m observations is then

$$\begin{aligned}\mathscr{L}(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}) \\ &= -(T/2)\log(2\pi) - (1/2)\sum_{t=1}^{T}\log(h_t) \\ &\quad - (1/2)\sum_{t=1}^{T}(y_t - \mathbf{x}'_t\boldsymbol{\beta})^2/h_t.\end{aligned} \qquad [21.1.20]$$

For a given numerical value for the parameter vector $\boldsymbol{\theta}$, the sequence of conditional variances can be calculated from [21.1.19] and used to evaluate the log likelihood function [21.1.20]. This can then be maximized numerically using the methods described in Section 5.7. The derivative of the log of the conditional likelihood of the tth observation with respect to the parameter vector $\boldsymbol{\theta}$, known as the tth score, is shown in Appendix 21.A to be given by

$$\begin{aligned}\underset{(a\times 1)}{\mathbf{s}_t(\boldsymbol{\theta})} &= \frac{\partial \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \{(u_t^2 - h_t)/(2h_t^2)\}\begin{bmatrix}\sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta})\end{bmatrix} + \begin{bmatrix}(\mathbf{x}_t u_t)/h_t \\ \mathbf{0}\end{bmatrix}.\end{aligned} \qquad [21.1.21]$$

The likelihood function can be maximized using the method of scoring as in Engle (1982, p. 997) or using the Berndt, Hall, Hall, and Hausman (1974) algorithm as in Bollerslev (1986, p. 317). Alternatively, the gradient of the log likelihood function can be calculated analytically from the sum of the scores,

$$\nabla\mathscr{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \mathbf{s}_t(\boldsymbol{\theta}),$$

or numerically by numerical differentiation of the log likelihood [21.1.20]. The analytically or numerically evaluated gradient could then be used with any of the numerical optimization procedures described in Section 5.7.

Imposing the stationarity condition $(\Sigma_{j=1}^{m}\alpha_j < 1)$ and the nonnegativity condition ($\alpha_j \geq 0$ for all j) can be difficult in practice. Typically, either the value of m is very small or else some ad hoc structure is imposed on the sequence $\{\alpha_j\}_{j=1}^{m}$ as in Engle (1982, equation (38)).

Maximum Likelihood Estimation with Non-Gaussian v_t

The preceding formulation of the likelihood function assumed that v_t has a Gaussian distribution. However, the unconditional distribution of many financial time series seems to have fatter tails than allowed by the Gaussian family. Some of this can be explained by the presence of *ARCH*; that is, even if v_t in [21.1.9] has a Gaussian distribution, the unconditional distribution of u_t is non-Gaussian with heavier tails than a Gaussian distribution (see Milhøj, 1985, or Bollerslev, 1986, p. 313). Even so, there is a fair amount of evidence that the conditional distribution of u_t is often non-Gaussian as well.

The same basic approach can be used with non-Gaussian distributions. For example, Bollerslev (1987) proposed that v_t in [21.1.9] might be drawn from a t distribution with ν degrees of freedom, where ν is regarded as a parameter to be estimated by maximum likelihood. If u_t has a t distribution with ν degrees of freedom and scale parameter M_t, then its density is given by

$$f(u_t) = \frac{\Gamma[(\nu + 1)/2]}{(\pi\nu)^{1/2}\Gamma(\nu/2)} M_t^{-1/2}\left[1 + \frac{u_t^2}{M_t\nu}\right]^{-(\nu+1)/2}, \qquad [21.1.22]$$

where $\Gamma(\cdot)$ is the gamma function described in the discussion following equation [12.1.18]. If $\nu > 2$, then v_t has mean zero and variance[2]

$$E(u_t^2) = M_t\nu/(\nu - 2).$$

Hence, a t variable with ν degrees of freedom and variance h_t is obtained by taking the scale parameter M_t to be

$$M_t = h_t(\nu - 2)/\nu,$$

for which the density [21.1.22] becomes

$$f(u_t) = \frac{\Gamma[(\nu + 1)/2]}{\pi^{1/2}\Gamma(\nu/2)} (\nu - 2)^{-1/2} h_t^{-1/2}\left[1 + \frac{u_t^2}{h_t(\nu - 2)}\right]^{-(\nu+1)/2}. \qquad [21.1.23]$$

This density can be used in place of the Gaussian specification [21.1.18] along with the same specification of the conditional mean and conditional variance used in [21.1.17] and [21.1.19]. The sample log likelihood conditional on the first m observations then becomes

$$\begin{aligned}\sum_{t=1}^{T} \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}) \\ &= T \log\left\{\frac{\Gamma[(\nu + 1)/2]}{\pi^{1/2}\Gamma(\nu/2)} (\nu - 2)^{-1/2}\right\} - (1/2) \sum_{t=1}^{T} \log(h_t) \qquad [21.1.24] \\ &\quad - [(\nu + 1)/2] \sum_{t=1}^{T} \log\left[1 + \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{h_t(\nu - 2)}\right],\end{aligned}$$

where

$$\begin{aligned}h_t &= \zeta + \alpha_1(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2 + \alpha_2(y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta})^2 + \cdots + \alpha_m(y_{t-m} - \mathbf{x}_{t-m}'\boldsymbol{\beta})^2 \\ &= [\mathbf{z}_t(\boldsymbol{\beta})]'\boldsymbol{\delta}.\end{aligned}$$

The log likelihood [21.1.24] is then maximized numerically with respect to ν, $\boldsymbol{\beta}$, and $\boldsymbol{\delta}$ subject to the constraint $\nu > 2$.

The same approach can be used with other distributions for v_t. Other distributions that have been employed with *ARCH*-related models include a Normal-Poisson mixture distribution (Jorion, 1988), power exponential distribution (Baillie and Bollerslev, 1989), Normal–log normal mixture (Hsieh, 1989), generalized exponential distribution (Nelson, 1991), and serially dependent mixture of Normals (Cai, forthcoming) or t variables (Hamilton and Susmel, forthcoming).

[2]See, for example, DeGroot (1970, p. 42).

Quasi-Maximum Likelihood Estimation

Even if the assumption that v_t is i.i.d. $N(0, 1)$ is invalid, we saw in [21.1.6] that the *ARCH* specification can still offer a reasonable model on which to base a linear forecast of the squared value of v_t. As shown in Weiss (1984, 1986), Bollerslev and Wooldridge (1992), and Glosten, Jagannathan, and Runkle (1989), maximization of the Gaussian log likelihood function [21.1.20] can provide consistent estimates of the parameters $\zeta, \alpha_1, \alpha_2, \ldots, \alpha_m$ of this linear representation even when the distribution of u_t is non-Gaussian, provided that v_t in [21.1.9] satisfies

$$E(v_t | \mathbf{x}_t, \mathcal{Y}_{t-1}) = 0$$

and

$$E(v_t^2 | \mathbf{x}_t, \mathcal{Y}_{t-1}) = 1.$$

However, the standard errors have to be adjusted. Let $\hat{\boldsymbol{\theta}}_T$ be the estimate that maximizes the Gaussian log likelihood [21.1.20], and let $\boldsymbol{\theta}$ be the true value that characterizes the linear representations [21.1.9], [21.1.17], and [21.1.19]. Then even when v_t is actually non-Gaussian, under certain regularity conditions

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \xrightarrow{L} N(\mathbf{0}, \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1}),$$

where

$$\mathbf{S} = \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} [\mathbf{s}_t(\boldsymbol{\theta})]\cdot[\mathbf{s}_t(\boldsymbol{\theta})]'$$

for $\mathbf{s}_t(\boldsymbol{\theta})$ the score vector as calculated in [21.1.21], and where

$$\begin{aligned}
\mathbf{D} &= \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} -E\left\{ \frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \middle| \mathbf{x}_t, \mathcal{Y}_{t-1} \right\} \\
&= \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \left\{ [1/(2h_t^2)] \begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix} \right. \\
&\quad \times \left. \begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j}' & [\mathbf{z}_t(\boldsymbol{\beta})]' \end{bmatrix} + (1/h_t) \begin{bmatrix} \mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right\},
\end{aligned} \qquad [21.1.25]$$

where

$$\mathcal{Y}_t = (y_t, y_{t-1}, \ldots, y_1, y_0, \ldots, y_{-m+1}, \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_1', \mathbf{x}_0', \ldots, \mathbf{x}_{-m+1}')'.$$

The second equality in [21.1.25] is established in Appendix 21.A. The matrix $\mathbf{S}$ can be consistently estimated by

$$\hat{\mathbf{S}}_T = T^{-1} \sum_{t=1}^{T} [\mathbf{s}_t(\hat{\boldsymbol{\theta}}_T)]\cdot[\mathbf{s}_t(\hat{\boldsymbol{\theta}}_T)]',$$

where $\mathbf{s}_t(\hat{\boldsymbol{\theta}}_T)$ indicates the vector given in [21.1.21] evaluated at $\hat{\boldsymbol{\theta}}_T$. Similarly, the matrix $\mathbf{D}$ can be consistently estimated by

$$\begin{aligned}
\hat{\mathbf{D}}_T = T^{-1} \sum_{t=1}^{T} &\left\{ [1/(2\hat{h}_t^2)] \begin{bmatrix} \sum_{j=1}^{m} -2\hat{\alpha}_j \hat{u}_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\hat{\boldsymbol{\beta}}) \end{bmatrix} \right. \\
&\times \left. \begin{bmatrix} \sum_{j=1}^{m} -2\hat{\alpha}_j \hat{u}_{t-j}\mathbf{x}_{t-j}' & [\mathbf{z}_t(\hat{\boldsymbol{\beta}})]' \end{bmatrix} + (1/\hat{h}_t) \begin{bmatrix} \mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right\}.
\end{aligned}$$

Standard errors for $\hat{\theta}_T$ that are robust to misspecification of the family of densities can thus be obtained from the square root of diagonal elements of

$$T^{-1}\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{S}}_T\hat{\mathbf{D}}_T^{-1}.$$

Recall that if the model is correctly specified so that the data were really generated by a Gaussian model, then $\mathbf{S} = \mathbf{D}$, and this simplifies to the usual asymptotic variance matrix for maximum likelihood estimation.

Estimation by Generalized Method of Moments

The *ARCH* regression model of [21.1.17] and [21.1.19] can be characterized by the assumptions that the residual in the regression equation is uncorrelated with the explanatory variables,

$$E[(y_t - \mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t] = \mathbf{0},$$

and that the implicit error in forecasting the squared residual is uncorrelated with lagged squared residuals,

$$E[(u_t^2 - h_t)\mathbf{z}_t] = \mathbf{0}.$$

As noted by Bates and White (1988), Mark (1988), Ferson (1989), Simon (1989), or Rich, Raymond, and Butler (1991), this means that the parameters of an *ARCH* model could be estimated by generalized method of moments,[3] choosing $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}')'$ so as to minimize

$$[\mathbf{g}(\boldsymbol{\theta}; \mathcal{Y}_T)]'\hat{\mathbf{S}}_T^{-1}[\mathbf{g}(\boldsymbol{\theta}; \mathcal{Y}_T)],$$

where

$$\mathbf{g}(\boldsymbol{\theta}; \mathcal{Y}_T) = \begin{bmatrix} T^{-1}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t \\ T^{-1}\sum_{t=1}^{T}\{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2 - [\mathbf{z}_t(\boldsymbol{\beta})]'\boldsymbol{\delta}\}\mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix}.$$

The matrix $\hat{\mathbf{S}}_T$, standard errors for parameter estimates, and tests of the model can be constructed using the methods described in Chapter 14. Any other variables believed to be uncorrelated with u_t or with $(u_t^2 - h_t)$ could be used as additional instruments.

Testing for ARCH

Fortunately, it is simple to test whether the residuals u_t from a regression model exhibit time-varying heteroskedasticity without actually having to estimate the *ARCH* parameters. Engle (1982, p. 1000) derived the following test based on the Lagrange multiplier principle. First the regression of [21.1.17] is estimated by *OLS* for observations $t = -m + 1, -m + 2, \ldots, T$ and the *OLS* sample residuals $\hat{u}_t$ are saved. Next, $\hat{u}_t^2$ is regressed on a constant and m of its own lagged values:

$$\hat{u}_t^2 = \zeta + \alpha_1\hat{u}_{t-1}^2 + \alpha_2\hat{u}_{t-2}^2 + \cdots + \alpha_m\hat{u}_{t-m}^2 + e_t, \qquad [21.1.26]$$

for $t = 1, 2, \ldots, T$. The sample size T times the uncentered R_u^2 from the regression

[3]As noted in Section 14.4, maximum likelihood estimation can itself be viewed as estimation by *GMM* in which the orthogonality condition is that the expected score is zero.

of [21.1.26] then converges in distribution to a χ^2 variable with m degrees of freedom under the null hypothesis that u_t is actually i.i.d. $N(0, \sigma^2)$.

Recalling that the $ARCH(m)$ specification can be regarded as an $AR(m)$ process for u_t^2, another approach developed by Bollerslev (1988) is to use the Box-Jenkins methods described in Section 4.8 to analyze the autocorrelations of u_t^2. Other tests for $ARCH$ are described in Bollerslev, Chou, and Kroner (1992, p. 8).

21.2. Extensions

Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

Equations [21.1.9] and [21.1.10] described an $ARCH(m)$ process (u_t) characterized by

$$u_t = \sqrt{h_t} \cdot v_t,$$

where v_t is i.i.d. with zero mean and unit variance and where h_t evolves according to

$$h_t = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2.$$

More generally, we can imagine a process for which the conditional variance depends on an infinite number of lags of u_{t-j}^2,

$$h_t = \zeta + \pi(L) u_t^2, \qquad [21.2.1]$$

where

$$\pi(L) = \sum_{j=1}^{\infty} \pi_j L^j.$$

A natural idea is to parameterize $\pi(L)$ as the ratio of two finite-order polynomials:

$$\pi(L) = \frac{\alpha(L)}{1 - \delta(L)} = \frac{\alpha_1 L^1 + \alpha_2 L^2 + \cdots + \alpha_m L^m}{1 - \delta_1 L^1 - \delta_2 L^2 - \cdots - \delta_r L^r}, \qquad [21.2.2]$$

where for now we assume that the roots of $1 - \delta(z) = 0$ are outside the unit circle. If [21.2.1] is multiplied by $1 - \delta(L)$, the result is

$$[1 - \delta(L)]h_t = [1 - \delta(1)]\zeta + \alpha(L)u_t^2$$

or

$$\begin{aligned} h_t = \kappa &+ \delta_1 h_{t-1} + \delta_2 h_{t-2} + \cdots + \delta_r h_{t-r} \\ &+ \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 \end{aligned} \qquad [21.2.3]$$

for $\kappa \equiv [1 - \delta_1 - \delta_2 - \cdots - \delta_r]\zeta$. Expression [21.2.3] is the *generalized autoregressive conditional heteroskedasticity* model, denoted $u_t \sim GARCH(r, m)$, proposed by Bollerslev (1986).

One's first guess from expressions [21.2.2] and [21.2.3] might be that $\delta(L)$ describes the "autoregressive" terms for the variance while $\alpha(L)$ captures the "moving average" terms. However, this is not the case. The easiest way to see why is to add u_t^2 to both sides of [21.2.3] and rewrite the resulting expression as

$$\begin{aligned} h_t + u_t^2 = \kappa &- \delta_1(u_{t-1}^2 - h_{t-1}) - \delta_2(u_{t-2}^2 - h_{t-2}) - \cdots \\ &- \delta_r(u_{t-r}^2 - h_{t-r}) + \delta_1 u_{t-1}^2 + \delta_2 u_{t-2}^2 + \cdots \\ &+ \delta_r u_{t-r}^2 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 + u_t^2 \end{aligned}$$

or

$$u_t^2 = \kappa + (\delta_1 + \alpha_1)u_{t-1}^2 + (\delta_2 + \alpha_2)u_{t-2}^2 + \cdots + (\delta_p + \alpha_p)u_{t-p}^2 + w_t - \delta_1 w_{t-1} - \delta_2 w_{t-2} - \cdots - \delta_r w_{t-r}, \qquad [21.2.4]$$

where $w_t \equiv u_t^2 - h_t$ and $p \equiv \max\{m, r\}$. We have further defined $\delta_j \equiv 0$ for $j > r$ and $\alpha_j \equiv 0$ for $j > m$. Notice that h_t is the forecast of u_t^2 based on its own lagged values and thus $w_t \equiv u_t^2 - h_t$ is the error associated with this forecast. Thus, w_t is a white noise process that is fundamental for u_t^2. Expression [21.2.4] will then be recognized as an $ARMA(p, r)$ process for u_t^2, in which the jth autoregressive coefficient is the sum of δ_j plus α_j while the jth moving average coefficient is the negative of δ_j. If u_t is described by a $GARCH(r, m)$ process, then u_t^2 follows an $ARMA(p, r)$ process, where p is the larger of r and m.

The nonnegativity requirement is satisfied if $\kappa > 0$ and $\alpha_j \geq 0$, $\delta_j \geq 0$ for $j = 1, 2, \ldots, p$. From our analysis of $ARMA$ processes, it then follows that u_t^2 is covariance-stationary provided that w_t has finite variance and that the roots of

$$1 - (\delta_1 + \alpha_1)z - (\delta_2 + \alpha_2)z^2 - \cdots - (\delta_p + \alpha_p)z^p = 0$$

are outside the unit circle. Given the nonnegativity restriction, this means that u_t^2 is covariance-stationary if

$$(\delta_1 + \alpha_1) + (\delta_2 + \alpha_2) + \cdots + (\delta_p + \alpha_p) < 1.$$

Assuming that this condition holds, the unconditional mean of u_t^2 is

$$E(u_t^2) = \sigma^2 = \kappa/[1 - (\delta_1 + \alpha_1) - (\delta_2 + \alpha_2) - \cdots - (\delta_p + \alpha_p)].$$

Nelson and Cao (1992) noted that the conditions $\alpha_j \geq 0$ and $\delta_j \geq 0$ are sufficient but not necessary to ensure nonnegativity of h_t. For example, for a $GARCH(1, 2)$ process, the $\pi(L)$ operator implied by [21.2.2] is given by

$$\begin{aligned}\pi(L) &= (1 - \delta_1 L)^{-1}(\alpha_1 L + \alpha_2 L^2) \\ &= (1 + \delta_1 L + \delta_1^2 L^2 + \delta_1^3 L^3 + \cdots)(\alpha_1 L + \alpha_2 L^2) \\ &= \alpha_1 L + (\delta_1\alpha_1 + \alpha_2)L^2 + \delta_1(\delta_1\alpha_1 + \alpha_2)L^3 \\ &\quad + \delta_1^2(\delta_1\alpha_1 + \alpha_2)L^4 + \cdots.\end{aligned}$$

The π_j coefficients are all nonnegative provided that $0 \leq \delta_1 < 1$, $\alpha_1 \geq 0$, and $(\delta_1\alpha_1 + \alpha_2) \geq 0$. Hence, α_2 could be negative as long as $-\alpha_2$ is less than $\delta_1\alpha_1$.

The forecast of u_{t+s}^2 based on $u_t^2, u_{t-1}^2, \ldots$, denoted $\hat{u}_{t+s|t}^2$, can be calculated as in [4.2.45] by iterating on

$$\hat{u}_{t+s|t}^2 - \sigma^2 = \begin{cases} (\delta_1 + \alpha_1)(\hat{u}_{t+s-1|t}^2 - \sigma^2) + (\delta_2 + \alpha_2)(\hat{u}_{t+s-2|t}^2 - \sigma^2) \\ + \cdots + (\delta_p + \alpha_p)(\hat{u}_{t+s-p|t}^2 - \sigma^2) - \delta_s\hat{w}_t - \delta_{s+1}\hat{w}_{t-1} \\ - \cdots - \delta_r\hat{w}_{t+s-r} & \text{for } s = 1, 2, \ldots, r \\ (\delta_1 + \alpha_1)(\hat{u}_{t+s-1|t}^2 - \sigma^2) + (\delta_2 + \alpha_2)(\hat{u}_{t+s-2|t}^2 - \sigma^2) \\ + \cdots + (\delta_p + \alpha_p)(\hat{u}_{t+s-p|t}^2 - \sigma^2) & \text{for } s = r + 1, r + 2, \ldots, \end{cases}$$

$$\begin{aligned}\hat{u}_{\tau|t}^2 &= u_\tau^2 && \text{for } \tau \leq t \\ \hat{w}_\tau &= u_\tau^2 - \hat{u}_{\tau|\tau-1}^2 && \text{for } \tau = t, t-1, \ldots, t-r+1.\end{aligned}$$

See Baillie and Bollerslev (1992) for further discussion of forecasts and mean squared errors for $GARCH$ processes.

Calculation of the sequence of conditional variances $\{h_t\}_{t=1}^T$ from [21.2.3] requires presample values for $h_{-p+1}, \ldots, h_0$ and $u_{-p+1}^2, \ldots, u_0^2$. If we have observations on y_t and $\mathbf{x}_t$ for $t = 1, 2, \ldots, T$, Bollerslev (1986, p. 316) suggested setting

$$h_j = u_j^2 = \hat{\sigma}^2 \qquad \text{for } j = -p + 1, \ldots, 0,$$

where

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2.$$

The sequence $\{h_t\}_{t=1}^{T}$ can be used to evaluate the log likelihood from the expression given in [21.1.20]. This can then be maximized numerically with respect to $\boldsymbol{\beta}$ and the parameters $\kappa, \delta_1, \ldots, \delta_r, \alpha_1, \ldots, \alpha_m$ of the *GARCH* process; for details, see Bollerslev (1986).

Integrated GARCH

Suppose that $u_t = \sqrt{h_t} \cdot v_t$, where v_t is i.i.d. with zero mean and unit variance and where h_t obeys the $GARCH(r, m)$ specification

$$\begin{aligned} h_t = \kappa &+ \delta_1 h_{t-1} + \delta_2 h_{t-2} + \cdots + \delta_r h_{t-r} \\ &+ \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2. \end{aligned}$$

We saw in [21.2.4] that this implies an *ARMA* process for u_t^2 where the jth autoregressive coefficient is given by $(\delta_j + \alpha_j)$. This *ARMA* process for u_t^2 would have a unit root if

$$\sum_{j=1}^{r} \delta_j + \sum_{j=1}^{m} \alpha_j = 1. \qquad [21.2.5]$$

Engle and Bollerslev (1986) referred to a model satisfying [21.2.5] as an *integrated GARCH* process, denoted *IGARCH*.

If u_t follows an *IGARCH* process, then the unconditional variance of u_t is infinite, so neither u_t nor u_t^2 satisfies the definition of a covariance-stationary process. However, it is still possible for u_t to come from a strictly stationary process in the sense that the unconditional density of u_t is the same for all t; see Nelson (1990).

The ARCH-*in-Mean Specification*

Finance theory suggests that an asset with a higher perceived risk would pay a higher return on average. For example, let r_t denote the ex post rate of return on some asset minus the return on a safe alternative asset. Suppose that r_t is decomposed into a component anticipated by investors at date $t - 1$ (denoted μ_t) and a component that was unanticipated (denoted u_t):

$$r_t = \mu_t + u_t.$$

Then the theory suggests that the mean return (μ_t) would be related to the variance of the return (h_t). In general, the *ARCH*-in-mean, or *ARCH-M*, regression model introduced by Engle, Lilien, and Robins (1987) is characterized by

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\beta} + \delta h_t + u_t \\ u_t &= \sqrt{h_t} \cdot v_t \\ h_t &= \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 \end{aligned}$$

for v_t i.i.d. with zero mean and unit variance. The effect that higher perceived variability of u_t has on the level of y_t is captured by the parameter δ.

Exponential GARCH

As before, let $u_t = \sqrt{h_t} \cdot v_t$ where v_t is i.i.d. with zero mean and unit variance. Nelson (1991) proposed the following model for the evolution of the conditional variance of u_t:

$$\log h_t = \zeta + \sum_{j=1}^{\infty} \pi_j \cdot \{|v_{t-j}| - E|v_{t-j}| + \aleph v_{t-j}\}. \qquad [21.2.6]$$

Nelson's model is sometimes referred to as *exponential GARCH*, or *EGARCH*. If $\pi_j > 0$, Nelson's model implies that a deviation of $|v_{t-j}|$ from its expected value causes the variance of u_t to be larger than otherwise, an effect similar to the idea behind the *GARCH* specification.

The $\aleph$ parameter allows this effect to be asymmetric. If $\aleph = 0$, then a positive surprise ($v_{t-j} > 0$) has the same effect on volatility as a negative surprise of the same magnitude. If $-1 < \aleph < 0$, a positive surprise increases volatility less than a negative surprise. If $\aleph < -1$, a positive surprise actually reduces volatility while a negative surprise increases volatility. A number of researchers have found evidence of asymmetry in stock price behavior—negative surprises seem to increase volatility more than positive surprises.[4] Since a lower stock price reduces the value of equity relative to corporate debt, a sharp decline in stock prices increases corporate leverage and could thus increase the risk of holding stocks. For this reason, the apparent finding that $\aleph < 0$ is sometimes described as the *leverage effect.*

One of the key advantages of Nelson's specification is that since [21.2.6] describes the log of h_t, the variance itself (h_t) will be positive regardless of whether the π_j coefficients are positive. Thus, in contrast to the *GARCH* model, no restrictions need to be imposed on [21.2.6] for estimation. This makes numerical optimization simpler and allows a more flexible class of possible dynamic models for the variance. Nelson (1991, p. 351) showed that [21.2.6] implies that $\log h_t$, h_t, and u_t are all strictly stationary provided that $\Sigma_{j=1}^{\infty} \pi_j^2 < \infty$.

A natural parameterization is to model $\pi(L)$ as the ratio of two finite-order polynomials as in the *GARCH(r, m)* specification:

$$\begin{aligned} \log h_t = \kappa &+ \delta_1 \log h_{t-1} + \delta_2 \log h_{t-2} + \cdots \\ &+ \delta_r \log h_{t-r} + \alpha_1\{|v_{t-1}| - E|v_{t-1}| + \aleph v_{t-1}\} \\ &+ \alpha_2\{|v_{t-2}| - E|v_{t-2}| + \aleph v_{t-2}\} + \cdots \\ &+ \alpha_m\{|v_{t-m}| - E|v_{t-m}| + \aleph v_{t-m}\}. \end{aligned} \qquad [21.2.7]$$

The *EGARCH* model can be estimated by maximum likelihood by specifying a density for v_t. Nelson proposed using the *generalized error distribution*, normalized to have zero mean and unit variance:

$$f(v_t) = \frac{\nu \exp[-(1/2)|v_t/\lambda|^{\nu}]}{\lambda \cdot 2^{[(\nu+1)/\nu]}\Gamma(1/\nu)}. \qquad [21.2.8]$$

Here $\Gamma(\cdot)$ is the gamma function, λ is a constant given by

$$\lambda = \left\{\frac{2^{(-2/\nu)}\Gamma(1/\nu)}{\Gamma(3/\nu)}\right\}^{1/2},$$

[4]See Pagan and Schwert (1990), Engle and Ng (1991), and the studies cited in Bollerslev, Chou, and Kroner (1992, p. 24).

and ν is a positive parameter governing the thickness of the tails. For $\nu = 2$, the constant $\lambda = 1$ and expression [21.2.8] is just the standard Normal density. If $\nu < 2$, the density has thicker tails than the Normal, whereas for $\nu > 2$ it has thinner tails. The expected absolute value of a variable drawn from this distribution is

$$E|v_t| = \frac{\lambda \cdot 2^{1/\nu}\Gamma(2/\nu)}{\Gamma(1/\nu)}.$$

For the standard Normal case ($\nu = 2$), this becomes

$$E|v_t| = \sqrt{2/\pi}.$$

As an illustration of how this model might be used, consider Nelson's analysis of stock return data. For r_t the daily return on stocks minus the daily interest rate on Treasury bills, Nelson estimated a regression model of the form

$$r_t = a + br_{t-1} + \delta h_t + u_t.$$

The residual u_t was modeled as $\sqrt{h_t} \cdot v_t$, where v_t is i.i.d. with density [21.2.8] and where h_t evolves according to

$$\begin{aligned} \log h_t - \zeta_t = \delta_1(\log h_{t-1} - \zeta_{t-1}) + \delta_2(\log h_{t-2} - \zeta_{t-2}) \\ + \alpha_1\{|v_{t-1}| - E|v_{t-1}| + \aleph v_{t-1}\} \\ + \alpha_2\{|v_{t-2}| - E|v_{t-2}| + \aleph v_{t-2}\}. \end{aligned} \qquad [21.2.9]$$

Nelson allowed ζ_t, the unconditional mean of $\log h_t$, to be a function of time:

$$\zeta_t = \zeta + \log(1 + \rho N_t),$$

where N_t denotes the number of nontrading days between dates $t - 1$ and t and ζ and ρ are parameters to be estimated by maximum likelihood. The sample log likelihood is then

$$\begin{aligned} \mathcal{L} = T\{\log(\nu/\lambda) - (1 + \nu^{-1})\log(2) - \log[\Gamma(1/\nu)]\} \\ - (1/2)\sum_{t=1}^{T} |(r_t - a - br_{t-1} - \delta h_t)/(\lambda \cdot \sqrt{h_t})|^{\nu} - (1/2)\sum_{t=1}^{T} \log(h_t). \end{aligned}$$

The sequence $\{h_t\}_{t=1}^{T}$ is obtained by iterating on [21.2.7] with

$$v_t = (r_t - a - br_{t-1} - \delta h_t)/\sqrt{h_t}$$

and with presample values of $\log h_t$ set to their unconditional expectations ζ_t.

Other Nonlinear ARCH *Specifications*

Asymmetric consequences of positive and negative innovations can also be captured with a simple modification of the linear *GARCH* framework. Glosten, Jagannathan, and Runkle (1989) proposed modeling $u_t = \sqrt{h_t} \cdot v_t$, where v_t is i.i.d. with zero mean and unit variance and

$$h_t = \kappa + \delta_1 h_{t-1} + \alpha_1 u_{t-1}^2 + \aleph u_{t-1}^2 \cdot I_{t-1}. \qquad [21.2.10]$$

Here, $I_{t-1} = 1$ if $u_{t-1} \geq 0$ and $I_{t-1} = 0$ if $u_{t-1} < 0$. Again, if the leverage effect holds, we expect to find $\aleph < 0$. The nonnegativity condition is satisfied provided that $\delta_1 \geq 0$ and $\alpha_1 + \aleph \geq 0$.

A variety of other nonlinear functional forms relating h_t to $\{u_{t-1}, u_{t-2}, \ldots\}$ have been proposed. Geweke (1986), Pantula (1986), and Milhøj (1987) suggested a specification in which the log of h_t depends linearly on past logs of the squared residuals. Higgins and Bera (1992) proposed a power transformation of the form

$$h_t = [\zeta^\delta + \alpha_1(u_{t-1}^2)^\delta + \alpha_2(u_{t-2}^2)^\delta + \cdots + \alpha_m(u_{t-m}^2)^\delta]^{1/\delta},$$

with $\zeta > 0$, $\delta > 0$, and $\alpha_i \geq 0$ for $i = 1, 2, \ldots, m$. Gourieroux and Monfort (1992) used a Markov chain to model the conditional variance as a general stepwise function of past realizations.

Multivariate GARCH *Models*

The preceding ideas can also be extended to an $(n \times 1)$ vector $\mathbf{y}_t$. Consider a system of n regression equations of the form

$$\underset{(n\times 1)}{\mathbf{y}_t} = \underset{(n\times k)}{\mathbf{\Pi}'} \cdot \underset{(k\times 1)}{\mathbf{x}_t} + \underset{(n\times 1)}{\mathbf{u}_t},$$

where $\mathbf{x}_t$ is a vector of explanatory variables and $\mathbf{u}_t$ is a vector of white noise residuals. Let $\mathbf{H}_t$ denote the $(n \times n)$ conditional variance-covariance matrix of the residuals:

$$\mathbf{H}_t = E(\mathbf{u}_t\mathbf{u}_t' | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots).$$

Engle and Kroner (1993) proposed the following vector generalization of a *GARCH*(r, m) specification:

$$\mathbf{H}_t = \mathbf{K} + \mathbf{\Delta}_1\mathbf{H}_{t-1}\mathbf{\Delta}_1' + \mathbf{\Delta}_2\mathbf{H}_{t-2}\mathbf{\Delta}_2' + \cdots + \mathbf{\Delta}_r\mathbf{H}_{t-r}\mathbf{\Delta}_r' + \mathbf{A}_1\mathbf{u}_{t-1}\mathbf{u}_{t-1}'\mathbf{A}_1'$$
$$+ \mathbf{A}_2\mathbf{u}_{t-2}\mathbf{u}_{t-2}'\mathbf{A}_2' + \cdots + \mathbf{A}_m\mathbf{u}_{t-m}\mathbf{u}_{t-m}'\mathbf{A}_m'.$$

Here $\mathbf{K}$, $\mathbf{\Delta}_s$, and $\mathbf{A}_s$ for $s = 1, 2, \ldots$ denote $(n \times n)$ matrices of parameters. An advantage of this parameterization is that $\mathbf{H}_t$ is guaranteed to be positive definite as long as $\mathbf{K}$ is positive definite, which can be ensured numerically by parameterizing $\mathbf{K}$ as $\mathbf{PP}'$, where $\mathbf{P}$ is a lower triangular matrix.

In practice, for reasonably sized n it is necessary to restrict the specification for $\mathbf{H}_t$ further to obtain a numerically tractable formulation. One useful special case restricts $\mathbf{\Delta}_s$ and $\mathbf{A}_s$ to be diagonal matrices for $s = 1, 2, \ldots$. In such a model, the conditional covariance between u_{it} and u_{jt} depends only on past values of $u_{i,t-s} \cdot u_{j,t-s}$, and not on the products or squares of other residuals.

Another popular approach introduced by Bollerslev (1990) assumes that the conditional correlations among the elements of $\mathbf{u}_t$ are constant over time. Let $h_{ii}^{(t)}$ denote the row i, column i element of $\mathbf{H}_t$. Thus, $h_{ii}^{(t)}$ represents the conditional variance of the ith element of $\mathbf{u}_t$:

$$h_{ii}^{(t)} = E(u_{it}^2 | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots).$$

This conditional variance might be modeled with a univariate *GARCH*$(1, 1)$ process driven by the lagged innovation in variable i:

$$h_{ii}^{(t)} = \kappa_i + \delta_i h_{ii}^{(t-1)} + \alpha_i u_{i,t-1}^2.$$

We might postulate n such *GARCH* specifications $(i = 1, 2, \ldots, n)$, one for each element of $\mathbf{u}_t$. The conditional covariance between u_{it} and u_{jt}, or the row i, column j element of $\mathbf{H}_t$, is then taken to be a constant correlation ρ_{ij} times the conditional standard deviations of u_{it} and u_{jt}:

$$h_{ij}^{(t)} = E(u_{it}u_{jt} | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots) = \rho_{ij} \cdot \sqrt{h_{ii}^{(t)}} \cdot \sqrt{h_{jj}^{(t)}}.$$

Maximum likelihood estimation of this specification turns out to be quite tractable; see Bollerslev (1990) for details.

Other multivariate models include a formulation for vech($\mathbf{H}_t$) proposed by Bollerslev, Engle, and Wooldridge (1988) and the factor *ARCH* specifications of Diebold and Nerlove (1989) and Engle, Ng, and Rothschild (1990).

Nonparametric Estimates

Pagan and Hong (1990) explored a nonparametric kernel estimate of the expected value of u_t^2. The estimate is based on an average value of those u_τ^2 whose preceding values of $u_{\tau-1}, u_{\tau-2}, \ldots, u_{\tau-m}$ were "close" to the values that preceded u_t^2:

$$h_t = \sum_{\substack{\tau=1 \\ \tau \neq t}}^{T} w_\tau(t) \cdot u_\tau^2.$$

The weights $\{w_\tau(t)\}_{\tau=1,\tau\neq t}^{T}$ are a set of $(T - 1)$ numbers that sum to unity. If the values of $u_{\tau-1}, u_{\tau-2}, \ldots, u_{\tau-m}$ that preceded u_τ were similar to the values $u_{t-1}, u_{t-2}, \ldots, u_{t-m}$ that preceded u_t, then u_τ^2 is viewed as giving useful information about $h_t = E(u_t^2 | u_{t-1}, u_{t-2}, \ldots, u_{t-m})$. In this case, the weight $w_\tau(t)$ would be large. If the values that preceded u_τ are quite different from those that preceded u_t, then u_τ^2 is viewed as giving little information about h_t and so $w_\tau(t)$ is small. One popular specification for the weight $w_\tau(t)$ is to use a Gaussian kernel:

$$\kappa_\tau(t) = \prod_{j=1}^{m} (2\pi)^{-1/2} \lambda_j^{-1} \exp[-(u_{\tau-j} - u_{t-j})^2/(2\lambda_j^2)].$$

The positive parameter λ_j is known as the *bandwidth*. The bandwidth calibrates the distance between $u_{\tau-j}$ and u_{t-j}—the smaller is λ_j, the closer $u_{\tau-j}$ must be to u_{t-j} before giving the value of u_τ^2 much weight in estimating h_t. To ensure that the weights $w_\tau(t)$ sum to unity, we take

$$w_\tau(t) = \frac{\kappa_\tau(t)}{\sum\limits_{\substack{\tau=1 \\ \tau \neq t}}^{T} \kappa_\tau(t)}.$$

The key difficulty with constructing this estimate is in choosing the bandwidth parameter λ_j. One approach is known as *cross-validation*. To illustrate this approach, suppose that the same bandwidth is selected for each lag ($\lambda_j = \lambda$ for $j = 1, 2, \ldots, m$). Then the nonparametric estimate of h_t is implicitly a function of the bandwidth parameter imposed, and accordingly could be denoted $h_t(\lambda)$. We might then choose λ so as to minimize

$$\sum_{t=1}^{T} [u_t^2 - h_t(\lambda)]^2.$$

Semiparametric Estimates

Other approaches to describing the conditional variance of u_t include general series expansions for the function $h_t = h(u_{t-1}, u_{t-2}, \ldots)$ as in Pagan and Schwert (1990, p. 278) or for the density $f(v_t)$ itself as in Gallant and Tauchen (1989) and Gallant, Hsieh, and Tauchen (1989). Engle and Gonzalez-Rivera (1991) combined a parametric specification for h_t with a nonparametric estimate of the density of v_t in [21.1.9].

Comparison of Alternative Models of Stock Market Volatility

A number of approaches have been suggested for comparing alternative *ARCH* specifications. One appealing measure is to see how well different models of heteroskedasticity forecast the value of u_t^2. Pagan and Schwert (1990) fitted a number of different models to monthly U.S. stock returns from 1834 to 1925. They found that the semiparametric and nonparametric methods did a good job in sample, though the parametric models yielded superior out-of-sample forecasts. Nelson's *EGARCH* specification was one of the best in overall performance from this comparison. Pagan and Schwert concluded that some benefits emerge from using parametric and nonparametric methods together.

Another approach is to calculate various specification tests of the fitted model. Tests can be constructed from the Lagrange mutiplier principle as in Engle, Lilien, and Robins (1987) or Higgins and Bera (1992), on moment tests and analysis of outliers as in Nelson (1991), or on the information matrix equality as in Bera and Zuo (1991). Related robust diagnostics were developed by Bollerslev and Wooldridge (1992). Other diagnostics are illustrated in Hsieh (1989). Engle and Ng (1991) suggested some particularly simple tests of the functional form of h_t related to Lagrange multiplier tests, from which they concluded that Nelson's *EGARCH* specification or Glosten, Jagannathan, and Runkle's modification of *GARCH* described in [21.2.10] best describes the asymmetry in the conditional volatility of Japanese stock returns.

Engle and Mustafa (1992) proposed another approach to assessing the usefulness of a given specification of the conditional variance based on the observed prices for security options. These financial instruments give an investor the right to buy or sell the security at some date in the future at a price agreed upon today. The value of such an option increases with the perceived variability of the security. If the term for which the option applies is sufficiently short that stock prices can be approximated by Brownian motion with constant variance, a well-known formula developed by Black and Scholes (1973) relates the price of the option to investors' perception of the variance of the stock price. The observed option prices can then be used to construct the market's implicit perception of h_t, which can be compared with the specification implied by a given time series model. The results of such comparisons are quite favorable to simple *GARCH* and *EGARCH* specifications. Studies by Day and Lewis (1992) and Lamoureux and Lastrapes (1993) suggest that *GARCH*(1, 1) or *EGARCH*(1, 1) models can improve on the market's implicit assessment of h_t. Related evidence in support of the *GARCH*(1, 1) formulation was provided by Engle, Hong, Kane, and Noh (1991) and West, Edison, and Cho (1993).

APPENDIX 21.A. *Derivation of Selected Equations for Chapter 21*

This appendix provides the details behind several of the assertions in the text.

■ **Derivation of [21.1.21].** Observe that

$$\frac{\partial \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\frac{\partial \log h_t}{\partial \boldsymbol{\theta}} - \frac{1}{2}\left\{\frac{1}{h_t}\frac{\partial (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\partial \boldsymbol{\theta}} - \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\right\}. \qquad [21.A.1]$$

But

$$\frac{\partial(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\partial\boldsymbol{\theta}} = \begin{bmatrix} -2\mathbf{x}_t u_t \\ \mathbf{0} \end{bmatrix} \qquad [21.A.2]$$

and

$$\begin{aligned}
\frac{\partial h_t}{\partial\boldsymbol{\theta}} &= \frac{\partial\left(\zeta + \sum_{j=1}^{m} \alpha_j u_{t-j}^2\right)}{\partial\boldsymbol{\theta}} \\
&= \partial\zeta/\partial\boldsymbol{\theta} + \sum_{j=1}^{m} (\partial\alpha_j/\partial\boldsymbol{\theta})\cdot u_{t-j}^2 + \sum_{j=1}^{m} \alpha_j\cdot(\partial u_{t-j}^2/\partial\boldsymbol{\theta}) \\
&= \begin{bmatrix} \mathbf{0} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ 0 \\ u_{t-1}^2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} \mathbf{0} \\ 0 \\ 0 \\ \vdots \\ u_{t-m}^2 \end{bmatrix} + \sum_{j=1}^{m} \alpha_j \begin{bmatrix} -2u_{t-j}\mathbf{x}_{t-j} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix}.
\end{aligned} \qquad [21.A.3]$$

Substituting [21.A.2] and [21.A.3] into [21.A.1] produces

$$\frac{\partial \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\left\{\frac{1}{2h_t} - \frac{u_t^2}{2h_t^2}\right\} \begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix} + \begin{bmatrix} (\mathbf{x}_t u_t)/h_t \\ \mathbf{0} \end{bmatrix},$$

as claimed. ■

■ **Derivation of [21.1.25].** Expression [21.A.1] can be written

$$\mathbf{s}_t(\boldsymbol{\theta}) = \frac{1}{2}\left\{\frac{u_t^2}{h_t} - 1\right\}\frac{\partial \log h_t}{\partial\boldsymbol{\theta}} - \frac{1}{2h_t}\frac{\partial u_t^2}{\partial\boldsymbol{\theta}},$$

from which

$$\begin{aligned}
\frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= \frac{1}{2}\frac{\partial \log h_t}{\partial\boldsymbol{\theta}}\left\{\frac{1}{h_t}\frac{\partial u_t^2}{\partial\boldsymbol{\theta}'} - \frac{u_t^2}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right\} + \frac{1}{2}\left\{\frac{u_t^2}{h_t} - 1\right\}\frac{\partial^2 \log h_t}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} \\
&\quad - \frac{1}{2h_t}\frac{\partial^2 u_t^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} + \frac{\partial u_t^2}{\partial\boldsymbol{\theta}}\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}.
\end{aligned} \qquad [21.A.4]$$

From expression [21.A.2],

$$\begin{aligned}
\frac{\partial^2 u_t^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} &= \begin{bmatrix} -2\mathbf{x}_t \\ \mathbf{0} \end{bmatrix}\frac{\partial u_t}{\partial\boldsymbol{\theta}'} \\
&= \begin{bmatrix} 2\mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.
\end{aligned}$$

Substituting this and [21.A.2] into [21.A.4] results in

$$\begin{aligned}
\frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= \frac{1}{2}\frac{\partial \log h_t}{\partial\boldsymbol{\theta}}\left\{\frac{1}{h_t}[-2u_t\mathbf{x}_t' \quad \mathbf{0}'] - \frac{u_t^2}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right\} + \frac{1}{2}\left\{\frac{u_t^2}{h_t} - 1\right\}\frac{\partial^2 \log h_t}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} \\
&\quad - \frac{1}{2h_t}\begin{bmatrix} 2\mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -2\mathbf{x}_t u_t \\ \mathbf{0} \end{bmatrix}\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}.
\end{aligned} \qquad [21.A.5]$$

Recall that conditional on $\mathbf{x}_t$ and on $\mathcal{Y}_{t-1}$, the magnitudes h_t and $\mathbf{x}_t$ are nonstochastic and

$$E(u_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = 0$$
$$E(u_t^2|\mathbf{x}_t, \mathcal{Y}_{t-1}) = h_t.$$

Thus, taking expectations of [21.A.5] conditional on $\mathbf{x}_t$ and $\mathcal{Y}_{t-1}$ results in

$$E\left\{\frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \,\middle|\, \mathbf{x}_t, \mathcal{Y}_{t-1}\right\} = -\frac{1}{2}\frac{\partial \log h_t}{\partial \boldsymbol{\theta}}\frac{\partial \log h_t}{\partial \boldsymbol{\theta}'} - \frac{1}{h_t}\begin{bmatrix} \mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$
$$= -\frac{1}{2h_t^2}\begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix}\begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j}' & [\mathbf{z}_t(\boldsymbol{\beta})]' \end{bmatrix}$$
$$- \frac{1}{h_t}\begin{bmatrix} \mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where the last equality follows from [21.A.3]. ■

Chapter 21 References

Baillie, Richard T., and Tim Bollerslev. 1989. "The Message in Daily Exchange Rates: A Conditional Variance Tale." *Journal of Business and Economic Statistics* 7:297–305.

——— and ———. 1992. "Prediction in Dynamic Models with Time-Dependent Conditional Variances." *Journal of Econometrics* 52:91–113.

Bates, Charles, and Halbert White. 1988. "Efficient Instrumental Variables Estimation of Systems of Implicit Heterogeneous Nonlinear Dynamic Equations with Nonspherical Errors," in William A. Barnett, Ernst R. Berndt, and Halbert White, eds., *Dynamic Econometric Modeling*. Cambridge, England: Cambridge University Press.

Bera, Anil K., and X. Zuo. 1991. "Specification Test for a Linear Regression Model with ARCH Process." University of Illinois at Champaign-Urbana. Mimeo.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman. 1974. "Estimation and Inference in Nonlinear Structural Models." *Annals of Economic and Social Measurement* 3:653–65.

Black, Fischer, and Myron Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81:637–54.

Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307–27.

———. 1987. "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics* 69:542–47.

———. 1988. "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process." *Journal of Time Series Analysis* 9:121–31.

———. 1990. "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model." *Review of Economics and Statistics* 72:498–505.

———, Ray Y. Chou, and Kenneth F. Kroner. 1992. "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics* 52:5–59.

———, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. "A Capital Asset Pricing Model with Time Varying Covariances." *Journal of Political Economy* 96:116–31.

——— and Jeffrey M. Wooldridge. 1992. "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances." *Econometric Reviews* 11:143–72.

Cai, Jun. Forthcoming. "A Markov Model of Unconditional Variance in ARCH." *Journal of Business and Economic Statistics.*

Day, Theodore E., and Craig M. Lewis. 1992. "Stock Market Volatility and the Information Content of Stock Index Options." *Journal of Econometrics* 52:267–87.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions.* New York: McGraw-Hill.

Diebold, Francis X., and Mark Nerlove. 1989. "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model." *Journal of Applied Econometrics* 4:1–21.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50:987–1007.

——— and Tim Bollerslev. 1986. "Modelling the Persistence of Conditional Variances." *Econometric Reviews* 5:1–50.

——— and Gloria Gonzalez-Rivera. 1991. "Semiparametric ARCH Models." *Journal of Business and Economic Statistics* 9:345–59.

———, Ted Hong, Alex Kane, and Jaesun Noh. 1991. "Arbitration Valuation of Variance Forecasts Using Simulated Options Markets." *Advances in Futures and Options Research* forthcoming.

——— and Kenneth F. Kroner. 1993. "Multivariate Simultaneous Generalized ARCH." UCSD. Mimeo.

———, David M. Lilien, and Russell P. Robins. 1987. "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model." *Econometrica* 55:391–407.

——— and Chowdhury Mustafa. 1992. "Implied ARCH Models from Options Prices." *Journal of Econometrics* 52:289–311.

——— and Victor K. Ng. 1991. "Measuring and Testing the Impact of News on Volatility." University of California, San Diego. Mimeo.

———, Victor K. Ng, and Michael Rothschild. 1990. "Asset Pricing with a FACTOR-ARCH Covariance Structure: Empirical Estimates for Treasury Bills." *Journal of Econometrics* 45:213–37.

Ferson, Wayne E. 1989. "Changes in Expected Security Returns, Risk, and the Level of Interest Rates." *Journal of Finance* 44:1191–1218.

Gallant, A. Ronald, David A. Hsieh, and George Tauchen. 1989. "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate 1974–83." Duke University. Mimeo.

——— and George Tauchen. 1989. "Semi Non-Parametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications." *Econometrica* 57:1091–1120.

Geweke, John. 1986. "Modeling the Persistence of Conditional Variances: A Comment." *Econometric Reviews* 5:57–61.

Glosten, Lawrence R., Ravi Jagannathan, and David Runkle. 1989. "Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." Northwestern University. Mimeo.

Gourieroux, Christian, and Alain Monfort. 1992. "Qualitative Threshold ARCH Models." *Journal of Econometrics* 52:159–99.

Hamilton, James D., and Raul Susmel. Forthcoming. "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics.*

Higgins, M. L., and A. K. Bera. 1992. "A Class of Nonlinear ARCH Models." *International Economic Review* 33:137–58.

Hsieh, David A. 1989. "Modeling Heteroscedasticity in Daily Foreign-Exchange Rates." *Journal of Business and Economic Statistics* 7:307–17.

Jorion, Philippe. 1988. "On Jump Processes in the Foreign Exchange and Stock Markets." *Review of Financial Studies* 1:427–45.

Lamoureux, Christopher G., and William D. Lastrapes. 1993. "Forecasting Stock Return Variance: Toward an Understanding of Stochastic Implied Volatilities." *Review of Financial Studies* 5:293–326.

Mark, Nelson. 1988. "Time Varying Betas and Risk Premia in the Pricing of Forward Foreign Exchange Contracts." *Journal of Financial Economics* 22:335–54.

Milhøj, Anders. 1985. "The Moment Structure of ARCH Processes." *Scandinavian Journal of Statistics* 12:281–92.

———. 1987. "A Multiplicative Parameterization of ARCH Models." Department of Statistics, University of Copenhagen. Mimeo.

Nelson, Daniel B. 1990. "Stationarity and Persistence in the GARCH(1, 1) Model." *Econometric Theory* 6:318–34.

———. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59:347–70.

——— and Charles Q. Cao. 1992. "Inequality Constraints in the Univariate GARCH Model." *Journal of Business and Economic Statistics* 10:229–35.

Pagan, Adrian R., and Y. S. Hong. 1990. "Non-Parametric Estimation and the Risk Premium," in W. Barnett, J. Powell, and G. Tauchen, eds., *Semiparametric and Nonparametric Methods in Econometrics and Statistics*. Cambridge, England: Cambridge University Press.

Pagan, Adrian R., and G. William Schwert. 1990. "Alternative Models for Conditional Stock Volatility." *Journal of Econometrics* 45:267–90.

Pagan, Adrian R., and Aman Ullah. 1988. "The Econometric Analysis of Models with Risk Terms." *Journal of Applied Econometrics* 3:87–105.

Pantula, Sastry G. 1986. "Modeling the Persistence of Conditional Variances: A Comment." *Econometric Reviews* 5:71–74.

Rich, Robert W., Jennie Raymond, and J. S. Butler. 1991. "Generalized Instrumental Variables Estimation of Autoregressive Conditional Heteroskedastic Models." *Economics Letters* 35:179–85.

Simon, David P. 1989. "Expectations and Risk in the Treasury Bill Market: An Instrumental Variables Approach." *Journal of Financial and Quantitative Analysis* 24:357–66.

Weiss, Andrew A. 1984. "ARMA Models with ARCH Errors." *Journal of Time Series Analysis* 5:129–43.

———. 1986. "Asymptotic Theory for ARCH Models: Estimation and Testing." *Econometric Theory* 2:107–31.

West, Kenneth D., Hali J. Edison, and Dongchul Cho. 1993. "A Utility Based Comparison of Some Models of Foreign Exchange Volatility." *Journal of International Economics*, forthcoming.

22 Modeling Time Series with Changes in Regime

22.1. *Introduction*

Many variables undergo episodes in which the behavior of the series seems to change quite dramatically. A striking example is provided by Figure 22.1, which is taken from Rogers's (1992) study of the volume of dollar-denominated accounts held in Mexican banks. The Mexican government adopted various measures in 1982 to try to discourage the use of such accounts, and the effects are quite dramatic in a plot of the series.

Similar dramatic breaks will be seen if one follows almost any macroeconomic or financial time series for a sufficiently long period. Such apparent changes in the time series process can result from events such as wars, financial panics, or significant changes in government policies.

How should we model a change in the process followed by a particular time series? For the data plotted in Figure 22.1, one simple idea would be that the constant term for the autoregression changed in 1982. For data prior to 1982 we might use a model such as

$$y_t - \mu_1 = \phi(y_{t-1} - \mu_1) + \varepsilon_t, \qquad [22.1.1]$$

while data after 1982 might be described by

$$y_t - \mu_2 = \phi(y_{t-1} - \mu_2) + \varepsilon_t, \qquad [22.1.2]$$

where $\mu_2 < \mu_1$.

The specification in [22.1.1] and [22.1.2] seems a plausible description of the data in Figure 22.1, but it is not altogether satisfactory as a time series model. For example, how are we to forecast a series that is described by [22.1.1] and [22.1.2]? If the process has changed in the past, clearly it could also change again in the future, and this prospect should be taken into account in forming a forecast. Moreover, the change in regime surely should not be regarded as the outcome of a perfectly foreseeable, deterministic event. Rather, the change in regime is itself a random variable. A complete time series model would therefore include a description of the probability law governing the change from μ_1 to μ_2.

These observations suggest that we might consider the process to be influenced by an unobserved random variable s_t^*, which will be called the *state* or *regime* that the process was in at date t. If $s_t^* = 1$, then the process is in regime 1, while $s_t^* = 2$ means that the process is in regime 2. Equations [22.1.1] and [22.1.2] can then equivalently be written as

$$y_t - \mu_{s_t^*} = \phi(y_{t-1} - \mu_{s_{t-1}^*}) + \varepsilon_t, \qquad [22.1.3]$$

where $\mu_{s_t^*}$ indicates μ_1 when $s_t^* = 1$ and indicates μ_2 when $s_t^* = 2$.

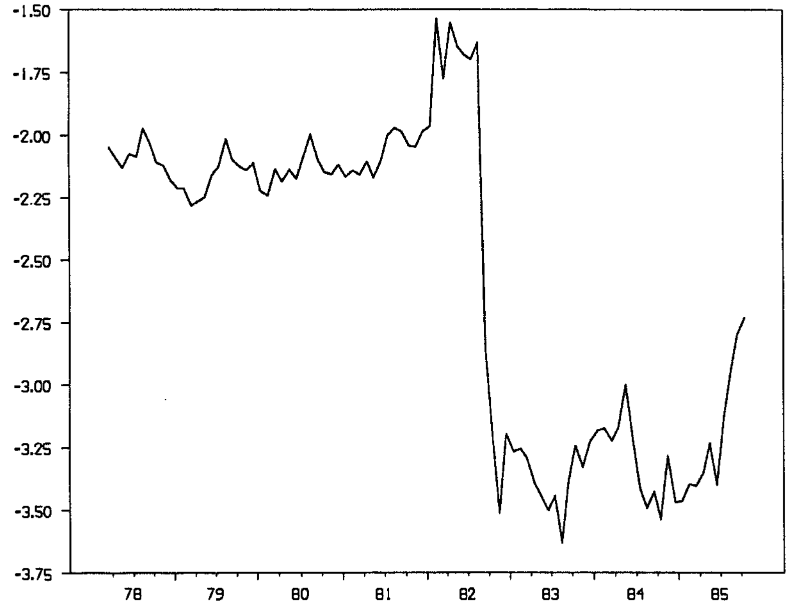


FIGURE 22.1 Log of the ratio of the peso value of dollar-denominated bank accounts in Mexico to the peso value of peso-denominated bank accounts in Mexico, monthly, 1978–85. (Rogers, 1992).

We then need a description of the time series process for the unobserved variable s_t^*. Since s_t^* takes on only discrete values (in this case, s_t^* is either 1 or 2), this will be a slightly different time series model from those for continuous-valued random variables considered elsewhere in this book.

The simplest time series model for a discrete-valued random variable is a *Markov chain*. The theory of Markov chains is reviewed in Section 22.2. In Section 22.4 this theory will be combined with a conventional time series model such as an autoregression that is assumed to characterize any given regime. Prior to doing so, however, it will be helpful to consider a special case of such processes, namely, that for which $\phi = 0$ in [22.1.3] and s_t^* is an i.i.d. discrete-valued random variable. Such a specification describes y_t as a simple mixture of different distributions, the statistical theory for which is reviewed in Section 22.3.

22.2. *Markov Chains*

Let s_t be a random variable that can assume only an integer value $\{1, 2, \ldots, N\}$. Suppose that the probability that s_t equals some particular value j depends on the past only through the most recent value s_{t-1}:

$$P\{s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots\} = P\{s_t = j | s_{t-1} = i\} = p_{ij}. \qquad [22.2.1]$$

Such a process is described as an N-state *Markov chain* with transition probabilities $\{p_{ij}\}_{i,j=1,2,\ldots,N}$. The transition probability p_{ij} gives the probability that state i will be followed by state j. Note that

$$p_{i1} + p_{i2} + \cdots + p_{iN} = 1. \qquad [22.2.2]$$

It is often convenient to collect the transition probabilities in an $(N \times N)$ matrix $\mathbf{P}$ known as the *transition matrix*:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}. \qquad [22.2.3]$$

The row j, column i element of $\mathbf{P}$ is the transition probability p_{ij}; for example, the row 2, column 1 element gives the probability that state 1 will be followed by state 2.

Representing a Markov Chain with a Vector Autoregression

A useful representation for a Markov chain is obtained by letting $\boldsymbol{\xi}_t$ denote a random $(N \times 1)$ vector whose jth element is equal to unity if $s_t = j$ and whose jth element equals zero otherwise. Thus, when $s_t = 1$, the vector $\boldsymbol{\xi}_t$ is equal to the first column of $\mathbf{I}_N$ (the $N \times N$ identity matrix); when $s_t = 2$, the vector $\boldsymbol{\xi}_t$ is the second column of $\mathbf{I}_N$; and so on:

$$\boldsymbol{\xi}_t = \begin{cases} (1, 0, 0, \ldots, 0)' & \text{when } s_t = 1 \\ (0, 1, 0, \ldots, 0)' & \text{when } s_t = 2 \\ \quad\vdots & \quad\vdots \\ (0, 0, 0, \ldots, 1)' & \text{when } s_t = N. \end{cases}$$

If $s_t = i$, then the jth element of $\boldsymbol{\xi}_{t+1}$ is a random variable that takes on the value unity with probability p_{ij} and takes on the value zero otherwise. Such a random variable has expectation p_{ij}. Thus, the conditional expectation of $\boldsymbol{\xi}_{t+1}$ given $s_t = i$ is given by

$$E(\boldsymbol{\xi}_{t+1}|s_t = i) = \begin{bmatrix} p_{i1} \\ p_{i2} \\ \vdots \\ p_{iN} \end{bmatrix}. \qquad [22.2.4]$$

This vector is simply the ith column of the matrix $\mathbf{P}$ in [22.2.3]. Moreover, when $s_t = i$, the vector $\boldsymbol{\xi}_t$ corresponds to the ith column of $\mathbf{I}_N$, in which case the vector in [22.2.4] could be described as $\mathbf{P}\boldsymbol{\xi}_t$. Hence, expression [22.2.4] implies that

$$E(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t) = \mathbf{P}\boldsymbol{\xi}_t,$$

and indeed, from the Markov property [22.2.1], it follows further that

$$E(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots) = \mathbf{P}\boldsymbol{\xi}_t. \qquad [22.2.5]$$

Result [22.2.5] implies that it is possible to express a Markov chain in the form

$$\boldsymbol{\xi}_{t+1} = \mathbf{P}\boldsymbol{\xi}_t + \mathbf{v}_{t+1}, \qquad [22.2.6]$$

where

$$\mathbf{v}_{t+1} \equiv \boldsymbol{\xi}_{t+1} - E(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots). \qquad [22.2.7]$$

Expression [22.2.6] has the form of a first-order vector autoregression for $\boldsymbol{\xi}_t$; note that [22.2.7] implies that the innovation $\mathbf{v}_t$ is a martingale difference sequence. Although the vector $\mathbf{v}_t$ can take on only a finite set of values, on average $\mathbf{v}_t$ is zero. Moreover, the value of $\mathbf{v}_t$ is impossible to forecast on the basis of previous states of the process.

Forecasts for a Markov Chain

Expression [22.2.6] implies that

$$\boldsymbol{\xi}_{t+m} = \mathbf{v}_{t+m} + \mathbf{P}\mathbf{v}_{t+m-1} + \mathbf{P}^2\mathbf{v}_{t+m-2} + \cdots + \mathbf{P}^{m-1}\mathbf{v}_{t+1} + \mathbf{P}^m\boldsymbol{\xi}_t, \quad [22.2.8]$$

where $\mathbf{P}^m$ indicates the transition matrix multiplied by itself m times. It follows from [22.2.8] that m-period-ahead forecasts for a Markov chain can be calculated from

$$E(\boldsymbol{\xi}_{t+m}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots) = \mathbf{P}^m\boldsymbol{\xi}_t. \quad [22.2.9]$$

Again, since the jth element of $\boldsymbol{\xi}_{t+m}$ will be unity if $s_{t+m} = j$ and zero otherwise, the jth element of the $(N \times 1)$ vector $E(\boldsymbol{\xi}_{t+m}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots)$ indicates the probability that s_{t+m} takes on the value j, conditional on the state of the system at date t. For example, if the process is in state i at date t, then [22.2.9] asserts that

$$\begin{bmatrix} P\{s_{t+m} = 1|s_t = i\} \\ P\{s_{t+m} = 2|s_t = i\} \\ \vdots \\ P\{s_{t+m} = N|s_t = i\} \end{bmatrix} = \mathbf{P}^m \cdot \mathbf{e}_i, \quad [22.2.10]$$

where $\mathbf{e}_i$ denotes the ith column of $\mathbf{I}_N$. Expression [22.2.10] indicates that the m-period-ahead transition probabilities for a Markov chain can be calculated by multiplying the matrix $\mathbf{P}$ by itself m times. Specifically, the probability that an observation from regime i will be followed m periods later by an observation from regime j, $P\{s_{t+m} = j|s_t = i\}$, is given by the row j, column i element of the matrix $\mathbf{P}^m$.

Reducible Markov Chains

For a two-state Markov chain, the transition matrix is

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}. \quad [22.2.11]$$

Suppose that $p_{11} = 1$, so that the matrix $\mathbf{P}$ is upper triangular. Then, once the process enters state 1, there is no possibility of ever returning to state 2. In such a case we would say that state 1 is an *absorbing state* and that the Markov chain is *reducible.*

More generally, an N-state Markov chain is said to be reducible if there exists a way to label the states (that is, a way to choose which state to call state 1, which to call state 2, and so on) such that the transition matrix can be written in the form

$$\mathbf{P} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{0} & \mathbf{D} \end{bmatrix},$$

where $\mathbf{B}$ denotes a $(K \times K)$ matrix for some $1 \le K < N$. If $\mathbf{P}$ is upper block-triangular, then so is $\mathbf{P}^m$ for any m. Hence, once such a process enters a state j such that $j \le K$, there is no possibility of ever returning to one of the states $K + 1, K + 2, \ldots, N$.

A Markov chain that is not reducihle is said to be *irreducible.* For example, a two-state chain is irreducible if $p_{11} < 1$ and $p_{22} < 1$.

Ergodic Markov Chains

Equation [22.2.2] requires that every column of $\mathbf{P}$ sum to unity, or

$$\mathbf{P}'\mathbf{1} = \mathbf{1}, \tag{22.2.12}$$

where $\mathbf{1}$ denotes an $(N \times 1)$ vector of 1s. Expression [22.2.12] implies that unity is an eigenvalue of the matrix $\mathbf{P}'$ and that $\mathbf{1}$ is the associated eigenvector. Since a matrix and its transpose share the same eigenvalues, it follows that unity is an eigenvalue of the transition matrix $\mathbf{P}$ for any Markov chain.

Consider an N-state irreducible Markov chain with transition matrix $\mathbf{P}$. Suppose that one of the eigenvalues of $\mathbf{P}$ is unity and that all other eigenvalues of $\mathbf{P}$ are inside the unit circle. Then the Markov chain is said to be *ergodic*. The $(N \times 1)$ vector of *ergodic probabilities* for an ergodic chain is denoted $\boldsymbol{\pi}$. This vector $\boldsymbol{\pi}$ is defined as the eigenvector of $\mathbf{P}$ associated with the unit eigenvalue; that is, the vector of ergodic probabilities $\boldsymbol{\pi}$ satisfies

$$\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\pi}. \tag{22.2.13}$$

The eigenvector $\boldsymbol{\pi}$ is normalized so that its elements sum to unity ($\mathbf{1}'\boldsymbol{\pi} = 1$). It can be shown that if $\mathbf{P}$ is the transition matrix for an ergodic Markov chain, then

$$\lim_{m\to\infty} \mathbf{P}^m = \boldsymbol{\pi}\cdot\mathbf{1}'. \tag{22.2.14}$$

We establish [22.2.14] here for the case when all the eigenvalues of $\mathbf{P}$ are distinct; a related argument based on the Jordan decomposition that is valid for ergodic chains with repeated eigenvalues is developed in Cox and Miller (1965, pp. 120–23). For the case of distinct eigenvalues, we know from [A.4.24] that $\mathbf{P}$ can always be written in the form

$$\mathbf{P} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^{-1}, \tag{22.2.15}$$

where $\mathbf{T}$ is an $(N \times N)$ matrix whose columns are the eigenvectors of $\mathbf{P}$ while $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal contains the corresponding eigenvalues of $\mathbf{P}$. It follows as in [1.2.19] that

$$\mathbf{P}^m = \mathbf{T}\boldsymbol{\Lambda}^m\mathbf{T}^{-1}. \tag{22.2.16}$$

Since the (1, 1) element of $\boldsymbol{\Lambda}$ is unity and all other elements of $\boldsymbol{\Lambda}$ are inside the unit circle, $\boldsymbol{\Lambda}^m$ converges to a matrix with unity in the (1, 1) position and zeros elsewhere. Hence,

$$\lim_{m\to\infty} \mathbf{P}^m = \mathbf{x}\cdot\mathbf{y}', \tag{22.2.17}$$

where $\mathbf{x}$ is the first column of $\mathbf{T}$ and $\mathbf{y}'$ is the first row of $\mathbf{T}^{-1}$.

The first column of $\mathbf{T}$ is the eigenvector of $\mathbf{P}$ corresponding to the unit eigenvalue, which eigenvector was denoted $\boldsymbol{\pi}$ in [22.2.13]:

$$\mathbf{x} = \boldsymbol{\pi}. \tag{22.2.18}$$

Moreover, the first row of $\mathbf{T}^{-1}$, when expressed as a column vector, corresponds to the eigenvector of $\mathbf{P}'$ associated with the unit eigenvalue, which eigenvector was seen to be proportional to the vector $\mathbf{1}$ in [22.2.12]:

$$\mathbf{y} = \alpha\cdot\mathbf{1}. \tag{22.2.19}$$

To verify [22.2.19], note from [22.2.15] that the matrix of eigenvectors $\mathbf{T}$ of the matrix $\mathbf{P}$ is characterized by

$$\mathbf{P}\mathbf{T} = \mathbf{T}\boldsymbol{\Lambda}. \tag{22.2.20}$$

Transposing [22.2.15] results in

$$\mathbf{P}' = (\mathbf{T}^{-1})'\mathbf{\Lambda}\mathbf{T}',$$

and postmultiplying by $(\mathbf{T}^{-1})'$ yields

$$\mathbf{P}'(\mathbf{T}^{-1})' = (\mathbf{T}^{-1})'\mathbf{\Lambda}. \qquad [22.2.21]$$

Comparing [22.2.21] with [22.2.20] confirms that the columns of $(\mathbf{T}^{-1})'$ correspond to eigenvectors of $\mathbf{P}'$. In particular, then, the first column of $(\mathbf{T}^{-1})'$ is proportional to the eigenvector of $\mathbf{P}'$ associated with the unit eigenvalue, which eigenvector was seen to be given by $\mathbf{1}$ in equation [22.2.12]. Since $\mathbf{y}$ was defined as the first column of $(\mathbf{T}^{-1})'$, this establishes the claim made in equation [22.2.19].

Substituting [22.2.18] and [22.2.19] into [22.2.17], it follows that

$$\lim_{m\to\infty} \mathbf{P}^m = \boldsymbol{\pi}\cdot\alpha\mathbf{1}'.$$

Since $\mathbf{P}^m$ can be interpreted as a matrix of transition probabilities, each column must sum to unity. Thus, since the vector of ergodic probabilities $\boldsymbol{\pi}$ was normalized by the condition that $\mathbf{1}'\boldsymbol{\pi} = 1$, it follows that the normalizing constant α must be unity, establishing the claim made in [22.2.14].

Result [22.2.14] implies that the long-run forecast for an ergodic Markov chain is independent of the current state, since, from [22.2.9],

$$E(\boldsymbol{\xi}_{t+m}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots) = \mathbf{P}^m\boldsymbol{\xi}_t \xrightarrow{p} \boldsymbol{\pi}\cdot\mathbf{1}'\boldsymbol{\xi}_t = \boldsymbol{\pi},$$

where the final equality follows from the observation that $\mathbf{1}'\boldsymbol{\xi}_t = 1$ regardless of the value of $\boldsymbol{\xi}_t$. The long-run forecast of $\boldsymbol{\xi}_{t+m}$ is given by the vector of ergodic probabilities $\boldsymbol{\pi}$ regardless of the current value of $\boldsymbol{\xi}_t$.

The vector of ergodic probabilities can also be viewed as indicating the unconditional probability of each of the N different states. To see this, suppose that we had used the symbol π_j to indicate the unconditional probability $P\{s_t = j\}$. Then the vector $\boldsymbol{\pi} \equiv (\pi_1, \pi_2, \ldots, \pi_N)'$ could be described as the unconditional expectation of $\boldsymbol{\xi}_t$:

$$\boldsymbol{\pi} = E(\boldsymbol{\xi}_t). \qquad [22.2.22]$$

If one takes unconditional expectations of [22.2.6], the result is

$$E(\boldsymbol{\xi}_{t+1}) = \mathbf{P}\cdot E(\boldsymbol{\xi}_t).$$

Assuming stationarity and using the definition [22.2.22], this becomes

$$\boldsymbol{\pi} = \mathbf{P}\cdot\boldsymbol{\pi},$$

which is identical to equation [22.2.13] characterizing $\boldsymbol{\pi}$ as the eigenvector of $\mathbf{P}$ associated with the unit eigenvalue. For an ergodic Markov chain, this eigenvector is unique, and so the vector of ergodic probabilities $\boldsymbol{\pi}$ can be interpreted as the vector of unconditional probabilities.

An ergodic Markov chain is a covariance-stationary process. Yet [22.2.6] takes the form of a *VAR* with a unit root, since one of the eigenvalues of $\mathbf{P}$ is unity. This *VAR* is stationary despite the unit root because the variance-covariance matrix of $\mathbf{v}_t$ is singular. In particular, since $\mathbf{1}'\boldsymbol{\xi}_t = 1$ for all t and since $\mathbf{1}'\mathbf{P} = \mathbf{1}'$, equation [22.2.6] implies that $\mathbf{1}'\mathbf{v}_t = 0$ for all t. Thus, from [22.2.19], the first element of the $(N \times 1)$ vector $\mathbf{T}^{-1}\mathbf{v}_t$ is always zero, meaning that from [22.2.16] the unit eigenvalue in $\mathbf{P}^m\mathbf{v}_t$ always has a coefficient of zero.

Further Discussion of Two-State Markov Chains

The eigenvalues of the transition matrix $\mathbf{P}$ for any N-state Markov chain are found from the solutions to $|\mathbf{P} - \lambda \mathbf{I}_N| = 0$. For the two-state Markov chain, the eigenvalues satisfy

$$\begin{aligned}
0 &= \begin{vmatrix} p_{11} - \lambda & 1 - p_{22} \\ 1 - p_{11} & p_{22} - \lambda \end{vmatrix} \\
&= (p_{11} - \lambda)(p_{22} - \lambda) - (1 - p_{11})(1 - p_{22}) \\
&= p_{11}p_{22} - (p_{11} + p_{22})\lambda + \lambda^2 - 1 + p_{11} + p_{22} - p_{11}p_{22} \\
&= \lambda^2 - (p_{11} + p_{22})\lambda - 1 + p_{11} + p_{22} \\
&= (\lambda - 1)(\lambda + 1 - p_{11} - p_{22}).
\end{aligned}$$

Thus, the eigenvalues for a two-state chain are given by $\lambda_1 = 1$ and $\lambda_2 = -1 + p_{11} + p_{22}$. The second eigenvalue, λ_2, will be inside the unit circle as long as $0 < p_{11} + p_{22} < 2$. We saw earlier that this chain is irreducible as long as $p_{11} < 1$ and $p_{22} < 1$. Thus, a two-state Markov chain is ergodic provided that $p_{11} < 1$, $p_{22} < 1$, and $p_{11} + p_{22} > 0$.

The eigenvector associated with λ_1 for the two-state chain turns out to be

$$\boldsymbol{\pi} = \begin{bmatrix} (1 - p_{22})/(2 - p_{11} - p_{22}) \\ (1 - p_{11})/(2 - p_{11} - p_{22}) \end{bmatrix}$$

(the reader is invited to confirm this and the claims that follow in Exercise 22.1). Thus, the unconditional probability that the process will be in regime 1 at any given date is given by

$$P\{s_t = 1\} = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}.$$

The unconditional probability that the process will be in regime 2, the second element of $\boldsymbol{\pi}$, is readily seen to be 1 minus this magnitude. The eigenvector associated with λ_2 is

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Thus, from [22.2.16], the matrix of m-period-ahead transition probabilities for an ergodic two-state Markov chain is given by

$$\begin{aligned}
\mathbf{P}^m &= \begin{bmatrix} \dfrac{1 - p_{22}}{2 - p_{11} - p_{22}} & -1 \\ \dfrac{1 - p_{11}}{2 - p_{11} - p_{22}} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \lambda_2^m \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \dfrac{-(1 - p_{11})}{2 - p_{11} - p_{22}} & \dfrac{1 - p_{22}}{2 - p_{11} - p_{22}} \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{(1 - p_{22}) + \lambda_2^m(1 - p_{11})}{2 - p_{11} - p_{22}} & \dfrac{(1 - p_{22}) - \lambda_2^m(1 - p_{22})}{(2 - p_{11} - p_{22})} \\ \dfrac{(1 - p_{11}) - \lambda_2^m(1 - p_{11})}{2 - p_{11} - p_{22}} & \dfrac{(1 - p_{11}) + \lambda_2^m(1 - p_{22})}{2 - p_{11} - p_{22}} \end{bmatrix}.
\end{aligned}$$

Thus, for example if the process is currently in state 1, the probability that m periods later it will be in state 2 is given by

$$P\{s_{t+m} = 2 | s_t = 1\} = \frac{(1 - p_{11}) - \lambda_2^m(1 - p_{11})}{2 - p_{11} - p_{22}},$$

where $\lambda_2 = -1 + p_{11} + p_{22}$.

A two-state Markov chain can also be represented by a simple scalar $AR(1)$ process, as follows. Let ξ_{1t} denote the first element of the vector $\boldsymbol{\xi}_t$; that is, ξ_{1t} is a random variable that is equal to unity when $s_t = 1$ and equal to zero otherwise. For the two-state chain, the second element of $\boldsymbol{\xi}_t$ is then $1 - \xi_{1t}$. Hence, [22.2.6] can be written

$$\begin{bmatrix} \xi_{1,t+1} \\ 1 - \xi_{1,t+1} \end{bmatrix} = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix} \begin{bmatrix} \xi_{1t} \\ 1 - \xi_{1t} \end{bmatrix} + \begin{bmatrix} v_{1,t+1} \\ v_{2,t+1} \end{bmatrix}. \qquad [22.2.23]$$

The first row of [22.2.23] states that

$$\xi_{1,t+1} = (1 - p_{22}) + (-1 + p_{11} + p_{22})\xi_{1t} + v_{1,t+1}. \qquad [22.2.24]$$

Expression [22.2.24] will be recognized as an $AR(1)$ process with constant term $(1 - p_{22})$ and autoregressive coefficient equal to $(-1 + p_{11} + p_{22})$. Note that this autoregressive coefficient turns out to be the second eigenvalue λ_2 of $\mathbf{P}$ calculated previously. When $p_{11} + p_{22} > 1$, the process is likely to persist in its current state and the variable ξ_{1t} would be positively serially correlated, whereas when $p_{11} + p_{22} < 1$, the process is more likely to switch out of a state than stay in it, producing negative serial correlation. Recall further from equation [3.4.3] that the mean of a first-order autoregression is given by $c/(1 - \phi)$. Hence, the representation [22.2.24] implies that

$$E(\xi_{1t}) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}},$$

which reproduces the earlier calculation of the value for the ergodic probability π_1.

Calculating Ergodic Probabilities for an N-state Markov Chain

For a general ergodic N-state process, the vector of unconditional probabilities represents a vector $\boldsymbol{\pi}$ with the properties that $\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\pi}$ and $\mathbf{1}'\boldsymbol{\pi} = 1$, where $\mathbf{1}$ denotes an $(N \times 1)$ vector of 1s. We thus seek a vector $\boldsymbol{\pi}$ satisfying

$$\mathbf{A}\boldsymbol{\pi} = \mathbf{e}_{N+1}. \qquad [22.2.25]$$

where $\mathbf{e}_{N+1}$ denotes the $(N + 1)$th column of $\mathbf{I}_{N+1}$ and where

$$\underset{(N+1)\times N}{\mathbf{A}} = \begin{bmatrix} \mathbf{I}_N - \mathbf{P} \\ \mathbf{1}' \end{bmatrix}.$$

Such a solution can be found by premultiplying [22.2.25] by $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$:

$$\boldsymbol{\pi} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{e}_{N+1}. \qquad [22.2.26]$$

In other words, $\boldsymbol{\pi}$ is the $(N + 1)$th column of the matrix $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$.

Periodic Markov Chains

If a Markov chain is irreducible, then there is one and only one eigenvalue equal to unity. However, there may be more than one eigenvalue on the unit circle, meaning that not all irreducible Markov chains are ergodic. For example, consider a two-state Markov chain in which $p_{11} = p_{22} = 0$:

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

The eigenvalues of this transition matrix are $\lambda_1 = 1$ and $\lambda_2 = -1$, both of which are on the unit circle. Thus, the matrix $\mathbf{P}^m$ does not converge to any fixed limit of the form $\boldsymbol{\pi}\cdot\mathbf{1}'$ for this case. Instead, if the process is in state 1 at date t, then it is certain to be there again for dates $t + 2$, $t + 4$, $t + 6$, . . . , with no tendency to converge as $m \rightarrow \infty$. Such a Markov chain is said to be *periodic* with period 2.

In general, it is possible to show that for any irreducible N-state Markov chain, all the eigenvalues of the transition matrix will be on or inside the unit circle. If there are K eigenvalues strictly on the unit circle with $K > 1$, then the chain is said to be periodic with period K. Such chains have the property that the states can be classified into K distinct classes, such that if the state at date t is from class α, then the state at date $t + 1$ is certain to be from class $\alpha + 1$ (where class $\alpha + 1$ for $\alpha = K$ is interpreted to be class 1). Thus, there is a zero probability of returning to the original state s_t, and indeed zero probability of returning to any member of the original class α, except at horizons that are integer multiples of the period (such as dates $t + K$, $t + 2K$, $t + 3K$, and so on). For further discussion of periodic Markov chains, see Cox and Miller (1965).

22.3. *Statistical Analysis of i.i.d. Mixture Distributions*

In Section 22.4, we will consider autoregressive processes in which the parameters of the autoregression can change as the result of a regime-shift variable. The regime itself will be described as the outcome of an unobserved Markov chain. Before analyzing such processes, it is instructive first to consider a special case of these processes known as i.i.d. *mixture distributions*.

Let the regime that a given process is in at date t be indexed by an unobserved random variable s_t, where there are N possible regimes ($s_t = 1, 2, \ldots,$ or N). When the process is in regime 1, the observed variable y_t is presumed to have been drawn from a $N(\mu_1, \sigma_1^2)$ distribution. If the process is in regime 2, then y_t is drawn from a $N(\mu_2, \sigma_2^2)$ distribution, and so on. Hence, the density of y_t conditional on the random variable s_t taking on the value j is

$$f(y_t|s_t = j; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_j}\exp\left\{\frac{-(y_t - \mu_j)^2}{2\sigma_j^2}\right\} \qquad [22.3.1]$$

for $j = 1, 2, \ldots, N$. Here $\boldsymbol{\theta}$ is a vector of population parameters that includes $\mu_1, \ldots, \mu_N$ and $\sigma_1^2, \ldots, \sigma_N^2$.

The unobserved regime $\{s_t\}$ is presumed to have been generated by some probability distribution, for which the unconditional probability that s_t takes on the value j is denoted π_j:

$$P\{s_t = j; \boldsymbol{\theta}\} = \pi_j \qquad \text{for } j = 1, 2, \ldots, N. \qquad [22.3.2]$$

The probabilities $\pi_1, \ldots, \pi_N$ are also included in $\boldsymbol{\theta}$; that is, $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta} \equiv (\mu_1, \ldots, \mu_N, \sigma_1^2, \ldots, \sigma_N^2, \pi_1, \ldots, \pi_N)'.$$

Recall that for any events A and B, the conditional probability of A given B is defined as

$$P\{A|B\} = \frac{P\{A \text{ and } B\}}{P\{B\}},$$

assuming that the probability that event B occurs is not zero. This expression implies that the joint probability of A and B occurring together can be calculated as

$$P\{A \text{ and } B\} = P\{A|B\} \cdot P\{B\}.$$

For example, if we were interested in the probability of the joint event that $s_t = j$ and that y_t falls within some interval $[c, d]$, this could be found by integrating

$$p(y_t, s_t = j; \boldsymbol{\theta}) = f(y_t|s_t = j; \boldsymbol{\theta}) \cdot P\{s_t = j; \boldsymbol{\theta}\} \qquad [22.3.3]$$

over all values of y_t between c and d. Expression [22.3.3] will be called the *joint density-distribution function* of y_t and s_t. From [22.3.1] and [22.3.2], this function is given by

$$p(y_t, s_t = j; \boldsymbol{\theta}) = \frac{\pi_j}{\sqrt{2\pi}\sigma_j} \exp\left\{\frac{-(y_t - \mu_j)^2}{2\sigma_j^2}\right\}. \qquad [22.3.4]$$

The unconditional density of y_t can be found by summing [22.3.4] over all possible values for j:

$$\begin{aligned} f(y_t; \boldsymbol{\theta}) &= \sum_{j=1}^{N} p(y_t, s_t = j; \boldsymbol{\theta}) \\ &= \frac{\pi_1}{\sqrt{2\pi}\sigma_1} \exp\left\{\frac{-(y_t - \mu_1)^2}{2\sigma_1^2}\right\} \\ &\quad + \frac{\pi_2}{\sqrt{2\pi}\sigma_2} \exp\left\{\frac{-(y_t - \mu_2)^2}{2\sigma_2^2}\right\} + \cdots \\ &\quad + \frac{\pi_N}{\sqrt{2\pi}\sigma_N} \exp\left\{\frac{-(y_t - \mu_N)^2}{2\sigma_N^2}\right\}. \end{aligned} \qquad [22.3.5]$$

Since the regime s_t is unobserved, expression [22.3.5] is the relevant density describing the actually observed data y_t. If the regime variable s_t is distributed i.i.d. across different dates t, then the log likelihood for the observed data can be calculated from [22.3.5] as

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(y_t; \boldsymbol{\theta}). \qquad [22.3.6]$$

The maximum likelihood estimate of $\boldsymbol{\theta}$ is obtained by maximizing [22.3.6] subject to the constraints that $\pi_1 + \pi_2 + \cdots + \pi_N = 1$ and $\pi_j \geq 0$ for $j = 1, 2, \ldots, N$. This can be achieved using the numerical methods described in Section 5.7, or using the *EM* algorithm developed later in this section.

Functions of the form of [22.3.5] can be used to represent a broad class of different densities. Figure 22.2 gives an example for $N = 2$. The joint density-distribution $p(y_t, s_t = 1; \boldsymbol{\theta})$ is π_1 times a $N(\mu_1, \sigma_1^2)$ density, while $p(y_t, s_t = 2; \boldsymbol{\theta})$ is π_2 times a $N(\mu_2, \sigma_2^2)$ density. The unconditional density for the observed variable $f(y_t; \boldsymbol{\theta})$ is the sum of these two magnitudes.

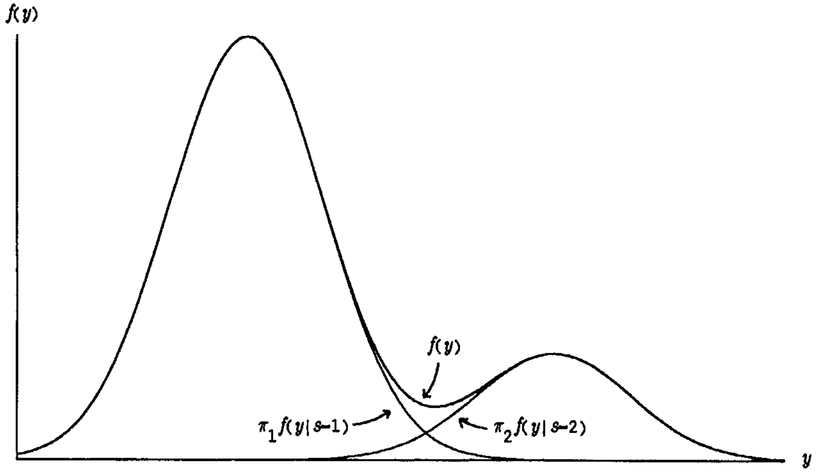


FIGURE 22.2 Density of mixture of two Gaussian distributions with $y_t|s_t = 1 \sim N(0, 1)$, $y_t|s_t = 2 \sim N(4, 1)$, and $P\{s_t = 1\} = 0.8$.

A mixture of two Gaussian variables need not have the bimodal appearance of Figure 22.2. Gaussian mixtures can also produce a unimodal density, allowing skew or kurtosis different from that of a single Gaussian variable, as in Figure 22.3.

Inference About the Unobserved Regime

Once one has obtained estimates of $\boldsymbol{\theta}$, it is possible to make an inference about which regime was more likely to have been responsible for producing the

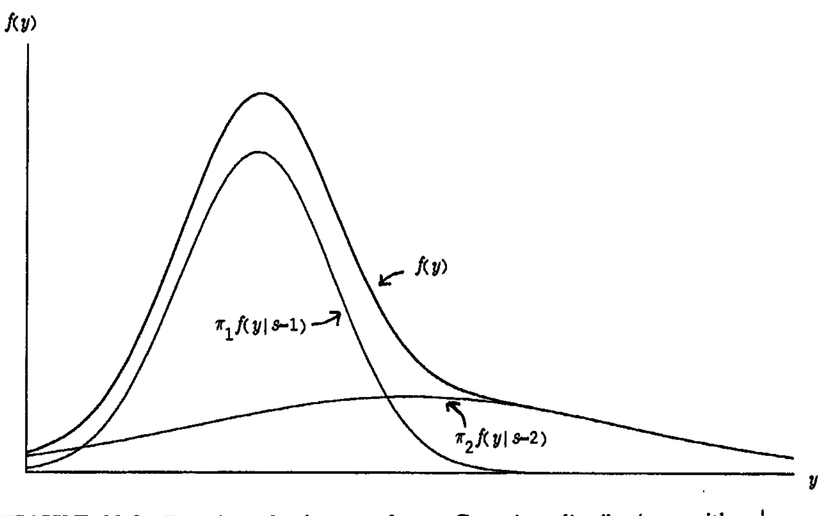


FIGURE 22.3 Density of mixture of two Gaussian distributions with $y_t|s_t = 1 \sim N(0, 1)$, $y_t|s_t = 2 \sim N(2, 8)$, and $P\{s_t = 1\} = 0.6$.

date t observation of y_t. Again, from the definition of a conditional probability, it follows that

$$P\{s_t = j|y_t; \boldsymbol{\theta}\} = \frac{p(y_t, s_t = j; \boldsymbol{\theta})}{f(y_t; \boldsymbol{\theta})} = \frac{\pi_j \cdot f(y_t|s_t = j; \boldsymbol{\theta})}{f(y_t; \boldsymbol{\theta})}. \qquad [22.3.7]$$

Given knowledge of the population parameters $\boldsymbol{\theta}$, it would be possible to use [22.3.1] and [22.3.5] to calculate the magnitude in [22.3.7] for each observation y_t in the sample. This number represents the probability, given the observed data, that the unobserved regime responsible for observation t was regime j. For example, for the mixture represented in Figure 22.2, if an observation y_t were equal to zero, one could be virtually certain that the observation had come from a $N(0, 1)$ distribution rather than a $N(4, 1)$ distribution, so that $P\{s_t = 1|y_t; \boldsymbol{\theta}\}$ for that date would be near unity. If instead y_t were around 2.3, it is equally likely that the observation might have come from either regime, so that $P\{s_t = 1|y_t; \boldsymbol{\theta}\}$ for such an observation would be close to 0.5.

Maximum Likelihood Estimates and the EM *Algorithm*

It is instructive to characterize analytically the maximum likelihood estimates of the population parameter $\boldsymbol{\theta}$. Appendix 22.A demonstrates that the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ represents a solution to the following system of nonlinear equations:

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} y_t \cdot P\{s_t = j|y_t; \hat{\boldsymbol{\theta}}\}}{\sum_{t=1}^{T} P\{s_t = j|y_t; \hat{\boldsymbol{\theta}}\}} \qquad \text{for } j = 1, 2, \ldots, N \qquad [22.3.8]$$

$$\hat{\sigma}_j^2 = \frac{\sum_{t=1}^{T} (y_t - \hat{\mu}_j)^2 \cdot P\{s_t = j|y_t; \hat{\boldsymbol{\theta}}\}}{\sum_{t=1}^{T} P\{s_t = j|y_t; \hat{\boldsymbol{\theta}}\}} \qquad \text{for } j = 1, 2, \ldots, N \qquad [22.3.9]$$

$$\hat{\pi}_j = T^{-1} \sum_{t=1}^{T} P\{s_t = j|y_t; \hat{\boldsymbol{\theta}}\} \qquad \text{for } j = 1, 2, \ldots, N. \qquad [22.3.10]$$

Suppose we were virtually certain which observations came from regime j and which did not, so that $P\{s_t = j|y_t; \boldsymbol{\theta}\}$ equaled unity for those observations that came from regime j and equaled zero for those observations that came from other regimes. Then the estimate of the mean for regime j in [22.3.8] would simply be the average value of y_t for those observations known to have come from regime j. In the more general case where $P\{s_t = j|y_t; \boldsymbol{\theta}\}$ is between 0 and 1 for some observations, the estimate $\hat{\mu}_j$ is a weighted average of all the observations in the sample, where the weight for observation y_t is proportional to the probability that date t's observation was generated by regime j. The more likely an observation is to have come from regime j, the bigger the weight given that observation in estimating μ_j. Similarly, $\hat{\sigma}_j^2$ is a weighted average of the squared deviations of y_t from $\hat{\mu}_j$, while $\hat{\pi}_j$ is essentially the fraction of observations that appear to have come from regime j.

Because equations [22.3.8] to [22.3.10] are nonlinear, it is not possible to solve them analytically for $\hat{\boldsymbol{\theta}}$ as a function of $\{y_1, y_2, \ldots, y_T\}$. However, these equations do suggest an appealing iterative algorithm for finding the maximum likelihood estimate. Starting from an arbitrary initial guess for the value of $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}^{(0)}$, one could calculate $P\{s_t = j|y_t; \boldsymbol{\theta}^{(0)}\}$ from [22.3.7]. One could then calculate the magnitudes on the right sides of [22.3.8] through [22.3.10] with $\boldsymbol{\theta}^{(0)}$ in place of $\hat{\boldsymbol{\theta}}$. The left sides of [22.3.8] through [22.3.10] would then produce a new estimate $\boldsymbol{\theta}^{(1)}$. This estimate $\boldsymbol{\theta}^{(1)}$ could be used to reevaluate $P\{s_t = j|y_t; \boldsymbol{\theta}^{(1)}\}$ and recalculate the expressions on the right sides of [22.3.8] through [22.3.10]. The left sides of [22.3.8] through [22.3.10] then can produce a new estimate $\boldsymbol{\theta}^{(2)}$. One continues iterating in this fashion until the change between $\boldsymbol{\theta}^{(m+1)}$ and $\boldsymbol{\theta}^{(m)}$ is smaller than some specified convergence criterion.

This algorithm turns out to be a special case of the *EM* principle developed by Dempster, Laird, and Rubin (1977). One can show that each iteration on this algorithm increases the value of the likelihood function. Clearly, if the iterations reach a point such that $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^{(m+1)}$, the algorithm has found the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$.

Further Discussion

The mixture density [22.3.5] has the property that a global maximum of the log likelihood [22.3.6] does not exist. A singularity arises whenever one of the distributions is imputed to have a mean exactly equal to one of the observations ($\mu_1 = y_1$, say) with no variance ($\sigma_1^2 \to 0$). At such a point the log likelihood becomes infinite.

Such singularities do not pose a major problem in practice, since numerical maximization procedures typically converge to a reasonable local maximum rather than a singularity. The largest local maximum with $\sigma_j > 0$ for all j is described as the *maximum likelihood estimate*. Kiefer (1978) showed that there exists a bounded local maximum of [22.3.6] that yields a consistent, asymptotically Gaussian estimate of $\boldsymbol{\theta}$ for which standard errors can be constructed using the usual formulas such as expression [5.8.3]. Hence, if a numerical maximization algorithm becomes stuck at a singularity, one satisfactory solution is simply to ignore the singularity and try again with different starting values.

Another approach is to maximize a slightly different objective function such as

$$Q(\boldsymbol{\theta}) = \mathscr{L}(\boldsymbol{\theta}) - \sum_{j=1}^{N} (a_j/2)\log(\sigma_j^2) - \sum_{j=1}^{N} b_j/(2\sigma_j^2) - \sum_{j=1}^{N} c_j(m_j - \mu_j)^2/(2\sigma_j^2), \qquad [22.3.11]$$

where $\mathscr{L}(\boldsymbol{\theta})$ is the log likelihood function described in [22.3.6]. If $a_j = c_j$, then expression [22.3.11] is the form the log likelihood would take if, in addition to the data, the analyst had a_j observations from regime j whose sample mean was m_j and whose sample variance was b_j/a_j. Thus, m_j represents the analyst's prior expectation of the value of μ_j, and b_j/a_j represents the analyst's prior expectation of the value of σ_j^2. The parameters a_j and c_j represent the strength of these priors, measured in terms of the confidence one would have if the priors were based on a_j or c_j direct observations of data known to have come from regime j. See Hamilton (1991) for further discussion of this approach.

Nice surveys of i.i.d. mixture distributions have been provided by Everitt and Hand (1981) and Titterington, Smith, and Makov (1985).

Description of the Process

We now return to the objective of developing a model that allows a given variable to follow a different time series process over different subsamples. As an illustration, consider a first-order autoregression in which both the constant term and the autoregressive coefficient might be different for different subsamples:

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t, \qquad [22.4.1]$$

where $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$. The proposal will be to model the regime s_t as the outcome of an unobserved N-state Markov chain with s_t independent of ε_τ for all t and τ.

Why might a Markov chain be a useful description of the process generating changes in regime? One's first thought could be that a change in regime such as that in Figure 22.1 is a permanent event. Such a permanent regime change could be modeled with a two-state Markov chain in which state 2 is an absorbing state. The advantage of using a Markov chain over a deterministic specification for such a process is that it allows one to generate meaningful forecasts prior to the change that take into account the possibility of the change from regime 1 to regime 2.

We might also want a time series model of changes in regime to account for unusual short-lived events such as World War II. Again, it is possible to choose parameters for a Markov chain such that, given 100 years of data, it is quite likely that we would have observed a single episode of regime 2 lasting for about 5 years. A Markov chain specification, of course, implies that given another 100 years we could well see another such event. One might argue that this is a sensible property to build into a model. The essence of the scientific method is the presumption that the future will in some sense be like the past.

While the Markov chain can describe such examples of changes in regime, a further advantage is its flexibility. There seems some value in specifying a probability law consistent with a broad range of different outcomes, and choosing particular parameters within that class on the basis of the data alone.

In any case, the approach described here readily generalizes to processes in which the probability that $s_t = j$ depends not only on the value of s_{t-1} but also on a vector of other observed variables—see Filardo (1992) and Diebold, Lee, and Weinbach (forthcoming).

The general model investigated in this section is as follows. Let $\mathbf{y}_t$ be an $(n \times 1)$ vector of observed endogenous variables and $\mathbf{x}_t$ a $(k \times 1)$ vector of observed exogenous variables. Let $\mathcal{Y}_t = (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{-m}', \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_{-m}')'$ be a vector containing all observations obtained through date t. If the process is governed by regime $s_t = j$ at date t, then the conditional density of $\mathbf{y}_t$ is assumed to be given by

$$f(\mathbf{y}_t | s_t = j, \mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\alpha}), \qquad [22.4.2]$$

where $\boldsymbol{\alpha}$ is a vector of parameters characterizing the conditional density. If there are N different regimes, then there are N different densities represented by [22.4.2] for $j = 1, 2, \ldots, N$. These densities will be collected in an $(N \times 1)$ vector denoted $\boldsymbol{\eta}_t$.

For the example of [22.4.1], y_t is a scalar ($n = 1$), the exogenous variables consist only of a constant term ($\mathbf{x}_t = 1$), and the unknown parameters in $\boldsymbol{\alpha}$ consist of $c_1, \ldots, c_N, \phi_1, \ldots, \phi_N$, and σ^2. With $N = 2$ regimes the two densities represented by [22.4.2] are

$$\boldsymbol{\eta}_t = \begin{bmatrix} f(y_t|s_t = 1, y_{t-1}; \boldsymbol{\alpha}) \\ f(y_t|s_t = 2, y_{t-1}; \boldsymbol{\alpha}) \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{\dfrac{-(y_t - c_1 - \phi_1 y_{t-1})^2}{2\sigma^2}\right\} \\ \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{\dfrac{-(y_t - c_2 - \phi_2 y_{t-1})^2}{2\sigma^2}\right\} \end{bmatrix}.$$

It is assumed in [22.4.2] that the conditional density depends only on the current regime s_t and not on past regimes:

$$\begin{aligned} &f(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}, s_t = j; \boldsymbol{\alpha}) \\ &\qquad = f(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}, s_t = j, s_{t-1} = i, s_{t-2} = k, \ldots; \boldsymbol{\alpha}), \end{aligned} \qquad [22.4.3]$$

though this is not really restrictive. Consider, for example, the specification in [22.1.3], where the conditional density of y_t depends on both s_t^* and s_{t-1}^* and where s_t^* is described by a two-state Markov chain. One can define a new variable s_t that characterizes the regime for date t in a way consistent with [22.4.2] as follows:

$$\begin{aligned} s_t &= 1 \quad \text{if } s_t^* = 1 \text{ and } s_{t-1}^* = 1 \\ s_t &= 2 \quad \text{if } s_t^* = 2 \text{ and } s_{t-1}^* = 1 \\ s_t &= 3 \quad \text{if } s_t^* = 1 \text{ and } s_{t-1}^* = 2 \\ s_t &= 4 \quad \text{if } s_t^* = 2 \text{ and } s_{t-1}^* = 2. \end{aligned}$$

If p_{ij}^* denotes $P\{s_t^* = j|s_{t-1}^* = i\}$, then s_t follows a four-state Markov chain with transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11}^* & 0 & p_{11}^* & 0 \\ p_{12}^* & 0 & p_{12}^* & 0 \\ 0 & p_{21}^* & 0 & p_{21}^* \\ 0 & p_{22}^* & 0 & p_{22}^* \end{bmatrix}.$$

Hence, [22.1.3] could be represented as a special case of this framework with $N = 4$, $\boldsymbol{\alpha} = (\mu_1, \mu_2, \phi, \sigma^2)'$ and with [22.4.2] representing the four densities

$$f(y_t|y_{t-1}, s_t = 1; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-[(y_t - \mu_1) - \phi(y_{t-1} - \mu_1)]^2}{2\sigma^2}\right\}$$

$$f(y_t|y_{t-1}, s_t = 2; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-[(y_t - \mu_2) - \phi(y_{t-1} - \mu_1)]^2}{2\sigma^2}\right\}$$

$$f(y_t|y_{t-1}, s_t = 3; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-[(y_t - \mu_1) - \phi(y_{t-1} - \mu_2)]^2}{2\sigma^2}\right\}$$

$$f(y_t|y_{t-1}, s_t = 4; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-[(y_t - \mu_2) - \phi(y_{t-1} - \mu_2)]^2}{2\sigma^2}\right\}.$$

It is assumed that s_t evolves according to a Markov chain that is independent of past observations on $\mathbf{y}_t$ or current or past $\mathbf{x}_t$:

$$P\{s_t = j|s_{t-1} = i, s_{t-2} = k, \ldots, \mathbf{x}_t, \mathcal{Y}_{t-1}\} = P\{s_t = j|s_{t-1} = i\} = p_{ij}. \qquad [22.4.4]$$

For generalizations of this assumption, see Lam (1990), Durland and McCurdy (1992), Filardo (1992), and Diebold, Lee, and Weinbach (forthcoming).

Optimal Inference About Regimes and Evaluation of the Likelihood Function

The population parameters that describe a time series governed by [22.4.2] and [22.4.4] consist of $\boldsymbol{\alpha}$ and the various transition probabilities p_{ij}. Collect these parameters in a vector $\boldsymbol{\theta}$. One important objective will be to estimate the value of $\boldsymbol{\theta}$ based on observation of $\mathcal{Y}_T$. Let us nevertheless put this objective on hold for the moment and suppose that the value of $\boldsymbol{\theta}$ is somehow known with certainty to the analyst. Even if we know the value of $\boldsymbol{\theta}$, we will not know which regime the process was in at every date in the sample. Instead the best we can do is to form a probabilistic inference that is a generalization of [22.3.7]. In the i.i.d. case, the analyst's inference about the value of s_t depends only on the value of y_t. In the more general class of time series models described here the inference typically depends on all the observations available.

Let $P\{s_t = j | \mathcal{Y}_t; \boldsymbol{\theta}\}$ denote the analyst's inference about the value of s_t based on data obtained through date t and based on knowledge of the population parameters $\boldsymbol{\theta}$. This inference takes the form of a conditional probability that the analyst assigns to the possibility that the tth observation was generated by regime j. Collect these conditional probabilities $P\{s_t = j | \mathcal{Y}_t; \boldsymbol{\theta}\}$ for $j = 1, 2, \ldots, N$ in an $(N \times 1)$ vector denoted $\hat{\boldsymbol{\xi}}_{t|t}$.

One could also imagine forming forecasts of how likely the process is to be in regime j in period $t + 1$ given observations obtained through date t. Collect these forecasts in an $(N \times 1)$ vector $\hat{\boldsymbol{\xi}}_{t+1|t}$, which is a vector whose jth element represents $P\{s_{t+1} = j | \mathcal{Y}_t; \boldsymbol{\theta}\}$.

The optimal inference and forecast for each date t in the sample can be found by iterating on the following pair of equations:

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)}{\mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)} \qquad [22.4.5]$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{P} \cdot \hat{\boldsymbol{\xi}}_{t|t}. \qquad [22.4.6]$$

Here $\boldsymbol{\eta}_t$ represents the $(N \times 1)$ vector whose jth element is the conditional density in [22.4.2], $\mathbf{P}$ represents the $(N \times N)$ transition matrix defined in [22.2.3], $\mathbf{1}$ represents an $(N \times 1)$ vector of 1s, and the symbol $\odot$ denotes element-by-element multiplication. Given a starting value $\hat{\boldsymbol{\xi}}_{1|0}$ and an assumed value for the population parameter vector $\boldsymbol{\theta}$, one can iterate on [22.4.5] and [22.4.6] for $t = 1, 2, \ldots, T$ to calculate the values of $\hat{\boldsymbol{\xi}}_{t|t}$ and $\hat{\boldsymbol{\xi}}_{t+1|t}$ for each date t in the sample. The log likelihood function $\mathcal{L}(\boldsymbol{\theta})$ for the observed data $\mathcal{Y}_T$ evaluated at the value of $\boldsymbol{\theta}$ that was used to perform the iterations can also be calculated as a by-product of this algorithm from

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t | \mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}), \qquad [22.4.7]$$

where

$$f(\mathbf{y}_t | \mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}) = \mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t). \qquad [22.4.8]$$

We now explain why this algorithm works.

Derivation of Equations [22.4.5] Through [22.4.8]

To see the basis for the algorithm just described, note that we have assumed that $\mathbf{x}_t$ is exogenous, by which we mean that $\mathbf{x}_t$ contains no information about s_t

beyond that contained in $\mathcal{Y}_{t-1}$. Hence, the jth element of $\hat{\boldsymbol{\xi}}_{t|t-1}$ could also be described as $P\{s_t = j|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}\}$. The jth element of $\boldsymbol{\eta}_t$ is $f(\mathbf{y}_t|s_t = j, \mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})$. The jth element of the $(N \times 1)$ vector $(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)$ is the product of these two magnitudes, which product can be interpreted as the conditional joint density-distribution of $\mathbf{y}_t$ and s_t:

$$P\{s_t = j|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}\} \times f(\mathbf{y}_t|s_t = j, \mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}) \qquad [22.4.9]$$
$$= p(\mathbf{y}_t, s_t = j|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}).$$

The density of the observed vector $\mathbf{y}_t$ conditioned on past observables is the sum of the N magnitudes in [22.4.9] for $j = 1, 2, \ldots, N$. This sum can be written in vector notation as

$$f(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}) = \mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t),$$

as claimed in [22.4.8]. If the joint density-distribution in [22.4.9] is divided by the density of $\mathbf{y}_t$ in [22.4.8], the result is the conditional distribution of s_t:

$$\frac{p(\mathbf{y}_t, s_t = j|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{f(\mathbf{y}_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})} = P\{s_t = j|\mathbf{y}_t, \mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}\}$$
$$= P\{s_t = j|\mathcal{Y}_t; \boldsymbol{\theta}\}.$$

Hence, from [22.4.8],

$$P\{s_t = j|\mathcal{Y}_t; \boldsymbol{\theta}\} = \frac{p(\mathbf{y}_t, s_t = j|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)}. \qquad [22.4.10]$$

But recall from [22.4.9] that the numerator in the expression on the right side of [22.4.10] is the jth element of the vector $(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)$, while the left side of [22.4.10] is the jth element of the vector $\hat{\boldsymbol{\xi}}_{t|t}$. Thus, collecting the equations in [22.4.10] for $j = 1, 2, \ldots, N$ into an $(N \times 1)$ vector produces

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)}{\mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)},$$

as claimed in [22.4.5].

To see the basis for [22.4.6], take expectations of [22.2.6] conditional on $\mathcal{Y}_t$:

$$E(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_t) = \mathbf{P}\cdot E(\boldsymbol{\xi}_t|\mathcal{Y}_t) + E(\mathbf{v}_{t+1}|\mathcal{Y}_t). \qquad [22.4.11]$$

Note that $\mathbf{v}_{t+1}$ is a martingale difference sequence with respect to $\mathcal{Y}_t$, so that [22.4.11] becomes

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{P}\cdot\hat{\boldsymbol{\xi}}_{t|t},$$

as claimed in [22.4.6].

Starting the Algorithm

Given a starting value $\hat{\boldsymbol{\xi}}_{1|0}$, one can use [22.4.5] and [22.4.6] to calculate $\hat{\boldsymbol{\xi}}_{t|t}$ for any t. Several options are available for choosing the starting value. One approach is to set $\hat{\boldsymbol{\xi}}_{1|0}$ equal to the vector of unconditional probabilities $\boldsymbol{\pi}$ described in equation [22.2.26]. Another option is to set

$$\hat{\boldsymbol{\xi}}_{1|0} = \boldsymbol{\rho}, \qquad [22.4.12]$$

where $\boldsymbol{\rho}$ is a fixed $(N \times 1)$ vector of nonnegative constants summing to unity, such as $\boldsymbol{\rho} = N^{-1} \cdot \mathbf{1}$. Alternatively, $\boldsymbol{\rho}$ could be estimated by maximum likelihood along with $\boldsymbol{\theta}$ subject to the constraint that $\mathbf{1}'\boldsymbol{\rho} = 1$ and $\rho_j \geq 0$ for $j = 1, 2, \ldots, N$.

Forecasts and Smoothed Inferences for the Regime

Generalizing the earlier notation, let $\hat{\boldsymbol{\xi}}_{t|\tau}$ represent the $(N \times 1)$ vector whose jth element is $P\{s_t = j | \mathcal{Y}_\tau; \boldsymbol{\theta}\}$. For $t > \tau$, this represents a forecast about the regime for some future period, whereas for $t < \tau$ it represents the *smoothed inference* about the regime the process was in at date t based on data obtained through some later date τ.

The optimal m-period-ahead forecast of $\boldsymbol{\xi}_{t+m}$ can be found by taking expectations of both sides of [22.2.8] conditional on information available at date t:

$$E(\boldsymbol{\xi}_{t+m} | \mathcal{Y}_t) = \mathbf{P}^m \cdot E(\boldsymbol{\xi}_t | \mathcal{Y}_t)$$

or

$$\hat{\boldsymbol{\xi}}_{t+m|t} = \mathbf{P}^m \cdot \hat{\boldsymbol{\xi}}_{t|t}, \qquad [22.4.13]$$

where $\hat{\boldsymbol{\xi}}_{t|t}$ is calculated from [22.4.5].

Smoothed inferences can be calculated using an algorithm developed by Kim (1993). In vector form, this algorithm can be written as

$$\hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} \odot \{\mathbf{P}' \cdot [\hat{\boldsymbol{\xi}}_{t+1|T} (\div) \hat{\boldsymbol{\xi}}_{t+1|t}]\}, \qquad [22.4.14]$$

where the sign $(\div)$ denotes element-by-element division. The smoothed probabilities $\hat{\boldsymbol{\xi}}_{t|T}$ are found by iterating on [22.4.14] backward for $t = T - 1, T - 2, \ldots, 1$. This iteration is started with $\hat{\boldsymbol{\xi}}_{T|T}$, which is obtained from [22.4.5] for $t = T$. This algorithm is valid only when s_t follows a first-order Markov chain as in [22.4.4], when the conditional density [22.4.2] depends on $s_t, s_{t-1}, \ldots$ only through the current state s_t, and when $\mathbf{x}_t$, the vector of explanatory variables other than the lagged values of $\mathbf{y}$, is strictly exogenous, meaning that $\mathbf{x}_t$ is independent of s_τ for all t and τ. The basis for Kim's algorithm is explained in Appendix 22.A at the end of the chapter.

Forecasts for the Observed Variables

From the conditional density [22.4.2] it is straightforward to forecast $\mathbf{y}_{t+1}$ conditional on knowing $\mathcal{Y}_t$, $\mathbf{x}_{t+1}$, and s_{t+1}. For example, for the $AR(1)$ specification $y_{t+1} = c_{s_{t+1}} + \phi_{s_{t+1}} y_t + \varepsilon_{t+1}$, such a forecast is given by

$$E(y_{t+1} | s_{t+1} = j, \mathcal{Y}_t; \boldsymbol{\theta}) = c_j + \phi_j y_t. \qquad [22.4.15]$$

There are N different conditional forecasts associated with the N possible values for s_{t+1}. Note that the unconditional forecast based on actual observable variables is related to these conditional forecasts by

$$\begin{aligned} &E(\mathbf{y}_{t+1} | \mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}) \\ &\quad = \int \mathbf{y}_{t+1} \cdot f(\mathbf{y}_{t+1} | \mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta})\, d\mathbf{y}_{t+1} \\ &\quad = \int \mathbf{y}_{t+1} \left\{ \sum_{j=1}^{N} p(\mathbf{y}_{t+1}, s_{t+1} = j | \mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}) \right\} d\mathbf{y}_{t+1} \end{aligned}$$

$$= \int \mathbf{y}_{t+1} \left\{ \sum_{j=1}^{N} [f(\mathbf{y}_{t+1}|s_{t+1} = j, \mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}) P\{s_{t+1} = j|\mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\}] \right\} d\mathbf{y}_{t+1}$$

$$= \sum_{j=1}^{N} P\{s_{t+1} = j|\mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}\} \int \mathbf{y}_{t+1} \cdot f(\mathbf{y}_{t+1}|s_{t+1} = j, \mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta})\, d\mathbf{y}_{t+1}$$

$$= \sum_{j=1}^{N} P\{s_{t+1} = j|\mathcal{Y}_t; \boldsymbol{\theta}\} E(\mathbf{y}_{t+1}|s_{t+1} = j, \mathbf{x}_{t+1}, \mathcal{Y}_t; \boldsymbol{\theta}).$$

Thus, the forecast appropriate for the jth regime is simply multiplied by the probability that the process will be in the jth regime, and the resulting N different products are added together. For example, if the $j = 1, 2, \ldots, N$ forecasts in [22.4.15] are collected in a $(1 \times N)$ vector $\mathbf{h}_t'$, then

$$E(y_{t+1}|\mathcal{Y}_t; \boldsymbol{\theta}) = \mathbf{h}_t' \hat{\boldsymbol{\xi}}_{t+1|t}.$$

Note that although the Markov chain itself admits the linear representation [22.2.6], the optimal forecast of $\mathbf{y}_{t+1}$ is a nonlinear function of observables, since the inference $\hat{\boldsymbol{\xi}}_{t|t}$ in [22.4.5] depends nonlinearly on $\mathcal{Y}_t$. Although one may use a linear model to form forecasts within a given regime, if an observation seems unlikely to have been generated by the same regime as preceding observations, the appearance of the outlier causes the analyst to switch to a new rule for forming future linear forecasts.

The Markov chain is clearly well suited for forming multiperiod forecasts as well. See Hamilton (1989, 1993b, 1993c) for further discussion.

Maximum Likelihood Estimation of Parameters

In the iteration on [22.4.5] and [22.4.6], the parameter vector $\boldsymbol{\theta}$ was taken to be a fixed, known vector. Once the iteration has been completed for $t = 1, 2, \ldots, T$ for a given fixed $\boldsymbol{\theta}$, the value of the log likelihood implied by that value of $\boldsymbol{\theta}$ is then known from [22.4.7]. The value of $\boldsymbol{\theta}$ that maximizes the log likelihood can be found numerically using the methods described in Section 5.7.

If the transition probabilities are restricted only by the conditions that $p_{ij} \geq 0$ and $(p_{i1} + p_{i2} + \cdots + p_{iN}) = 1$ for all i and j, and if the initial probability $\hat{\boldsymbol{\xi}}_{1|0}$ is taken to be a fixed value $\boldsymbol{\rho}$ unrelated to the other parameters, then it is shown in Hamilton (1990) that the maximum likelihood estimates for the transition probabilities satisfy

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} P\{s_t = j, s_{t-1} = i|\mathcal{Y}_T; \hat{\boldsymbol{\theta}}\}}{\sum_{t=2}^{T} P\{s_{t-1} = i|\mathcal{Y}_T; \hat{\boldsymbol{\theta}}\}}, \qquad [22.4.16]$$

where $\hat{\boldsymbol{\theta}}$ denotes the full vector of maximum likelihood estimates. Thus, the estimated transition probability $\hat{p}_{ij}$ is essentially the number of times state i seems to have been followed by state j divided by the number of times the process was in state i. These counts are estimated on the basis of the smoothed probabilities.

If the vector of initial probabilities $\boldsymbol{\rho}$ is regarded as a separate vector of parameters constrained only by $\mathbf{1}'\boldsymbol{\rho} = 1$ and $\boldsymbol{\rho} \geq \mathbf{0}$, the maximum likelihood estimate of $\boldsymbol{\rho}$ turns out to be the smoothed inference about the initial state:

$$\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\xi}}_{1|T}. \qquad [22.4.17]$$

The maximum likelihood estimate of the vector $\boldsymbol{\alpha}$ that governs the conditional density [22.4.2] is characterized by

$$\sum_{t=1}^{T}\left(\frac{\partial \log \boldsymbol{\eta}_t}{\partial \boldsymbol{\alpha}'}\right)'\hat{\boldsymbol{\xi}}_{t|T} = \mathbf{0}. \qquad [22.4.18]$$

Here $\boldsymbol{\eta}_t$ is the $(N \times 1)$ vector obtained by vertically stacking the densities in [22.4.2] for $j = 1, 2, \ldots, N$ and $(\partial \log \boldsymbol{\eta}_t)/\partial\boldsymbol{\alpha}'$ is the $(N \times k)$ matrix of derivatives of the logs of these densities, where k represents the number of parameters in $\boldsymbol{\alpha}$. For example, consider a Markov-switching regression model of the form

$$y_t = \mathbf{z}_t'\boldsymbol{\beta}_{s_t} + \varepsilon_t, \qquad [22.4.19]$$

where $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$ and where $\mathbf{z}_t$ is a vector of explanatory variables that could include lagged values of y. The coefficient vector for this regression is $\boldsymbol{\beta}_1$ when the process is in regime 1, $\boldsymbol{\beta}_2$ when the process is in regime 2, and so on. For this example, the vector $\boldsymbol{\eta}_t$ would be

$$\boldsymbol{\eta}_t = \begin{bmatrix} \dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{\dfrac{-(y_t - \mathbf{z}_t'\boldsymbol{\beta}_1)^2}{2\sigma^2}\right\} \\ \vdots \\ \dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{\dfrac{-(y_t - \mathbf{z}_t'\boldsymbol{\beta}_N)^2}{2\sigma^2}\right\} \end{bmatrix},$$

and for $\boldsymbol{\alpha} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_N', \sigma^2)'$, condition [22.4.18] becomes

$$\sum_{t=1}^{T}(y_t - \mathbf{z}_t'\hat{\boldsymbol{\beta}}_j)\mathbf{z}_t \cdot P\{s_t = j|\mathcal{Y}_T; \hat{\boldsymbol{\theta}}\} = \mathbf{0} \qquad \text{for } j = 1, 2, \ldots, N \qquad [22.4.20]$$

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N}(y_t - \mathbf{z}_t'\hat{\boldsymbol{\beta}}_j)^2 \cdot P\{s_t = j|\mathcal{Y}_T; \hat{\boldsymbol{\theta}}\}. \qquad [22.4.21]$$

Equation [22.4.20] describes $\hat{\boldsymbol{\beta}}_j$ as satisfying a weighted *OLS* orthogonality condition where each observation is weighted by the probability that it came from regime j. In particular, the estimate $\hat{\boldsymbol{\beta}}_j$ can be found from an *OLS* regression of $\tilde{y}_t(j)$ on $\tilde{\mathbf{z}}_t(j)$:

$$\hat{\boldsymbol{\beta}}_j = \left[\sum_{t=1}^{T}[\tilde{\mathbf{z}}_t(j)][\tilde{\mathbf{z}}_t(j)]'\right]^{-1}\left[\sum_{t=1}^{T}[\tilde{\mathbf{z}}_t(j)]\tilde{y}_t(j)\right], \qquad [22.4.22]$$

where

$$\begin{aligned} \tilde{y}_t(j) &= y_t \cdot \sqrt{P\{s_t = j|\mathcal{Y}_T; \hat{\boldsymbol{\theta}}\}} \\ \tilde{\mathbf{z}}_t(j) &= \mathbf{z}_t \cdot \sqrt{P\{s_t = j|\mathcal{Y}_T; \hat{\boldsymbol{\theta}}\}}. \end{aligned} \qquad [22.4.23]$$

The estimate of σ^2 in [22.4.21] is just $(1/T)$ times the combined sum of the squared residuals from these N different regressions.

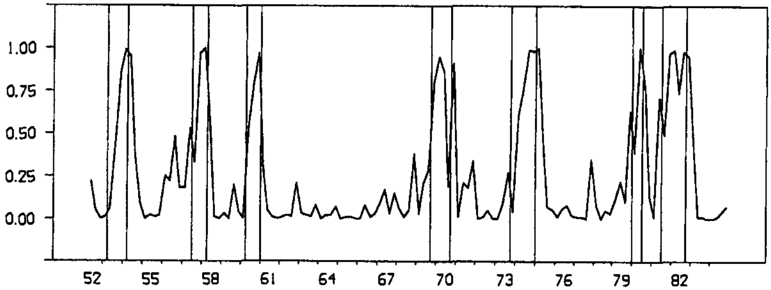
Again, this suggests an appealing algorithm for finding maximum likelihood estimates. For the case when $\boldsymbol{\rho}$ is fixed a priori, given an initial guess for the parameter vector $\boldsymbol{\theta}^{(0)}$ one could evaluate [22.4.16], [22.4.22], and [22.4.21] to generate a new estimate $\boldsymbol{\theta}^{(1)}$. One then iterates in the same fashion described in equations [22.3.8] through [22.3.10] to calculate $\boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(3)}, \ldots$. This again turns out to be an application of the *EM* algorithm. Alternatively, if $\boldsymbol{\rho}$ is to be estimated by maximum likelihood, equation [22.4.17] would be added to the equations that are reevaluated with each iteration. See Hamilton (1990) for details.

Illustration: The Behavior of U.S. Real GNP

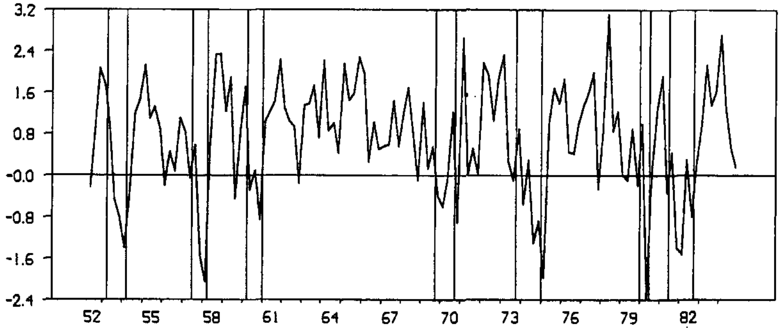
As an illustration of this method, consider the data on U.S. real GNP growth analyzed in Hamilton (1989). These data are plotted in the bottom panel of Figure 22.4. The following switching model was fitted to these data by maximum likelihood:

$$\begin{aligned} y_t - \mu_{s_t^*} = {} & \phi_1(y_{t-1} - \mu_{s_{t-1}^*}) + \phi_2(y_{t-2} - \mu_{s_{t-2}^*}) \\ & + \phi_3(y_{t-3} - \mu_{s_{t-3}^*}) + \phi_4(y_{t-4} - \mu_{s_{t-4}^*}) + \varepsilon_t, \end{aligned} \qquad [22.4.24]$$

with $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$ and with s_t^* presumed to follow a two-state Markov chain with transition probabilities p_{ij}^*. Maximum likelihood estimates of parameters are reported in Table 22.1. In the regime represented by $s_t^* = 1$, the average growth rate is $\mu_1 = 1.2\%$ per quarter, while when $s_t^* = 2$, the average growth rate is $\mu_2 = -0.4\%$. Each regime is highly persistent. The probability that expansion will be followed by another quarter of expansion is $p_{11}^* = 0.9$, so that this regime will persist on average for $1/(1 - p_{11}^*) = 10$ quarters. The probability that a contraction will be followed by contraction is $p_{22}^* = 0.75$, which episodes will typically persist for $1/(1 - p_{22}^*) = 4$ quarters.



(a) Probability that economy is in contraction state, or $P\{s_t^* = 2 | y_t, y_{t-1}, \ldots, y_{-4}; \hat{\boldsymbol{\theta}}\}$ plotted as a function of t.



(b) Quarterly rate of growth of U.S. real GNP, 1952–84.

FIGURE 22.4 Output growth and recession probabilities.

TABLE 22.1
Maximum Likelihood Estimates of Parameters for Markov-Switching Model of U.S. GNP (Standard Errors in Parentheses)

$\hat{\mu}_1 = 1.16$ (0.07)	$\hat{\mu}_2 = -0.36$ (0.26)	$\hat{p}_{11}^* = 0.90$ (0.04)	$\hat{p}_{22}^* = 0.75$ (0.10)	$\hat{\sigma}^2 = 0.59$ (0.10)
$\hat{\phi}_1 = 0.01$ (0.12)	$\hat{\phi}_2 = -0.06$ (0.14)	$\hat{\phi}_3 = -0.25$ (0.11)	$\hat{\phi}_4 = -0.21$ (0.11)	

In order to write [22.4.24] in a form where y_t depends only on the current value of the regime, a variable s_t was defined that takes on one of 32 different values representing the 32 possible combinations for $s_t^*, s_{t-1}^*, \ldots, s_{t-4}^*$. For example, $s_t = 1$ when $s_t^*, s_{t-1}^*, \ldots,$ and s_{t-4}^* all equal 1, $s_t = 2$ when $s_t^* = 2$ and $s_{t-1}^* = \cdots = s_{t-4}^* = 1$, and so on. The vector $\hat{\xi}_{t|t}$ calculated from [22.4.5] is thus a (32×1) vector that contains the probabilities of each of these 32 joint events conditional on data observed through date t.

The inference about the value of s_t^* for a single date t is obtained by summing together the relevant joint probabilities. For example, the inference

$$\begin{aligned} &P\{s_t^* = 2 | y_t, y_{t-1}, \ldots, y_{-4}; \hat{\boldsymbol{\theta}}\} \\ &= \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \sum_{i_3=1}^{2} \sum_{i_4=1}^{2} P\{s_t^* = 2, s_{t-1}^* = i_1, s_{t-2}^* = i_2, s_{t-3}^* = i_3, s_{t-4}^* = i_4 | y_t, \\ &\qquad y_{t-1}, \ldots, y_{-4}; \hat{\boldsymbol{\theta}}\} \end{aligned} \tag{22.4.25}$$

is obtained by iterating on [22.4.5] and [22.4.6] with $\boldsymbol{\theta}$ equal to the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. One then sums together the elements in the even-numbered rows of $\hat{\xi}_{t|t}$ to obtain $P\{s_t^* = 2 | y_t, y_{t-1}, \ldots, y_{-4}; \hat{\boldsymbol{\theta}}\}$.

A probabilistic inference in the form of [22.4.25] can be calculated for each date t in the sample. The resulting series is plotted as a function of t in panel (a) of Figure 22.4. The vertical lines in the figure indicate the dates at which economic recessions were determined to begin and end according to the National Bureau of Economic Research. These determinations are made informally on the basis of a large number of time series and are usually made some time after the event. Although these business cycle dates were not used in any way to estimate parameters or form inferences about s_t^*, it is interesting that the traditional business cycle dates correspond fairly closely to the expansion and contraction phases as described by the model in [22.4.24].

Determining the Number of States

One of the most important hypotheses that one would want to test for such models concerns the number of different regimes N that characterize the data. Unfortunately, this hypothesis cannot be tested using the usual likelihood ratio test. One of the regularity conditions for the likelihood ratio test to have an asymptotic χ^2 distribution is that the information matrix $\mathcal{I}$ be nonsingular. This condition fails to hold if the analyst tries to fit an N-state model when the true process has $N - 1$ states, since under the null hypothesis the parameters that describe the Nth state are unidentified. Tests that get around the problems with the regularity conditions have been proposed by Davies (1977), Hansen (1993), Andrews and Ploberger (1992), and Stinchcombe and White (1993). Another approach is to take the $(N-1)$-state model as the null and conduct a variety of tests of the validity of that specification as one way of seeing whether an N-state model is needed; Hamilton (1993a) proposed a number of such tests. Studies that illustrate the use of such tests include Engel and Hamilton (1990), Hansen (1992), and Goodwin (1993).

APPENDIX 22.A. *Derivation of Selected Equations for Chapter 22*

■ **Derivation of [22.3.8] through [22.3.10].** The maximum likelihood estimates are obtained by forming the Lagrangean

$$J(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) + \lambda(1 - \pi_1 - \pi_2 - \cdots - \pi_N) \qquad [22.A.1]$$

and setting the derivative with respect to $\boldsymbol{\theta}$ equal to zero. From [22.3.6], the derivative of the log likelihood is given by

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \frac{1}{f(y_t; \boldsymbol{\theta})} \times \frac{\partial f(y_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \qquad [22.A.2]$$

Observe from [22.3.5] that

$$\begin{aligned} \frac{\partial f(y_t; \boldsymbol{\theta})}{\partial \pi_j} &= \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{\frac{-(y_t - \mu_j)^2}{2\sigma_j^2}\right\} \\ &= f(y_t|s_t = j; \boldsymbol{\theta}), \end{aligned} \qquad [22.A.3]$$

while

$$\frac{\partial f(y_t; \boldsymbol{\theta})}{\partial \mu_j} = \frac{y_t - \mu_j}{\sigma_j^2} \times p(y_t, s_t = j; \boldsymbol{\theta}) \qquad [22.A.4]$$

and

$$\frac{\partial f(y_t; \boldsymbol{\theta})}{\partial \sigma_j^2} = \left\{-\frac{1}{2}\sigma_j^{-2} + \frac{(y_t - \mu_j)^2}{2\sigma_j^4}\right\} \times p(y_t, s_t = j; \boldsymbol{\theta}). \qquad [22.A.5]$$

Thus, [22.A.2] becomes

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \pi_j} = \sum_{t=1}^{T} \frac{1}{f(y_t; \boldsymbol{\theta})} f(y_t|s_t = j; \boldsymbol{\theta}) \qquad [22.A.6]$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \mu_j} = \sum_{t=1}^{T} \frac{1}{f(y_t; \boldsymbol{\theta})} \times \frac{y_t - \mu_j}{\sigma_j^2} p(y_t, s_t = j; \boldsymbol{\theta}) \qquad [22.A.7]$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma_j^2} = \sum_{t=1}^{T} \frac{1}{f(y_t; \boldsymbol{\theta})} \left\{-\frac{1}{2}\sigma_j^{-2} + \frac{(y_t - \mu_j)^2}{2\sigma_j^4}\right\} p(y_t, s_t = j; \boldsymbol{\theta}). \qquad [22.A.8]$$

Recalling [22.3.7], the derivatives in [22.A.6] through [22.A.8] can be written

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \pi_j} = \pi_j^{-1} \sum_{t=1}^{T} P\{s_t = j|y_t; \boldsymbol{\theta}\} \qquad [22.A.9]$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \mu_j} = \sum_{t=1}^{T} \frac{y_t - \mu_j}{\sigma_j^2} P\{s_t = j|y_t; \boldsymbol{\theta}\} \qquad [22.A.10]$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \sigma_j^2} = \sum_{t=1}^{T} \left\{-\frac{1}{2}\sigma_j^{-2} + \frac{(y_t - \mu_j)^2}{2\sigma_j^4}\right\} P\{s_t = j|y_t; \boldsymbol{\theta}\}. \qquad [22.A.11]$$

Setting the derivative of the Lagrangean in [22.A.1] with respect to μ_j equal to zero means setting [22.A.10] equal to zero, from which

$$\sum_{t=1}^{T} y_t \cdot P\{s_t = j|y_t; \boldsymbol{\theta}\} = \mu_j \sum_{t=1}^{T} P\{s_t = j|y_t; \boldsymbol{\theta}\}.$$

Equation [22.3.8] follows immediately from this condition. Similarly, the first-order conditions for maximization with respect to σ_j^2 are found by setting [22.A.11] equal to zero:

$$\sum_{t=1}^{T} \{-\sigma_j^2 + (y_t - \mu_j)^2\} P\{s_t = j | \mathcal{Y}_t; \boldsymbol{\theta}\} = 0,$$

from which [22.3.9] follows. Finally, from [22.A.9], the derivative of [22.A.1] with respect to π_j is given by

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \pi_j} = \pi_j^{-1} \sum_{t=1}^{T} P\{s_t = j | \mathbf{y}_t; \boldsymbol{\theta}\} - \lambda = 0,$$

from which

$$\sum_{t=1}^{T} P\{s_t = j | \mathbf{y}_t; \boldsymbol{\theta}\} = \lambda \pi_j. \qquad [22.A.12]$$

Summing [22.A.12] over $j = 1, 2, \ldots, N$ produces

$$\sum_{t=1}^{T} \left[P\{s_t = 1 | \mathbf{y}_t; \boldsymbol{\theta}\} + \cdots + P\{s_t = N | \mathbf{y}_t; \boldsymbol{\theta}\} \right] = \lambda(\pi_1 + \pi_2 + \cdots + \pi_N)$$

or

$$\sum_{t=1}^{T} \{1\} = \lambda \cdot (1),$$

implying that $T = \lambda$. Replacing λ with T in [22.A.12] produces [22.3.10]. ■

■ **Derivation of [22.4.14].** Recall first that under the maintained assumptions, the regime s_t depends on past observations $\mathcal{Y}_{t-1}$ only through the value of s_{t-1}. Similarly, s_t depends on future observations only through the value of s_{t+1}:

$$P\{s_t = j | s_{t+1} = i, \mathcal{Y}_T; \boldsymbol{\theta}\} = P\{s_t = j | s_{t+1} = i, \mathcal{Y}_t; \boldsymbol{\theta}\}. \qquad [22.A.13]$$

The validity of [22.A.13] is formally established as follows (the implicit dependence on $\boldsymbol{\theta}$ will be suppressed to simplify the notation). Observe that

$$\begin{aligned} &P\{s_t = j | s_{t+1} = i, \mathcal{Y}_{t+1}\} \\ &= P\{s_t = j | s_{t+1} = i, \mathbf{y}_{t+1}, \mathbf{x}_{t+1}, \mathcal{Y}_t\} \\ &= \frac{p(\mathbf{y}_{t+1}, s_t = j | s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t)}{f(\mathbf{y}_{t+1} | s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t)} \\ &= \frac{f(\mathbf{y}_{t+1} | s_t = j, s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t) \cdot P\{s_t = j | s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t\}}{f(\mathbf{y}_{t+1} | s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t)}, \end{aligned} \qquad [22.A.14]$$

which simplifies to

$$P\{s_t = j | s_{t+1} = i, \mathcal{Y}_{t+1}\} = P\{s_t = j | s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t\}, \qquad [22.A.15]$$

provided that

$$f(\mathbf{y}_{t+1} | s_t = j, s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t) = f(\mathbf{y}_{t+1} | s_{t+1} = i, \mathbf{x}_{t+1}, \mathcal{Y}_t), \qquad [22.A.16]$$

which is indeed the case, since the specification assumes that $\mathbf{y}_{t+1}$ depends on $\{s_{t+1}, s_t, \ldots\}$ only through the current value s_{t+1}. Since $\mathbf{x}$ is exogenous, [22.A.15] further implies that

$$P\{s_t = j | s_{t+1} = i, \mathcal{Y}_{t+1}\} = P\{s_t = j | s_{t+1} = i, \mathcal{Y}_t\}. \qquad [22.A.17]$$

By similar reasoning, it must be the case that

$$\begin{aligned} &P\{s_t = j | s_{t+1} = i, \mathcal{Y}_{t+2}\} \\ &= P\{s_t = j | s_{t+1} = i, \mathbf{y}_{t+2}, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}\} \\ &= \frac{p(\mathbf{y}_{t+2}, s_t = j | s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1})}{f(\mathbf{y}_{t+2} | s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1})} \\ &= \frac{f(\mathbf{y}_{t+2} | s_t = j, s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}) \cdot P\{s_t = j | s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}\}}{f(\mathbf{y}_{t+2} | s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1})}, \end{aligned}$$

which simplifies to

$$P\{s_t = j|s_{t+1} = i, \mathcal{Y}_{t+2}\} = P\{s_t = j|s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}\}, \qquad [22.A.18]$$

provided that

$$f(\mathbf{y}_{t+2}|s_t = j, s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}) = f(\mathbf{y}_{t+2}|s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}). \qquad [22.A.19]$$

In this case, [22.A.19] is established from the fact that

$$\begin{aligned}
&f(\mathbf{y}_{t+2}|s_t = j, s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}) \\
&= \sum_{k=1}^{N} p(\mathbf{y}_{t+2}, s_{t+2} = k|s_t = j, s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}) \\
&= \sum_{k=1}^{N} [f(\mathbf{y}_{t+2}|s_{t+2} = k, s_t = j, s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}) \\
&\qquad \times P\{s_{t+2} = k|s_t = j, s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}\}] \\
&= \sum_{k=1}^{N} [f(\mathbf{y}_{t+2}|s_{t+2} = k, s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}) \\
&\qquad \times P\{s_{t+2} = k|s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}\}] \\
&= f(\mathbf{y}_{t+2}|s_{t+1} = i, \mathbf{x}_{t+2}, \mathcal{Y}_{t+1}).
\end{aligned}$$

Again, exogeneity of $\mathbf{x}$ means that [22.A.18] can be written

$$P\{s_t = j|s_{t+1} = i, \mathcal{Y}_{t+2}\} = P\{s_t = j|s_{t+1} = i, \mathcal{Y}_{t+1}\} = P\{s_t = j|s_{t+1} = i, \mathcal{Y}_t\},$$

where the last equality follows from [22.A.17].

Proceeding inductively, the same argument can be used to establish that

$$P\{s_t = j|s_{t+1} = i, \mathcal{Y}_{t+m}\} = P\{s_t = j|s_{t+1} = i, \mathcal{Y}_t\}$$

for $m = 1, 2, \ldots$, from which [22.A.13] follows.

Note next that

$$\begin{aligned}
P\{s_t = j|s_{t+1} = i, \mathcal{Y}_t\} &= \frac{P\{s_t = j, s_{t+1} = i|\mathcal{Y}_t\}}{P\{s_{t+1} = i|\mathcal{Y}_t\}} \\
&= \frac{P\{s_t = j|\mathcal{Y}_t\}\cdot P\{s_{t+1} = i|s_t = j\}}{P\{s_{t+1} = i|\mathcal{Y}_t\}} \\
&= \frac{p_{ji}\cdot P\{s_t = j|\mathcal{Y}_t\}}{P\{s_{t+1} = i|\mathcal{Y}_t\}}.
\end{aligned} \qquad [22.A.20]$$

It is therefore the case that

$$\begin{aligned}
P\{s_t = j, s_{t+1} = i|\mathcal{Y}_T\} &= P\{s_{t+1} = i|\mathcal{Y}_T\}\cdot P\{s_t = j|s_{t+1} = i, \mathcal{Y}_T\} \\
&= P\{s_{t+1} = i|\mathcal{Y}_T\}\cdot P\{s_t = j|s_{t+1} = i, \mathcal{Y}_t\} \\
&= P\{s_{t+1} = i|\mathcal{Y}_T\}\frac{p_{ji}\cdot P\{s_t = j|\mathcal{Y}_t\}}{P\{s_{t+1} = i|\mathcal{Y}_t\}},
\end{aligned} \qquad [22.A.21]$$

where the second equality follows from [22.A.13] and the third follows from [22.A.20].

The smoothed inference for date t is the sum of [22.A.21] over $i = 1, 2, \ldots, N$:

$$\begin{aligned}
P\{s_t = j|\mathcal{Y}_T\} &= \sum_{i=1}^{N} P\{s_t = j, s_{t+1} = i|\mathcal{Y}_T\} \\
&= \sum_{i=1}^{N} P\{s_{t+1} = i|\mathcal{Y}_T\}\frac{p_{ji}\cdot P\{s_t = j|\mathcal{Y}_t\}}{P\{s_{t+1} = i|\mathcal{Y}_t\}} \\
&= P\{s_t = j|\mathcal{Y}_t\}\sum_{i=1}^{N}\frac{p_{ji}\cdot P\{s_{t+1} = i|\mathcal{Y}_T\}}{P\{s_{t+1} = i|\mathcal{Y}_t\}} \\
&= P\{s_t = j|\mathcal{Y}_t\}[p_{j1} \quad p_{j2} \quad \cdots \quad p_{jN}] \\
&\qquad \times \begin{bmatrix} P\{s_{t+1} = 1|\mathcal{Y}_T\}/P\{s_{t+1} = 1|\mathcal{Y}_t\} \\ P\{s_{t+1} = 2|\mathcal{Y}_T\}/P\{s_{t+1} = 2|\mathcal{Y}_t\} \\ \vdots \\ P\{s_{t+1} = N|\mathcal{Y}_T\}/P\{s_{t+1} = N|\mathcal{Y}_t\} \end{bmatrix} \\
&= P\{s_t = j|\mathcal{Y}_t\}\mathbf{p}_j'(\hat{\boldsymbol{\xi}}_{t+1|T}\,(\div)\,\hat{\boldsymbol{\xi}}_{t+1|t}),
\end{aligned} \qquad [22.A.22]$$

where the $(1 \times N)$ vector $\mathbf{p}_j'$ denotes the jth row of the matrix $\mathbf{P}'$ and the sign $(\div)$ indicates element-by-element division. When the equations represented by [22.A.22] for $j = 1, 2, \ldots, N$ are collected in an $(N \times 1)$ vector, the result is

$$\hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} \odot \{\mathbf{P}'(\hat{\boldsymbol{\xi}}_{t+1|T} (\div) \hat{\boldsymbol{\xi}}_{t+1|t})\},$$

as claimed. ∎

Chapter 22 Exercise

22.1. Let s_t be described by an ergodic two-state Markov chain with transition matrix $\mathbf{P}$ given by [22.2.11]. Verify that the matrix of eigenvectors of this matrix is given by

$$\mathbf{T} = \begin{bmatrix} (1 - p_{22})/(2 - p_{11} - p_{22}) & -1 \\ (1 - p_{11})/(2 - p_{11} - p_{22}) & 1 \end{bmatrix}$$

with inverse

$$\mathbf{T}^{-1} = \begin{bmatrix} 1 & 1 \\ -(1 - p_{11})/(2 - p_{11} - p_{22}) & (1 - p_{22})/(2 - p_{11} - p_{22}) \end{bmatrix}.$$

Chapter 22 References

Andrews, Donald W. K., and Werner Ploberger. 1992. "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative." Yale University. Mimeo.

Cox, D. R., and H. D. Miller. 1965. *The Theory of Stochastic Processes*. London: Methuen.

Davies, R. B. 1977. "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative." *Biometrika* 64:247–54.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society* Series B, 39:1–38.

Diebold, Francis X., Joon-Haeng Lee, and Gretchen C. Weinbach. Forthcoming. "Regime Switching with Time-Varying Transition Probabilities," in C. Hargreaves, ed., *Nonstationary Time Series Analysis and Cointegration*. Oxford: Oxford University Press.

Durland, J. Michael, and Thomas H. McCurdy. 1992. "Modelling Duration Dependence in Cyclical Data Using a Restricted Semi-Markov Process." Queen's University, Kingston, Ontario. Mimeo.

Engel, Charles, and James D. Hamilton. 1990. "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?" *American Economic Review* 80:689–713.

Everitt, B. S., and D. J. Hand. 1981. *Finite Mixture Distributions*. London: Chapman and Hall.

Filardo, Andrew J. 1992. "Business Cycle Phases and Their Transitional Dynamics." Federal Reserve Bank of Kansas City. Mimeo.

Goodwin, Thomas H. 1993. "Business Cycle Analysis with a Markov-Switching-Model." *Journal of Business and Economic Statistics* 11:331–39.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57:357–84.

———. 1990. "Analysis of Time Series Subject to Changes in Regime." *Journal of Econometrics* 45:39–70.

———. 1991. "A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions." *Journal of Business and Economic Statistics* 9:27–39.

———. 1993a. "Specification Testing in Markov-Switching Time Series Models." University of California, San Diego. Mimeo.

———. 1993b. "Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime," in G. S. Maddala, C. R. Rao, and H. D. Vinod, eds., *Handbook of Statistics*, Vol. 11. New York: North-Holland.

———. 1993c. "State-Space Models," in Robert Engle and Daniel McFadden, eds., *Handbook of Econometrics*, Vol. 4. New York: North-Holland.

Hansen, Bruce E. 1992. "The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP." *Journal of Applied Econometrics* 7:S61–82.

———. 1993. "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis." University of Rochester. Mimeo.

Kiefer, Nicholas M. 1978. "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model." *Econometrica* 46:427–34.

Kim, Chang-Jin. 1993. "Dynamic Linear Models with Markov-Switching." *Journal of Econometrics*, forthcoming.

Lam, Pok-sang. 1990. "The Hamilton Model with a General Autoregressive Component: Estimation and Comparison with Other Models of Economic Time Series." *Journal of Monetary Economics* 26:409–32.

Rogers, John H. 1992. "The Currency Substitution Hypothesis and Relative Money Demand in Mexico and Canada." *Journal of Money, Credit, and Banking* 24:300–18.

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

Titterington, D. M., A. F. M. Smith, and U. E. Makov. 1985. *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.

A Mathematical Review

This book assumes some familiarity with elementary trigonometry, complex numbers, calculus, matrix algebra, and probability. Introductions to the first three topics by Chiang (1974) or Thomas (1972) are adequate; Marsden (1974) treated these issues in more depth. No matrix algebra is required beyond the level of standard econometrics texts such as Theil (1971) or Johnston (1984); for more detailed treatments, see O'Nan (1976), Strang (1976), and Magnus and Neudecker (1988). The concepts of probability and statistics from standard econometrics texts are also sufficient for getting through this book; for more complete introductions, see Lindgren (1976) or Hoel, Port, and Stone (1971).

This appendix reviews the necessary mathematical concepts and results. The reader familiar with these topics is invited to skip this material, or consult subheadings for desired coverage.

A.1. Trigonometry

Definitions

Figure A.1 displays a circle with unit radius centered at the origin in (x, y)-space. Let (x_0, y_0) denote some point on this unit circle, and consider the angle θ between this point and the x-axis. The *sine* of θ is defined as the y-coordinate of the point, and the *cosine* is the x-coordinate:

$$\sin(\theta) = y_0 \qquad [A.1.1]$$

$$\cos(\theta) = x_0. \qquad [A.1.2]$$

This text always measures angles in *radians*. The radian measure of the angle θ is defined as the distance traveled counterclockwise along the unit circle starting at the x-axis before reaching (x_0, y_0). The circumference of a circle with unit radius is 2π. A rotation one-quarter of the way around the unit circle would therefore correspond to radian measure of $\theta = \frac{1}{4}(2\pi) = \pi/2$. An angle whose radian measure is $\pi/2$ is more commonly described as a right angle or a 90° angle. A 45° angle has radian measure of $\pi/4$, a 180° angle has radian measure of π, and so on.

Polar Coordinates

Consider a smaller triangle—say, the triangle with vertex (x_1, y_1) shown in Figure A.1—that shares the same angle θ as the original triangle with vertex

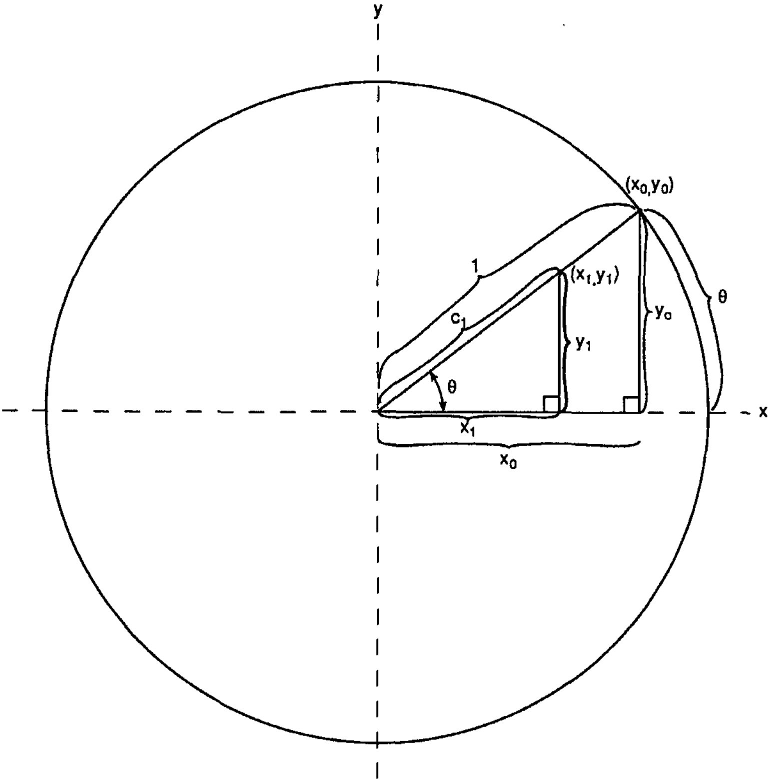


FIGURE A.1 Trigonometric functions as distances in (x, y)-space.

(x_0, y_0). The ratio of any two sides of such a smaller triangle will be the same as that for the larger triangle:

$$y_1/c_1 = y_0/1 \quad [A.1.3]$$

$$x_1/c_1 = x_0/1. \quad [A.1.4]$$

Comparing [A.1.3] with [A.1.1], the y-coordinate of any point such as (x_1, y_1) in (x, y)-space may be expressed as

$$y_1 = c_1 \cdot \sin(\theta), \quad [A.1.5]$$

where c_1 is the distance from the origin to (x_1, y_1) and θ is the angle that the point (x_1, y_1) makes with the x-axis. Comparing [A.1.4] with [A.1.2], the x-coordinate of (x_1, y_1) can be expressed as

$$x_1 = c_1 \cdot \cos(\theta). \quad [A.1.6]$$

Recall further that the magnitude c_1, which represents the distance from the origin to the point (x_1, y_1), is given by the formula

$$c_1 = \sqrt{x_1^2 + y_1^2}. \quad [A.1.7]$$

Taking a point in (x, y)-space and writing it as $(c \cdot \cos(\theta),\ c \cdot \sin(\theta))$ is called describing the point in terms of its *polar coordinates* c and θ.

Properties of Sine and Cosine Functions

The functions $\sin(\theta)$ and $\cos(\theta)$ are called *trigonometric* or *sinusoidal* functions. Viewed as a function of θ, the sine function starts out at zero:

$$\sin(0) = 0.$$

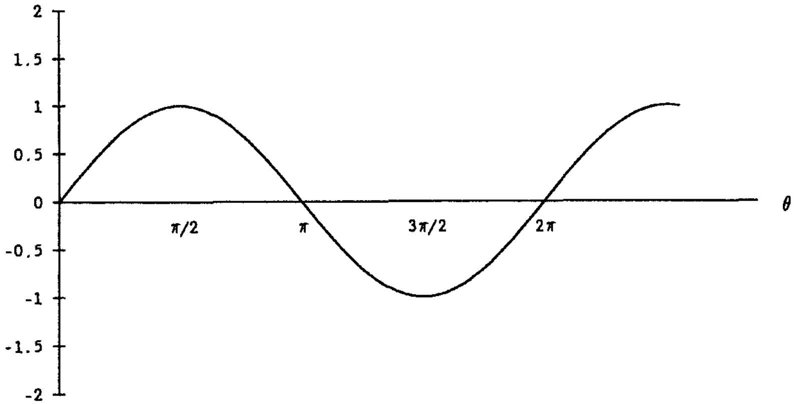
The sine function rises to 1 as θ increases to $\pi/2$ and then falls back to zero as θ increases further to π; see panel (a) of Figure A.2. The function reaches its minimum value of -1 at $\theta = 3\pi/2$ and then begins climbing back up.

If we travel a distance of 2π radians around the unit circle, we are right back where we started, and the function repeats itself:

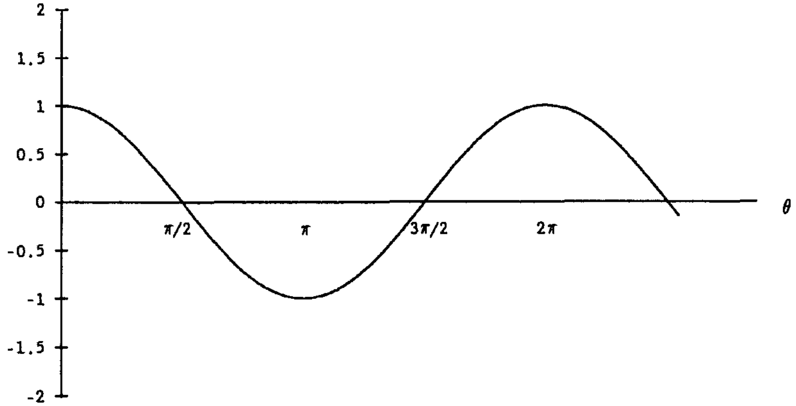
$$\sin(2\pi + \theta) = \sin(\theta).$$

The function would again repeat itself if we made two full revolutions around the unit circle. Indeed for any integer j,

$$\sin(2\pi j + \theta) = \sin(\theta). \qquad [A.1.8]$$



(a) $\sin(\theta)$



(b) $\cos(\theta)$

FIGURE A.2 Sine and cosine functions.

The sine function is thus *periodic* and is for this reason often useful for describing a time series that repeats itself in a particular cycle.

The cosine function starts out at unity and falls to zero as θ increases to $\pi/2$; see panel (b) of Figure A.2. It turns out simply to be a horizontal shift of the sine function:

$$\cos(\theta) = \sin\left(\theta + \frac{\pi}{2}\right). \qquad [A.1.9]$$

The sine or cosine function can also be evaluated for negative values of θ, defined as a clockwise rotation around the unit circle from the x-axis. Clearly,

$$\sin(-\theta) = -\sin(\theta) \qquad [A.1.10]$$

$$\cos(-\theta) = \cos(\theta). \qquad [A.1.11]$$

For (x_0, y_0) a point on the unit circle, [A.1.7] implies that

$$1 = \sqrt{x_0^2 + y_0^2},$$

or, squaring both sides and using [A.1.1] and [A.1.2],

$$1 = [\cos(\theta)]^2 + [\sin(\theta)]^2. \qquad [A.1.12]$$

Using Trigonometric Functions to Represent Cycles

Suppose we construct the function $g(\theta)$ by first multiplying θ by 2 and then evaluating the sine of the product:

$$g(\theta) = \sin(2\theta).$$

This doubles the frequency at which the function cycles. When θ goes from 0 to π, 2θ goes from 0 to 2π, and so $g(\theta)$ is back to its original value (see Figure A.3). In general, the function $\sin(k\theta)$ would go through k cycles in the time it takes $\sin(\theta)$ to complete a single cycle.

We will sometimes describe the value a variable y takes on at date t as a function of sines or cosines, such as

$$y_t = R \cdot \cos(\omega t + \alpha). \qquad [A.1.13]$$

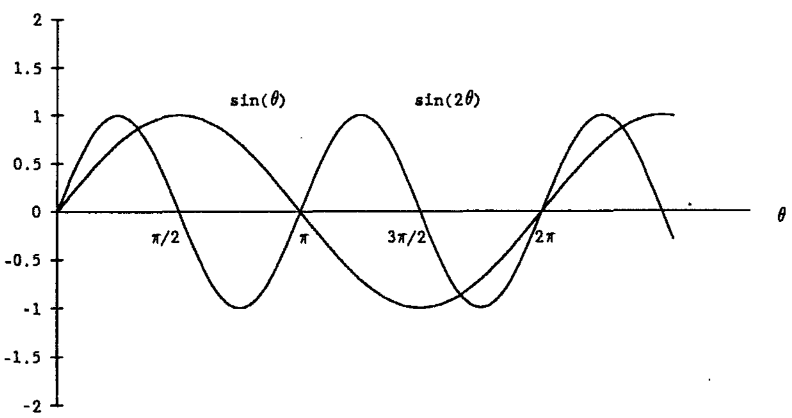


FIGURE A.3 Effect of changing frequency of a periodic function.

The parameter R gives the *amplitude* of [A.1.13]. The variable y_t will attain a maximum value of $+R$ and a minimum value of $-R$. The parameter α is the *phase*. The phase determines where in the cycle y_t would be at $t = 0$. The parameter ω governs how quickly the variable cycles, which can be summarized by either of two measures. The *period* is the length of time required for the process to repeat a full cycle. The period of [A.1.13] is $2\pi/\omega$. For example, if $\omega = 1$ then y repeats itself every 2π periods, whereas if $\omega = 2$ the process repeats itself every π periods. The *frequency* summarizes how frequently the process cycles compared with the simple function $\cos(t)$; thus, it measures the number of cycles completed during 2π periods. The frequency of $\cos(t)$ is unity, and the frequency of [A.1.13] is ω. For example, if $\omega = 2$, the cycles are completed twice as quickly as those for $\cos(t)$. There is a simple relation between these two measures of the speed of cycles—the period is equal to 2π divided by the frequency.

A.2. Complex Numbers

Definitions

Consider the following expression:

$$x^2 = 1. \tag{A.2.1}$$

There are two values of x that satisfy [A.2.1], namely, $x = 1$ and $x = -1$.

Suppose instead that we were given the following equation:

$$x^2 = -1. \tag{A.2.2}$$

No real number satisfies [A.2.2]. However, let us consider an imaginary number (denoted i) that does:

$$i^2 = -1. \tag{A.2.3}$$

We assume that i can be multiplied by a real number and manipulated using standard rules of algebra. For example,

$$2i + 3i = 5i$$

and

$$(2i)\cdot(3i) = (6)i^2 = -6.$$

This last property implies that a second solution to [A.2.2] is given by $x = -i$:

$$(-i)^2 = (-1)^2(i)^2 = -1.$$

Thus, [A.2.1] has two real roots ($+1$ and -1), whereas [A.2.2] has two imaginary roots (i and $-i$).

For any real numbers a and b, we can construct the expression

$$a + bi. \tag{A.2.4}$$

If $b = 0$, then [A.2.4] is a *real* number; whereas if $a = 0$ and b is nonzero, then [A.2.4] is an *imaginary* number. A number written in the general form of [A.2.4] is called a *complex* number.

Rules for Manipulating Complex Numbers

Complex numbers are manipulated using standard rules of algebra. Two complex numbers are added as follows:

$$(a_1 + b_1 i) + (a_2 + b_2 i) = (a_1 + a_2) + (b_1 + b_2)i.$$

Complex numbers are multiplied this way:

$$\begin{aligned}(a_1 + b_1 i)\cdot(a_2 + b_2 i) &= a_1a_2 + a_1b_2i + b_1a_2i + b_1b_2i^2 \\ &= (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)i.\end{aligned}$$

Note that the resulting expressions are always simplified by separating the real component (such as $[a_1a_2 - b_1b_2]$) from the imaginary component (such as $[a_1b_2 + b_1a_2]i$).

Graphical Representation of Complex Numbers

A complex number $(a + bi)$ is sometimes represented graphically in an *Argand diagram* as in Figure A.4. The value of the real component (a) is plotted on the horizontal axis, and the imaginary component (b) is plotted on the vertical axis. The size, or *modulus*, of a complex number is measured the same way as the distance from the origin of a real element in (x, y)-space (see equation [A.1.7]):

$$|a + bi| \equiv \sqrt{a^2 + b^2}. \qquad [A.2.5]$$

The *complex unit circle* is the set of all complex numbers whose modulus is 1. For example, the real number +1 is on the complex unit circle (represented by

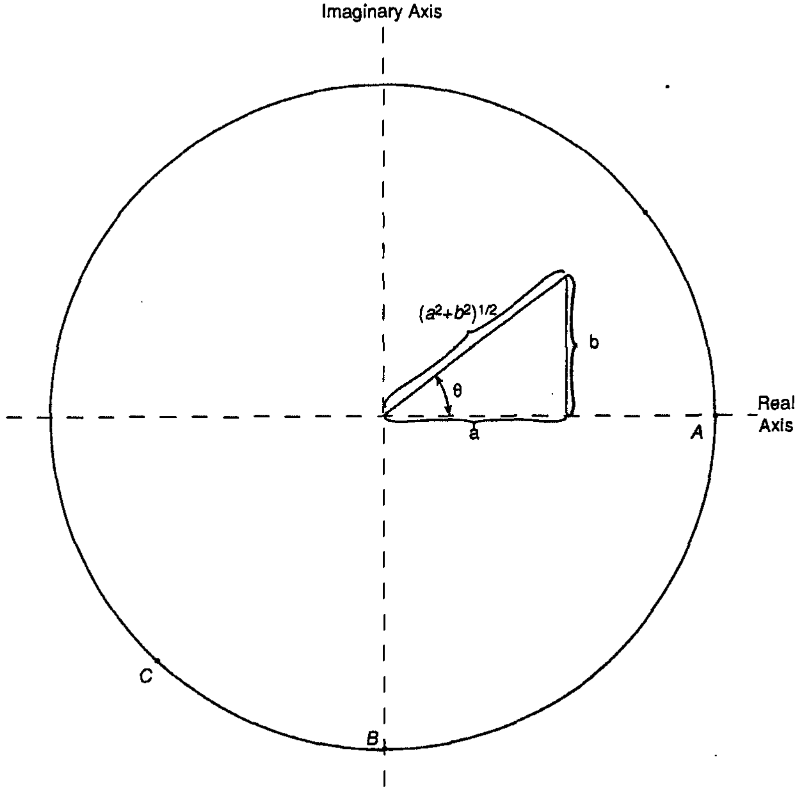


FIGURE A.4 Argand diagram and the complex unit circle.

the point A in Figure A.4). So are the imaginary number $-i$ (point B) and the complex number $(-0.6 - 0.8i)$ (point C).

We will often be interested in whether a complex number is less than 1 in modulus, in which case the number is said to be inside the unit circle. For example, $(-0.3 + 0.4i)$ has modulus 0.5, so it lies inside the unit circle, whereas $(3 + 4i)$, with modulus 5, lies outside the unit circle.

Polar Coordinates

Just as a point in (x, y)-space can be represented by its distance c from the origin and its angle θ with the x-axis, the complex number $a + bi$ can be represented by the distance of (a, b) from the origin (the modulus of the complex number),

$$R = \sqrt{a^2 + b^2},$$

and by the angle θ that the point (a, b) makes with the real axis, characterized by

$$\cos(\theta) = a/R$$
$$\sin(\theta) = b/R.$$

Thus, the complex number $a + bi$ is written in polar coordinate form as

$$[R \cdot \cos(\theta) + i \cdot R \cdot \sin(\theta)] = R[\cos(\theta) + i \cdot \sin(\theta)]. \qquad [A.2.6]$$

Complex Conjugates

The *complex conjugate* of $(a + bi)$ is given by $(a - bi)$. The numbers $(a + bi)$ and $(a - bi)$ are described as a *conjugate pair.* Notice that adding a conjugate pair produces a real result:

$$(a + bi) + (a - bi) = 2a.$$

The product of a conjugate pair is also real:

$$(a + bi) \cdot (a - bi) = a^2 + b^2. \qquad [A.2.7]$$

Comparing this with [A.2.5], we see that the modulus of a complex number $(a + bi)$ can be thought of as the square root of the product of the number with its complex conjugate:

$$|a + bi| = \sqrt{(a + bi)(a - bi)}. \qquad [A.2.8]$$

Quadratic Equations

A *quadratic equation*

$$\alpha x^2 + \beta x + \gamma = 0 \qquad [A.2.9]$$

with $\alpha \neq 0$ has two solutions:

$$x_1 = \frac{-\beta + (\beta^2 - 4\alpha\gamma)^{1/2}}{2\alpha} \qquad [A.2.10]$$

$$x_2 = \frac{-\beta - (\beta^2 - 4\alpha\gamma)^{1/2}}{2\alpha}. \qquad [A.2.11]$$

When $(\beta^2 - 4\alpha\gamma) \geq 0$, both these roots are real, whereas when $(\beta^2 - 4\alpha\gamma) < 0$, the roots are complex. Notice that when the roots are complex they appear as a conjugate pair:

$$x_1 = \{-\beta/[2\alpha]\} + \{(1/[2\alpha])(4\alpha\gamma - \beta^2)^{1/2}\}i$$
$$x_2 = \{-\beta/[2\alpha]\} - \{(1/[2\alpha])(4\alpha\gamma - \beta^2)^{1/2}\}i.$$

A.3. *Calculus*

Continuity

A function $f(x)$ is said to be *continuous* at $x = c$ if $f(c)$ is finite and if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ whenever $|x - c| < \delta$.

Derivatives of Some Simple Functions

The *derivative* of $f(\cdot)$ with respect to x is defined by

$$\frac{df}{dx} = \lim_{\Delta\to 0} \frac{f(x + \Delta) - f(x)}{\Delta},$$

provided that this limit exists.

If $f(\cdot)$ is linear in x, or

$$f(x) = \alpha + \beta x,$$

then the derivative is just the coefficient on x:

$$\frac{df}{dx} = \lim_{\Delta\to 0} \frac{[\alpha + \beta(x + \Delta)] - [\alpha + \beta x]}{\Delta} = \lim_{\Delta\to 0} \frac{\beta\Delta}{\Delta} = \beta.$$

For a quadratic function

$$f(x) = x^2,$$

the derivative is

$$\begin{aligned}\frac{df}{dx} &= \lim_{\Delta\to 0} \frac{[x + \Delta]^2 - x^2}{\Delta} \\ &= \lim_{\Delta\to 0} \frac{[x^2 + 2x\Delta + \Delta^2] - x^2}{\Delta} \\ &= \lim_{\Delta\to 0} \{2x + \Delta\} \\ &= 2x,\end{aligned}$$

and in general,

$$\frac{dx^k}{dx} = kx^{k-1}. \qquad [A.3.1]$$

For the trigonometric functions, it can be shown that when x is measured in radians,

$$\frac{d\sin(x)}{dx} = \cos(x) \qquad [A.3.2]$$

$$\frac{d\cos(x)}{dx} = -\sin(x). \qquad [A.3.3]$$

The derivative $df(x)/dx$ is itself a function of x. Often we want to specify the point at which the derivative should be evaluated, say, c. This is indicated by

$$\left.\frac{df(x)}{dx}\right|_{x=c}.$$

For example,

$$\left.\frac{dx^2}{dx}\right|_{x=3} = 2x|_{x=3} = 6.$$

Note that this notation refers to taking the derivative first and then evaluating the derivative at a particular point such as $x = 3$.

Chain Rule

The *chain rule* states that for composite functions such as

$$g(x) = f(u(x)),$$

the derivative is

$$\frac{dg(x)}{dx} = \frac{df}{du}\cdot\frac{du}{dx}. \qquad \text{[A.3.4]}$$

For example, to evaluate

$$\frac{d(\alpha + \beta x)^k}{dx},$$

we let $f(u) = u^k$ and $u(x) = \alpha + \beta x$. Then

$$\frac{df}{du}\cdot\frac{du}{dx} = ku^{k-1}\cdot\beta.$$

Thus,

$$\frac{d(\alpha + \beta x)^k}{dx} = \beta k(\alpha + \beta x)^{k-1}.$$

Higher-Order Derivatives

The *second derivative* is defined by

$$\frac{d^2f(x)}{dx^2} = \frac{d}{dx}\left[\frac{df(x)}{dx}\right].$$

For example,

$$\frac{d^2x^k}{dx^2} = \frac{d[kx^{k-1}]}{dx} = k(k - 1)x^{k-2}$$

and

$$\frac{d^2 \sin(x)}{dx^2} = \frac{d \cos(x)}{dx} = -\sin(x). \qquad \text{[A.3.5]}$$

In general, the jth-order derivative is the derivative of the $(j - 1)$th-order derivative.

Geometric Series

Consider the sum

$$s_T = 1 + \phi + \phi^2 + \phi^3 + \cdots + \phi^T. \qquad [A.3.6]$$

Multiplying both sides of [A.3.6] by ϕ,

$$\phi s_T = \phi + \phi^2 + \phi^3 + \cdots + \phi^T + \phi^{T+1}. \qquad [A.3.7]$$

Subtracting [A.3.7] from [A.3.6] produces

$$(1 - \phi)s_T = 1 - \phi^{T+1}. \qquad [A.3.8]$$

For any $\phi \neq 1$, both sides of [A.3.8] can be divided by $(1 - \phi)$. Hence, the sum in [A.3.6] is equal to

$$s_T = \begin{cases} \dfrac{1 - \phi^{T+1}}{1 - \phi} & \phi \neq 1 \\ T + 1 & \phi = 1. \end{cases} \qquad [A.3.9]$$

From [A.3.9],

$$\lim_{T \to \infty} s_T = \frac{1}{1 - \phi} \qquad |\phi| < 1,$$

and so

$$(1 + \phi + \phi^2 + \phi^3 + \cdots) = \frac{1}{1 - \phi} \qquad |\phi| < 1. \qquad [A.3.10]$$

Taylor Series Approximations

Suppose that the first through the $(r + 1)$th derivatives of a function $f(x)$ exist and are continuous in a neighborhood of c. *Taylor's theorem* states that the value of $f(x)$ at $x = c + \Delta$ is given by

$$\begin{aligned} f(c + \Delta) = f(c) &+ \left.\frac{df}{dx}\right|_{x=c} \cdot \Delta + \frac{1}{2!}\left.\frac{d^2f}{dx^2}\right|_{x=c} \cdot \Delta^2 \\ &+ \frac{1}{3!}\left.\frac{d^3f}{dx^3}\right|_{x=c} \cdot \Delta^3 + \cdots + \frac{1}{r!}\left.\frac{d^rf}{dx^r}\right|_{x=c} \cdot \Delta^r + R_r(c, x), \end{aligned} \qquad [A.3.11]$$

where $r!$ denotes *r factorial:*

$$r! \equiv r(r - 1)\cdot(r - 2)\cdots 2\cdot 1.$$

The remainder $R_r(c, x)$ is given by

$$R_r(c, x) = \frac{1}{(r + 1)!} \cdot \left.\frac{d^{r+1}f}{dx^{r+1}}\right|_{x=\delta} \cdot \Delta^{r+1},$$

where δ is a number between c and x. Notice that the remainder vanishes for small Δ:

$$\lim_{\Delta \to 0} \frac{R_r(c, x)}{\Delta^r} = 0.$$

Setting $R_r(c, x) = 0$ and $x = c + \Delta$ in [A.3.11] produces an *rth-order Taylor series approximation* to the function $f(x)$ in the neighborhood of $x = c$:

$$f(x) \cong f(c) + \left.\frac{df}{dx}\right|_{x=c} \cdot (x - c) + \frac{1}{2!}\left.\frac{d^2 f}{dx^2}\right|_{x=c} \cdot (x - c)^2 + \cdots + \frac{1}{r!}\left.\frac{d^r f}{dx^r}\right|_{x=c} \cdot (x - c)^r. \qquad [A.3.12]$$

Power Series

If the remainder $R_r(c, x)$ in [A.3.11] converges to zero for all x as $r \to \infty$, a power series can be used to characterize the function $f(x)$. To find a power series, we choose a particular value c around which to center the expansion, such as $c = 0$. We then use [A.3.12] with $r \to \infty$. For example, consider the sine function. The first two derivatives are given by [A.3.2] and [A.3.5], with the following higher-order derivatives:

$$\frac{d^3 \sin(x)}{dx^3} = -\cos(x)$$
$$\frac{d^4 \sin(x)}{dx^4} = \sin(x)$$
$$\frac{d^5 \sin(x)}{dx^5} = \cos(x),$$

and so on. Evaluated at $x = 0$, we have

$$f(0) = \sin(0) = 0$$
$$\left.\frac{df}{dx}\right|_{x=0} = \cos(0) = 1$$
$$\left.\frac{d^2 f}{dx^2}\right|_{x=0} = -\sin(0) = 0$$
$$\left.\frac{d^3 f}{dx^3}\right|_{x=0} = -\cos(0) = -1$$
$$\left.\frac{d^4 f}{dx^4}\right|_{x=0} = \sin(0) = 0$$
$$\left.\frac{d^5 f}{dx^5}\right|_{x=0} = \cos(0) = 1.$$

Substituting into [A.3.12] with $c = 0$ and letting $r \to \infty$ produces a power series for the sine function:

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \cdots. \qquad [A.3.13]$$

Similar calculations give a power series for the cosine function:

$$\cos(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \cdots. \qquad [A.3.14]$$

Exponential Functions

A number γ raised to the power x,

$$f(x) = \gamma^x,$$

is called an *exponential function* of x. The number γ is called the *base* of this function, and x is called the *exponent*. To multiply two exponential functions that share the same base, the exponents are added:

$$(\gamma^x)\cdot(\gamma^y) = \gamma^{(x+y)}. \qquad [A.3.15]$$

For example,

$$(\gamma^2)\cdot(\gamma^3) = (\gamma\cdot\gamma)\cdot(\gamma\cdot\gamma\cdot\gamma) = \gamma^5.$$

To raise an exponential function to the power k, the exponents are multiplied:

$$[\gamma^x]^k = \gamma^{xk}. \qquad [A.3.16]$$

For example,

$$[\gamma^2]^3 = [\gamma^2]\cdot[\gamma^2]\cdot[\gamma^2] = \gamma^6.$$

Exponentiation is distributive over multiplication:

$$(\alpha\cdot\beta)^x = (\alpha^x)\cdot(\beta^x). \qquad [A.3.17]$$

Negative exponents denote reciprocals:

$$\gamma^{-k} = (1/\gamma^k).$$

Any number raised to the power 0 is taken to be equal to unity:

$$\gamma^0 = 1. \qquad [A.3.18]$$

This convention is sensible, since if $y = -x$ in [A.3.15],

$$(\gamma^x)(\gamma^{-x}) = \gamma^0$$

and

$$(\gamma^x)(\gamma^{-x}) = \frac{\gamma^x}{\gamma^x} = 1.$$

The Number e

The base for the natural logarithms is denoted e. The number e has the property that an exponential function with base e equals its own derivative:

$$\frac{de^x}{dx} = e^x. \qquad [A.3.19]$$

Clearly, all the higher-order derivatives of e^x are equal to e^x as well:

$$\frac{d^r e^x}{dx^r} = e^x. \qquad [A.3.20]$$

We sometimes use the expression "exp[x]" to represent "e raised to the power x":

$$\exp[x] \equiv e^x.$$

If $u(x)$ denotes a separate function of x, the derivative of the compound function $e^{u(x)}$ can be evaluated using the chain rule:

$$\frac{de^{u(x)}}{dx} = \frac{de^u}{du}\cdot\frac{du}{dx} = e^{u(x)}\frac{du}{dx}. \qquad [A.3.21]$$

To find a power series for the function $f(x) = e^x$, notice from [A.3.20] that

$$\frac{d^r f}{dx^r} = e^x,$$

and so, from [A.3.18],

$$\left.\frac{d^r f}{dx^r}\right|_{x=0} = e^0 = 1 \qquad \text{[A.3.22]}$$

for all r. Substituting [A.3.22] into [A.3.12] with $c = 0$ yields a power series for the function $f(x) = e^x$:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots. \qquad \text{[A.3.23]}$$

Setting $x = 1$ in [A.3.23] gives a numerical procedure for calculating the value of e:

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots = 2.71828\ldots.$$

Euler Relations and De Moivre's Theorem

Suppose we evaluate the power series [A.3.23] at the imaginary number $x = i\theta$, where $i = \sqrt{-1}$ and θ is some real angle measured in radians:

$$\begin{aligned} e^{i\theta} &= 1 + (i\theta) + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \cdots \\ &= \left\{1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right\} + i\cdot\left\{\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right\}. \end{aligned} \qquad \text{[A.3.24]}$$

Reflecting on [A.3.13] and [A.3.14] gives another interpretation of [A.3.24]:

$$e^{i\theta} = \cos(\theta) + i\cdot\sin(\theta). \qquad \text{[A.3.25]}$$

Similarly,

$$\begin{aligned} e^{-i\theta} &= 1 + (-i\theta) + \frac{(-i\theta)^2}{2!} + \frac{(-i\theta)^3}{3!} + \frac{(-i\theta)^4}{4!} + \frac{(-i\theta)^5}{5!} + \cdots \\ &= \left\{1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right\} - i\cdot\left\{\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right\} \\ &= \cos(\theta) - i\cdot\sin(\theta). \end{aligned} \qquad \text{[A.3.26]}$$

To raise a complex number $(a + bi)$ to the kth power, the complex number is written in polar coordinate form as in [A.2.6]:

$$a + bi = R[\cos(\theta) + i\cdot\sin(\theta)].$$

Using [A.3.25], this can then be treated as an exponential function of θ:

$$a + bi = R\cdot e^{i\theta}. \qquad \text{[A.3.27]}$$

Now raise both sides of [A.3.27] to the kth power, recalling [A.3.17] and [A.3.16]:

$$(a + bi)^k = R^k\cdot[e^{i\theta}]^k = R^k\cdot e^{i\theta k}. \qquad \text{[A.3.28]}$$

Finally, use [A.3.25] in reverse,

$$e^{i(\theta k)} = \cos(\theta k) + i\cdot\sin(\theta k),$$

to deduce that [A.3.28] can be written

$$(a + bi)^k = R^k \cdot [\cos(\theta k) + i \cdot \sin(\theta k)]. \qquad [A.3.29]$$

Definition of Natural Logarithm

The *natural logarithm* (denoted throughout the text simply by "log") is the inverse of the function e^x:

$$\log(e^x) \equiv x.$$

Notice from [A.3.18] that $e^0 = 1$ and therefore $\log(1) = 0$.

Properties of Logarithms

For any $x > 0$, it is also the case that

$$x = e^{\log(x)}. \qquad [A.3.30]$$

From [A.3.30] and [A.3.15], we see that the log of the product of two numbers is equal to the sum of the logs:

$$\log(a \cdot b) = \log[(e^{\log(a)}) \cdot (e^{\log(b)})] = \log[e^{(\log(a) + \log(b))}] = \log(a) + \log(b).$$

Also, use [A.3.16] to write

$$x^a = [e^{\log(x)}]^a = e^{a \cdot \log(x)}. \qquad [A.3.31]$$

Taking logs of both sides of [A.3.31] reveals that the log of a number raised to the a power is equal to a times the log of the number:

$$\log(x^a) = a \cdot \log(x).$$

Derivatives of Natural Logarithms

Let $u(x) = \log(x)$, and write the right side of [A.3.30] as $e^{u(x)}$. Differentiating both sides of [A.3.30] using [A.3.21] reveals that

$$\frac{dx}{dx} = e^{\log(x)} \cdot \frac{d \log(x)}{dx}$$

or

$$1 = x \frac{d \log(x)}{dx}.$$

Thus,

$$\frac{d \log(x)}{dx} = \frac{1}{x}. \qquad [A.3.32]$$

Logarithms and Elasticities

It is sometimes also useful to differentiate a function $f(x)$ with respect to the variable $\log(x)$. To do so, write $f(x)$ as $f(u(x))$, where

$$u(x) = \exp[\log(x)].$$

Now use the chain rule to differentiate:

$$\frac{df(x)}{d\log(x)} = \frac{df}{du}\cdot\frac{du}{d\log(x)}. \qquad [A.3.33]$$

But from [A.3.21],

$$\frac{du}{d\log(x)} = \exp[\log(x)]\frac{d\log(x)}{d\log(x)} = x. \qquad [A.3.34]$$

Substituting [A.3.34] into [A.3.33] gives

$$\frac{df(x)}{d\log(x)} = x\frac{df}{dx}.$$

It follows from [A.3.32] that

$$\frac{d\log f(x)}{d\log x} = \frac{1}{f}x\frac{df}{dx} \cong \frac{[f(x+\Delta) - f(x)]/f(x)}{[(x+\Delta) - x]/x},$$

which has the interpretation as the *elasticity* of f with respect to x, or the percent change in f resulting from a 1% increase in x.

Logarithms and Percent

An approximation to the natural log function is obtained from a first-order Taylor series around $c = 1$:

$$\log(1+\Delta) \cong \log(1) + \left.\frac{d\log(x)}{dx}\right|_{x=1}\cdot\Delta. \qquad [A.3.35]$$

But $\log(1) = 0$, and

$$\left.\frac{d\log(x)}{dx}\right|_{x=1} = \left.\frac{1}{x}\right|_{x=1} = 1.$$

Thus, for Δ close to zero, an excellent approximation is provided by

$$\log(1+\Delta) \cong \Delta. \qquad [A.3.36]$$

An implication of [A.3.36] is the following. Let r denote the net interest rate measured as a fraction of 1; for example, $r = 0.05$ corresponds to a 5% interest rate. Then $(1 + r)$ denotes the gross interest rate (principal plus net interest). Equation [A.3.36] says that the log of the gross interest rate $(1 + r)$ is essentially the same number as the net interest rate (r).

Definition of Indefinite Integral

Integration (indicated by $\int dx$) is the inverse operation from differentiation. For example,

$$\int x\,dx = x^2/2, \qquad [A.3.37]$$

because

$$\frac{d(x^2/2)}{dx} = x. \qquad [A.3.38]$$

The function $x^2/2$ is not the only function satisfying [A.3.38]; the function

$$(x^2/2) + C$$

also works for any constant C. The term C is referred to as the *constant of integration*.

Some Useful Indefinite Integrals

The following integrals can be confirmed from [A.3.1], [A.3.32], [A.3.2], [A.3.3], and [A.3.21]:

$$\int x^k \, dx = \frac{x^{k+1}}{k+1} + C \qquad k \neq -1 \qquad [A.3.39]$$

$$\int x^{-1} \, dx = \begin{cases} \log(x) + C & x > 0 \\ \log(-x) + C & x < 0 \end{cases} \qquad [A.3.40]$$

$$\int \cos(x) \, dx = \sin(x) + C \qquad [A.3.41]$$

$$\int \sin(x) \, dx = -\cos(x) + C \qquad [A.3.42]$$

$$\int e^{ax} \, dx = (1/a) \cdot e^{ax} + C. \qquad [A.3.43]$$

It is also straightforward to demonstrate that for constants a and b not depending on x,

$$\int [a \cdot f(x) + b \cdot g(x)] \, dx = a \int f(x) \, dx + b \int g(x) \, dx + C.$$

Definite Integrals

Consider the continuous function $f(x)$ plotted in Figure A.5. Define the function $A(x; a)$ to be the area under $f(x)$ between a and x, viewed as a function of x. Thus, $A(b; a)$ would be the area between a and b. Suppose we increase b by a small amount Δ. This is approximately the same as adding a rectangle of height $f(b)$ and width Δ to the area $A(b; a)$:

$$A(b + \Delta; a) \cong A(b; a) + f(b) \cdot \Delta,$$

or

$$\frac{A(b + \Delta; a) - A(b; a)}{\Delta} \cong f(b).$$

In the limit as $\Delta \to 0$,

$$\left. \frac{dA(x; a)}{dx} \right|_{x=b} = f(b). \qquad [A.3.44]$$

Now, [A.3.44] has to hold for any value of $b > a$ that we might have chosen,

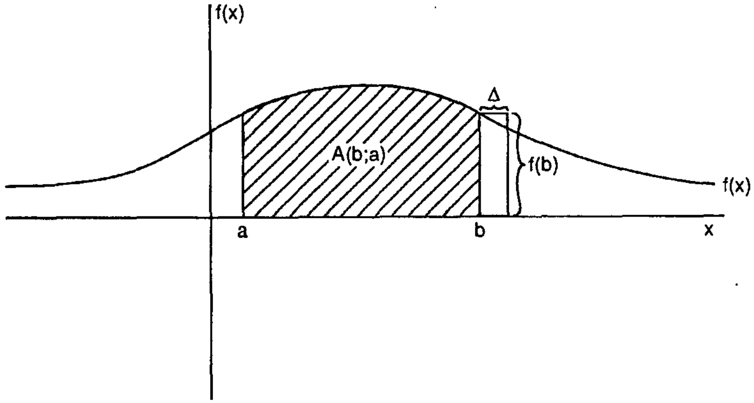


FIGURE A.5 The definite integral as the area under a function.

implying that the area function $A(x; a)$ is the inverse of differentiation:

$$A(x; a) = F(x) + C, \qquad [A.3.45]$$

where

$$\frac{dF(x)}{dx} = f(x).$$

To find the value of C, notice that $A(a; a)$ in [A.3.45] should be equal to zero:

$$A(a; a) = 0 = F(a) + C.$$

For this to be true,

$$C = -F(a). \qquad [A.3.46]$$

Evaluating [A.3.45] at $x = b$, the area between a and b is given by

$$A(b; a) = F(b) + C;$$

or using [A.3.46],

$$A(b; a) = F(b) - F(a), \qquad [A.3.47]$$

where $F(x)$ satisfies $dF/dx = f(x)$:

$$F(x) = \int f(x)\, dx.$$

Equation [A.3.47] is known as the *fundamental theorem of calculus*.

The operation in [A.3.47] is known as calculating a *definite integral*:

$$\int_a^b f(x)\, dx \equiv \left[\int f(x)\, dx\right]\Bigg|_{x=b} - \left[\int f(x)\, dx\right]\Bigg|_{x=a}.$$

For example, to find the area under the sine function between $\theta = 0$ and $\theta = \pi/2$, we use [A.3.42]:

$$\int_0^{\pi/2} \sin(x)\,dx = [-\cos(x)]|_{x=\pi/2} - [-\cos(x)]|_{x=0}$$
$$= [-\cos(\pi/2)] + [\cos(0)]$$
$$= 0 + 1$$
$$= 1.$$

To find the area between 0 and 2π, we take

$$\int_0^{2\pi} \sin(x)\,dx = [-\cos(2\pi)] + \cos(0)$$
$$= -1 + 1$$
$$= 0.$$

The positive values for $\sin(x)$ between 0 and π exactly cancel out the negative values between π and 2π.

A.4. *Matrix Algebra*

Definitions

An $(m \times n)$ matrix is an array of numbers ordered into m rows and n columns:

$$\underset{(m\times n)}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

If there is only one column ($n = 1$), then $\mathbf{A}$ is described as a *column vector*, whereas with only one row ($m = 1$), $\mathbf{A}$ is called a *row vector*. A single number ($n = 1$ and $m = 1$) is called a *scalar*.

If the number of rows equals the number of columns ($m = n$), the matrix is said to be *square*. The diagonal running through $(a_{11}, a_{22}, \ldots, a_{nn})$ in a square matrix is called the *principal diagonal*. If all elements off the principal diagonal are zero, the matrix is said to be *diagonal*.

A matrix is sometimes specified by describing the element in row i, column j:

$$\mathbf{A} = [a_{ij}].$$

Summation and Multiplication

Two $(m \times n)$ matrices are added element by element:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix};$$

or, more compactly,

$$\underset{(m\times n)}{\mathbf{A}} + \underset{(m\times n)}{\mathbf{B}} = [a_{ij} + b_{ij}].$$

The product of an $(m \times n)$ matrix and an $(n \times q)$ matrix is an $(m \times q)$ matrix:

$$\underset{(m\times n)}{\mathbf{A}} \times \underset{(n\times q)}{\mathbf{B}} = \underset{(m\times q)}{\mathbf{C}},$$

where the row i, column j element of $\mathbf{C}$ is given by $\Sigma_{k=1}^{n} a_{ik} b_{kj}$. Notice that multiplication requires that the number of columns of $\mathbf{A}$ be the same as the number of rows of $\mathbf{B}$.

To multiply $\mathbf{A}$ by a scalar α, each element of $\mathbf{A}$ is multiplied by α:

$$\underset{(1\times 1)}{\alpha} \times \underset{(m\times n)}{\mathbf{A}} = \underset{(m\times n)}{\mathbf{C}},$$

with

$$\mathbf{C} = [\alpha a_{ij}].$$

It is easy to show that addition is commutative:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A};$$

whereas multiplication is not:

$$\mathbf{AB} \neq \mathbf{BA}.$$

Indeed, the product $\mathbf{BA}$ will not exist unless $m = q$, and even where it exists, $\mathbf{AB}$ would be equal to $\mathbf{BA}$ only in rather special cases.

Both addition and multiplication are associative:

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$
$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}).$$

Identity Matrix

The *identity matrix* of order n (denoted $\mathbf{I}_n$) is an $(n \times n)$ matrix with 1s along the principal diagonal and 0s elsewhere:

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

For any $(m \times n)$ matrix $\mathbf{A}$,

$$\mathbf{A} \times \mathbf{I}_n = \mathbf{A}$$

and also

$$\mathbf{I}_m \times \mathbf{A} = \mathbf{A}.$$

Powers of Matrices

For an $(n \times n)$ matrix $\mathbf{A}$, the expression $\mathbf{A}^2$ denotes $\mathbf{A} \cdot \mathbf{A}$. The expression $\mathbf{A}^k$ indicates the matrix $\mathbf{A}$ multiplied by itself k times, with $\mathbf{A}^0$ interpreted as the $(n \times n)$ identity matrix.

Transposition

Let a_{ij} denote the row i, column j element of a matrix $\mathbf{A}$:

$$\mathbf{A} = [a_{ij}].$$

The *transpose* of $\mathbf{A}$ (denoted $\mathbf{A}'$) is given by

$$\mathbf{A}' = [a_{ji}].$$

For example, the transpose of

$$\begin{bmatrix} 2 & 4 & 6 \\ 3 & 5 & 7 \\ 1 & 2 & 3 \end{bmatrix}$$

is

$$\begin{bmatrix} 2 & 3 & 1 \\ 4 & 5 & 2 \\ 6 & 7 & 3 \end{bmatrix}.$$

The transpose of a row vector is a column vector.

It is easy to verify the following:

$$(\mathbf{A}')' = \mathbf{A} \qquad [A.4.1]$$

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}' \qquad [A.4.2]$$

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'. \qquad [A.4.3]$$

Symmetric Matrices

A square matrix satisfying $\mathbf{A} = \mathbf{A}'$ is said to be *symmetric*.

Trace of a Matrix

The *trace* of an $(n \times n)$ matrix is defined as the sum of the elements along the principal diagonal:

$$\text{trace}(\mathbf{A}) \equiv a_{11} + a_{22} + \cdots + a_{nn}.$$

If $\mathbf{A}$ is an $(m \times n)$ matrix and $\mathbf{B}$ is an $(n \times m)$ matrix, then $\mathbf{AB}$ is an $(m \times m)$ matrix whose trace is

$$\text{trace}(\mathbf{AB}) = \sum_{j=1}^{n} a_{1j}b_{j1} + \sum_{j=1}^{n} a_{2j}b_{j2} + \cdots + \sum_{j=1}^{n} a_{mj}b_{jm} = \sum_{k=1}^{m}\sum_{j=1}^{n} a_{kj}b_{jk}.$$

The product $\mathbf{BA}$ is an $(n \times n)$ matrix whose trace is

$$\text{trace}(\mathbf{BA}) = \sum_{k=1}^{m} b_{1k}a_{k1} + \sum_{k=1}^{m} b_{2k}a_{k2} + \cdots + \sum_{k=1}^{m} b_{nk}a_{kn} = \sum_{j=1}^{n}\sum_{k=1}^{m} b_{jk}a_{kj}.$$

Thus,

$$\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA}).$$

If $\mathbf{A}$ and $\mathbf{B}$ are both $(n \times n)$ matrices, then

$$\text{trace}(\mathbf{A} + \mathbf{B}) = \text{trace}(\mathbf{A}) + \text{trace}(\mathbf{B}).$$

If $\mathbf{A}$ is an $(n \times n)$ matrix and λ is a scalar, then

$$\text{trace}(\lambda\mathbf{A}) = \sum_{i=1}^{n} \lambda a_{ii} = \lambda \cdot \sum_{i=1}^{n} a_{ii} = \lambda \cdot \text{trace}(\mathbf{A}).$$

Partitioned Matrices

A *partitioned matrix* is a matrix whose individual elements are themselves matrices. For example, the (3×4) matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

could be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{a}_1' & \mathbf{a}_2' \end{bmatrix}$$

where

$$\mathbf{A}_1 \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad \mathbf{A}_2 \equiv \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix}$$

$$\mathbf{a}_1' \equiv \begin{bmatrix} a_{31} & a_{32} \end{bmatrix} \qquad \mathbf{a}_2' \equiv \begin{bmatrix} a_{33} & a_{34} \end{bmatrix}.$$

Partitioned matrices are added or multiplied as if the individual elements were scalars, provided that the row and column dimensions permit the appropriate matrix operations. For example,

$$\begin{bmatrix} \underset{(m_1 \times n_1)}{\mathbf{A}_1} & \underset{(m_1 \times n_2)}{\mathbf{A}_2} \\ \underset{(m_2 \times n_1)}{\mathbf{A}_3} & \underset{(m_2 \times n_2)}{\mathbf{A}_4} \end{bmatrix} + \begin{bmatrix} \underset{(m_1 \times n_1)}{\mathbf{B}_1} & \underset{(m_1 \times n_2)}{\mathbf{B}_2} \\ \underset{(m_2 \times n_1)}{\mathbf{B}_3} & \underset{(m_2 \times n_2)}{\mathbf{B}_4} \end{bmatrix} = \begin{bmatrix} \underset{(m_1 \times n_1)}{\mathbf{A}_1 + \mathbf{B}_1} & \underset{(m_1 \times n_2)}{\mathbf{A}_2 + \mathbf{B}_2} \\ \underset{(m_2 \times n_1)}{\mathbf{A}_3 + \mathbf{B}_3} & \underset{(m_2 \times n_2)}{\mathbf{A}_4 + \mathbf{B}_4} \end{bmatrix}.$$

Similarly,

$$\begin{bmatrix} \underset{(m_1 \times n_1)}{\mathbf{A}_1} & \underset{(m_1 \times n_2)}{\mathbf{A}_2} \\ \underset{(m_2 \times n_1)}{\mathbf{A}_3} & \underset{(m_2 \times n_2)}{\mathbf{A}_4} \end{bmatrix} \times \begin{bmatrix} \underset{(n_1 \times q_1)}{\mathbf{B}_1} & \underset{(n_1 \times q_2)}{\mathbf{B}_2} \\ \underset{(n_2 \times q_1)}{\mathbf{B}_3} & \underset{(n_2 \times q_2)}{\mathbf{B}_4} \end{bmatrix} = \begin{bmatrix} \underset{(m_1 \times q_1)}{\mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_3} & \underset{(m_1 \times q_2)}{\mathbf{A}_1\mathbf{B}_2 + \mathbf{A}_2\mathbf{B}_4} \\ \underset{(m_2 \times q_1)}{\mathbf{A}_3\mathbf{B}_1 + \mathbf{A}_4\mathbf{B}_3} & \underset{(m_2 \times q_2)}{\mathbf{A}_3\mathbf{B}_2 + \mathbf{A}_4\mathbf{B}_4} \end{bmatrix}.$$

Definition of Determinant

The *determinant* of a 2×2 matrix is given by the following scalar:

$$|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}. \qquad \text{[A.4.4]}$$

The determinant of an $n \times n$ matrix can be defined recursively. Let $\mathbf{A}_{ij}$ denote the $(n-1) \times (n-1)$ matrix formed by deleting row i and column j from $\mathbf{A}$. The determinant of $\mathbf{A}$ is given by

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} |\mathbf{A}_{1j}|. \qquad \text{[A.4.5]}$$

For example, the determinant of a 3×3 matrix is

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Properties of Determinants

A square matrix is said to be *lower triangular* if all the elements above the principal diagonal are zero ($a_{ij} = 0$ for $j > i$):

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}.$$

The determinant of a lower triangular matrix is simply the product of the terms along the principal diagonal:

$$|\mathbf{A}| = a_{11}a_{22} \cdots a_{nn}. \qquad [A.4.6]$$

That [A.4.6] holds for $n = 2$ follows immediately from [A.4.4] Given that it holds for a matrix of order $n - 1$, equation [A.4.5] implies that it holds for n:

$$|\mathbf{A}| = a_{11} \begin{vmatrix} a_{22} & 0 & 0 & \cdots & 0 \\ a_{32} & a_{33} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nn} \end{vmatrix} + 0 \cdot |\mathbf{A}_{12}| + \cdots + 0 \cdot |\mathbf{A}_{1n}|.$$

An immediate implication of [A.4.6] is that the determinant of the identity matrix is unity:

$$|\mathbf{I}_n| = 1. \qquad [A.4.7]$$

Another useful fact about determinants is that if an $n \times n$ matrix $\mathbf{A}$ is multiplied by a scalar α, the effect is to multiply the determinant by α^n:

$$|\alpha\mathbf{A}| = \alpha^n|\mathbf{A}|. \qquad [A.4.8]$$

Again, [A.4.8] is immediately apparent for the $n = 2$ case from [A.4.4]:

$$\begin{aligned} |\alpha\mathbf{A}| &= \begin{vmatrix} \alpha a_{11} & \alpha a_{12} \\ \alpha a_{21} & \alpha a_{22} \end{vmatrix} \\ &= (\alpha a_{11}\alpha a_{22}) - (\alpha a_{12}\alpha a_{21}) \\ &= \alpha^2(a_{11}a_{22} - a_{12}a_{21}) \\ &= \alpha^2|\mathbf{A}|. \end{aligned}$$

Given that it holds for $n - 1$, it is simple to verify for n using [A.4.5].

By contrast, if a single row of $\mathbf{A}$ is multiplied by the constant α (as opposed to multiplying the entire matrix by α), then the determinant is multiplied by α. If the row that is multiplied by α is the first row, then this result is immediately apparent from [A.4.5]. If only the ith row of $\mathbf{A}$ is multiplied by α, the result can be shown by recursively applying [A.4.5] until the elements of the ith row appear explicitly in the formula.

Suppose that some constant c times the second row of a 2×2 matrix is added to the first row. This operation has no effect on the determinant:

$$\begin{aligned} \begin{vmatrix} a_{11} + ca_{21} & a_{12} + ca_{22} \\ a_{21} & a_{22} \end{vmatrix} &= (a_{11} + ca_{21})a_{22} - (a_{12} + ca_{22})a_{21} \\ &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned}$$

Similarly, if some constant c times the third row of a 3×3 matrix is added to the second row, the determinant will again be unchanged:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} + ca_{31} & a_{22} + ca_{32} & a_{23} + ca_{33} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11}\begin{vmatrix} a_{22} + ca_{32} & a_{23} + ca_{33} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} + ca_{31} & a_{23} + ca_{33} \\ a_{31} & a_{33} \end{vmatrix}$$

$$+ a_{13}\begin{vmatrix} a_{21} + ca_{31} & a_{22} + ca_{32} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

In general, if any row of an $n \times n$ matrix is multiplied by c and added to another row, the new matrix will have the same determinant as the original. Similarly, multiplying any column by c and adding the result to another column will not change the determinant.

This can be viewed as a special case of the following result. If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times n$ matrices, then

$$|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|. \qquad [\text{A.4.9}]$$

Adding c times the second column of a 2×2 matrix $\mathbf{A}$ to the first column can be thought of as postmultiplying $\mathbf{A}$ by the following matrix:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}.$$

Since $\mathbf{B}$ is lower triangular with 1s along the principal diagonal, its determinant is unity, and so, from [A.4.9],

$$|\mathbf{AB}| = |\mathbf{A}|.$$

Thus, the fact that adding a multiple of one column to another does not alter the determinant can be viewed as an implication of [A.4.9].

If two rows of a matrix are switched, the determinant changes signs. To switch the ith row with the jth, multiply the ith row by -1; this changes the sign of the determinant. Then subtract row i from row j, add the new j back to i, and subtract i from j once again. These last operations complete the switch and do not affect the determinant further. For example, let $\mathbf{A}$ be a (4×4) matrix written in partitioned form as

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \\ \mathbf{a}_4' \end{bmatrix},$$

where the (1×4) vector $\mathbf{a}_i'$ represents the ith row of $\mathbf{A}$. The determinant when rows 1 and 4 are switched can be calculated from

$$\begin{vmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \\ \mathbf{a}_4' \end{vmatrix} = -\begin{vmatrix} -\mathbf{a}_1' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \\ \mathbf{a}_4' \end{vmatrix} = -\begin{vmatrix} -\mathbf{a}_1' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \\ \mathbf{a}_1' + \mathbf{a}_4' \end{vmatrix} = -\begin{vmatrix} \mathbf{a}_4' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \\ \mathbf{a}_1' + \mathbf{a}_4' \end{vmatrix} = -\begin{vmatrix} \mathbf{a}_4' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \\ \mathbf{a}_1' \end{vmatrix}.$$

This result permits calculation of the determinant of **A** in reference to any row of an $(n \times n)$ matrix **A**:

$$|\mathbf{A}| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |\mathbf{A}_{ij}|. \qquad \text{[A.4.10]}$$

To derive [A.4.10], define $\mathbf{A}^*$ as

$$\mathbf{A}^* \equiv \begin{bmatrix} \mathbf{a}'_i \\ \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_{i-1} \\ \mathbf{a}'_{i+1} \\ \vdots \\ \mathbf{a}'_n \end{bmatrix}.$$

Then, from [A.4.5],

$$|\mathbf{A}^*| = \sum_{j=1}^{n} (-1)^{j+1} a^*_{1j} |\mathbf{A}^*_{1j}| = \sum_{j=1}^{n} (-1)^{j+1} a_{ij} |\mathbf{A}_{ij}|.$$

Moreover, $\mathbf{A}^*$ is obtained from **A** by $(i - 1)$ row switches, such as switching i with $i - 1$, $i - 1$ with $i - 2$, . . . , and 2 with 1. Hence,

$$|\mathbf{A}| = (-1)^{i-1} |\mathbf{A}^*| = (-1)^{i-1} \sum_{j=1}^{n} (-1)^{j+1} a_{ij} |\mathbf{A}_{ij}|,$$

as claimed in [A.4.10].

An immediate implication of [A.4.10] is that if any row of a matrix contains all zeros, then the determinant of the matrix is zero.

It can also be shown that the transpose of a matrix has the same determinant as the original matrix:

$$|\mathbf{A}'| = |\mathbf{A}|. \qquad \text{[A.4.11]}$$

This means, for example, that if the kth column of a matrix consists entirely of zeros, then the determinant of the matrix is zero. It also implies that the determinant of an *upper triangular matrix* (one for which $a_{ij} = 0$ for all $j < i$) is the product of the terms on the principal diagonal.

Adjoint of a Matrix

Let **A** denote an $(n \times n)$ matrix, and as before let $\mathbf{A}_{ji}$ denote the $[(n - 1) \times (n - 1)]$ matrix that results from deleting row j and column i of **A**. The *adjoint* of **A** is the $(n \times n)$ matrix whose row i, column j element is given by $(-1)^{i+j} |\mathbf{A}_{ji}|$.

Inverse of a Matrix

If the determinant of an $n \times n$ matrix **A** is not equal to zero, its *inverse* (an $n \times n$ matrix denoted $\mathbf{A}^{-1}$) exists and is found by dividing the adjoint by the determinant:

$$\mathbf{A}^{-1} = (1/|\mathbf{A}|) \cdot [(-1)^{i+j} |\mathbf{A}_{ji}|]. \qquad \text{[A.4.12]}$$

For example, for $n = 2$,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = (1/[a_{11}a_{22} - a_{12}a_{21}]) \cdot \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \qquad [A.4.13]$$

A matrix whose inverse exists is said to be *nonsingular*. A matrix whose determinant is zero is *singular* and has no inverse.

When an inverse exists,

$$\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{I}_n. \qquad [A.4.14]$$

Taking determinants of both sides of [A.4.14] and using [A.4.9] and [A.4.7],

$$|\mathbf{A}| \cdot |\mathbf{A}^{-1}| = 1,$$

so

$$|\mathbf{A}^{-1}| = 1/|\mathbf{A}|. \qquad [A.4.15]$$

Alternatively, taking the transpose of both sides of [A.4.14] and recalling [A.4.3],

$$(\mathbf{A}^{-1})'\mathbf{A}' = \mathbf{I}_n,$$

which means that $(\mathbf{A}^{-1})'$ is the inverse of $\mathbf{A}'$:

$$(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}.$$

For α a nonzero scalar and $\mathbf{A}$ a nonsingular matrix,

$$[\alpha\mathbf{A}]^{-1} = \alpha^{-1}\mathbf{A}^{-1}.$$

Also, for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ all nonsingular $(n \times n)$ matrices,

$$[\mathbf{AB}]^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

and

$$[\mathbf{ABC}]^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Linear Dependence

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ be a set of k different $(n \times 1)$ vectors. The vectors are said to be *linearly dependent* if there exists a set of k scalars $(c_1, c_2, \ldots, c_k)$, not all of which are zero, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k = \mathbf{0}.$$

If no such set of nonzero numbers $(c_1, c_2, \ldots, c_k)$ exists, then the vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ are said to be *linearly independent.*

Suppose the vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ are collected in an $(n \times k)$ matrix $\mathbf{T}$, written in partitioned form as

$$\mathbf{T} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_k].$$

If the number of vectors (k) is equal to the dimension of each vector (n), then there is a simple relation between the notion of linear dependence and the determinant of the $(n \times n)$ matrix $\mathbf{T}$; specifically, if $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ are linearly dependent, then $|\mathbf{T}| = 0$. To see this, suppose that $\mathbf{x}_1$ is one of the vectors that have a nonzero value of c_i. Then linear dependence means that

$$\mathbf{x}_1 = -(c_2/c_1)\mathbf{x}_2 - (c_3/c_1)\mathbf{x}_3 - \cdots - (c_n/c_1)\mathbf{x}_n.$$

Then the determinant of $\mathbf{T}$ is equal to

$$|\mathbf{T}| = |[-(c_2/c_1)\mathbf{x}_2 - (c_3/c_1)\mathbf{x}_3 - \cdots - (c_n/c_1)\mathbf{x}_n] \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n|.$$

But if we add (c_n/c_1) times the nth column to the first column, (c_{n-1}/c_1) times the $(n - 1)$th column to the first column, . . . , and (c_2/c_1) times the second column to the first column, the result is

$$\begin{aligned}|\mathbf{T}| &= |\mathbf{0} \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n| \\ &= 0.\end{aligned}$$

The converse can also be shown to be true: if $|\mathbf{T}| = 0$, then $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ are linearly dependent.

Eigenvalues and Eigenvectors

Suppose that an $n \times n$ matrix $\mathbf{A}$, a nonzero $n \times 1$ vector $\mathbf{x}$, and a scalar λ are related by

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \qquad [A.4.16]$$

Then $\mathbf{x}$ is called an *eigenvector* of $\mathbf{A}$ and λ the associated *eigenvalue*. Equation [A.4.16] can be written

$$\mathbf{A}\mathbf{x} - \lambda\mathbf{I}_n\mathbf{x} = \mathbf{0}$$

or

$$(\mathbf{A} - \lambda\mathbf{I}_n)\mathbf{x} = \mathbf{0}. \qquad [A.4.17]$$

Suppose that the matrix $(\mathbf{A} - \lambda\mathbf{I}_n)$ were nonsingular. Then $(\mathbf{A} - \lambda\mathbf{I}_n)^{-1}$ would exist and we could premultiply [A.4.17] by $(\mathbf{A} - \lambda\mathbf{I}_n)^{-1}$ to deduce that

$$\mathbf{x} = \mathbf{0}.$$

Thus, if a nonzero vector $\mathbf{x}$ exists that satisfies [A.4.16], then it must be associated with a value of λ such that $(\mathbf{A} - \lambda\mathbf{I}_n)$ is singular. An eigenvalue of the matrix $\mathbf{A}$ is therefore a number λ such that

$$|\mathbf{A} - \lambda\mathbf{I}_n| = 0. \qquad [A.4.18]$$

Eigenvalues of Triangular Matrices

Notice that if $\mathbf{A}$ is upper triangular or lower triangular, then $\mathbf{A} - \lambda\mathbf{I}_n$ is as well, and its determinant is just the product of terms along the principal diagonal:

$$|\mathbf{A} - \lambda\mathbf{I}_n| = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

Thus, for a triangular matrix, the eigenvalues (the values of λ for which this expression equals zero) are just the values of $\mathbf{A}$ along the principal diagonal.

Linear Independence of Eigenvectors

A useful result is that if the eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are all distinct, then the associated eigenvectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ are linearly independent. To see this for the case $n = 2$, consider any numbers c_1 and c_2 such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = \mathbf{0}. \qquad [A.4.19]$$

Premultiplying both sides of [A.4.19] by $\mathbf{A}$ produces

$$c_1\mathbf{A}\mathbf{x}_1 + c_2\mathbf{A}\mathbf{x}_2 = c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 = \mathbf{0}. \qquad [A.4.20]$$

If [A.4.19] is multiplied by λ_1 and subtracted from [A.4.20], the result is

$$c_2(\lambda_2 - \lambda_1)\mathbf{x}_2 = \mathbf{0}. \qquad [A.4.21]$$

But $\mathbf{x}_2$ is an eigenvector of $\mathbf{A}$, and so it cannot be the zero vector. Also, $\lambda_2 - \lambda_1$ cannot be zero, since $\lambda_2 \neq \lambda_1$. Equation [A.4.21] therefore implies that $c_2 = 0$. A parallel set of calculations show that $c_1 = 0$. Thus, the only values of c_1 and c_2 consistent with [A.4.19] are $c_1 = 0$ and $c_2 = 0$, which means that $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent. A similar argument for $n > 2$ can be made by induction.

A Useful Decomposition

Suppose an $n \times n$ matrix $\mathbf{A}$ has n distinct eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Collect these in a diagonal matrix $\mathbf{\Lambda}$:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Collect the eigenvectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ in an $(n \times n)$ matrix $\mathbf{T}$:

$$\mathbf{T} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n].$$

Applying the formula for multiplying partitioned matrices,

$$\mathbf{AT} = [\mathbf{A}\mathbf{x}_1 \quad \mathbf{A}\mathbf{x}_2 \quad \cdots \quad \mathbf{A}\mathbf{x}_n].$$

But since $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ are eigenvectors, equation [A.4.16] implies that

$$\mathbf{AT} = [\lambda_1\mathbf{x}_1 \quad \lambda_2\mathbf{x}_2 \quad \cdots \quad \lambda_n\mathbf{x}_n]. \qquad [A.4.22]$$

A second application of the formula for multiplying partitioned matrices shows that the right side of [A.4.22] is in turn equal to

$$\begin{aligned} &[\lambda_1\mathbf{x}_1 \quad \lambda_2\mathbf{x}_2 \quad \cdots \quad \lambda_n\mathbf{x}_n] \\ &= [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\ &= \mathbf{T\Lambda}. \end{aligned}$$

Thus, [A.4.22] can be written

$$\mathbf{AT} = \mathbf{T\Lambda}. \qquad [A.4.23]$$

Now, since the eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are taken to be distinct, the eigenvectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ are known to be linearly independent. Thus, $|\mathbf{T}| \neq 0$ and $\mathbf{T}^{-1}$ exists. Postmultiplying [A.4.23] by $\mathbf{T}^{-1}$ reveals a useful decomposition of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{T\Lambda T}^{-1}. \qquad [A.4.24]$$

The Jordan Decomposition

The decomposition in [A.4.24] required the $(n \times n)$ matrix $\mathbf{A}$ to have n linearly independent eigenvectors. This will be true whenever $\mathbf{A}$ has n distinct eigenvalues, and could still be true even if $\mathbf{A}$ has some repeated eigenvalues. In the completely general case when $\mathbf{A}$ has $s \leq n$ linearly independent eigenvectors, there always exists a decomposition similar to [A.4.24], known as the *Jordan decomposition*. Specifically, for such a matrix $\mathbf{A}$ there exists a nonsingular $(n \times n)$ matrix $\mathbf{M}$ such that

$$\mathbf{A} = \mathbf{M}\mathbf{J}\mathbf{M}^{-1}, \qquad [\text{A.4.25}]$$

where the $(n \times n)$ matrix $\mathbf{J}$ takes the form

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_s \end{bmatrix} \qquad [\text{A.4.26}]$$

with

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}. \qquad [\text{A.4.27}]$$

Thus, $\mathbf{J}_i$ has the eigenvalue λ_i repeated along the principal diagonal and has unity repeated along the diagonal above the principal diagonal. The same eigenvalue λ_i can appear in two different Jordan blocks $\mathbf{J}_i$ and $\mathbf{J}_k$ if it corresponds to several linearly independent eigenvectors.

Some Further Results on Eigenvalues

Suppose that λ is an eigenvalue of the $(n \times n)$ matrix $\mathbf{A}$. Then λ is also an eigenvalue of $\mathbf{SAS}^{-1}$ for any nonsingular $(n \times n)$ matrix $\mathbf{S}$. To see this, note that

$$(\mathbf{A} - \lambda\mathbf{I}_n)\mathbf{x} = \mathbf{0}$$

implies that

$$\mathbf{S}(\mathbf{A} - \lambda\mathbf{I}_n)\mathbf{S}^{-1}\mathbf{S}\mathbf{x} = \mathbf{0}$$

or

$$(\mathbf{SAS}^{-1} - \lambda\mathbf{I}_n)\mathbf{x}^* = \mathbf{0} \qquad [\text{A.4.28}]$$

for $\mathbf{x}^* \equiv \mathbf{Sx}$. Thus, λ is an eigenvalue of $\mathbf{SAS}^{-1}$ associated with the eigenvector $\mathbf{x}^*$.

From [A.4.25], this implies that the determinant of any $(n \times n)$ matrix $\mathbf{A}$ is the same as the determinant of its Jordan matrix $\mathbf{J}$ defined in [A.4.26]. Since $\mathbf{J}$ is upper triangular, its determinant is the product of terms along the principal diagonal, which were just the eigenvalues of $\mathbf{A}$. Thus, the determinant of any matrix $\mathbf{A}$ is given by the product of its eigenvalues.

It is also clear that the eigenvalues of $\mathbf{A}$ are the same as those of $\mathbf{A}'$. Taking the transpose of [A.4.25],

$$\mathbf{A}' = (\mathbf{M}')^{-1}\mathbf{J}'\mathbf{M}',$$

we see that the eigenvalues of $\mathbf{A}'$ are the eigenvalues of $\mathbf{J}'$. Since $\mathbf{J}'$ is lower triangular, its eigenvalues are the elements on its principal diagonal. But $\mathbf{J}'$ and $\mathbf{J}$ have the same principal diagonal, meaning that $\mathbf{A}'$ and $\mathbf{A}$ have the same eigenvalues.

Matrix Geometric Series

The results of [A.3.6] through [A.3.10] generalize readily to geometric series involving square matrices. Consider the sum

$$\mathbf{S}_T = \mathbf{I}_n + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \cdots + \mathbf{A}^T \qquad [A.4.29]$$

for $\mathbf{A}$ an $(n \times n)$ matrix. Premultiplying both sides of [A.4.29] by $\mathbf{A}$, we see that

$$\mathbf{A}\mathbf{S}_T = \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \cdots + \mathbf{A}^T + \mathbf{A}^{T+1}. \qquad [A.4.30]$$

Subtracting [A.4.30] from [A.4.29], we find that

$$(\mathbf{I}_n - \mathbf{A})\mathbf{S}_T = \mathbf{I}_n - \mathbf{A}^{T+1}. \qquad [A.4.31]$$

Notice from [A.4.18] that if $|\mathbf{I}_n - \mathbf{A}| = 0$, then $\lambda = 1$ would be an eigenvalue of $\mathbf{A}$. Assuming that none of the eigenvalues of $\mathbf{A}$ is equal to unity, the matrix $(\mathbf{I}_n - \mathbf{A})$ is nonsingular and [A.4.31] implies that

$$\mathbf{S}_T = (\mathbf{I}_n - \mathbf{A})^{-1}(\mathbf{I}_n - \mathbf{A}^{T+1}) \qquad [A.4.32]$$

if no eigenvalue of $\mathbf{A}$ equals 1. If all the eigenvalues of $\mathbf{A}$ are strictly less than 1 in modulus, it can be shown that $\mathbf{A}^{T+1} \to \mathbf{0}$ as $T \to \infty$, implying that

$$(\mathbf{I}_n + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \cdots) = (\mathbf{I}_n - \mathbf{A})^{-1} \qquad [A.4.33]$$

assuming that the eigenvalues of $\mathbf{A}$ are all inside unit circle.

Kronecker Products

For $\mathbf{A}$ an $(m \times n)$ matrix and $\mathbf{B}$ a $(p \times q)$ matrix, the *Kronecker product* of $\mathbf{A}$ and $\mathbf{B}$ is defined as the following $(mp) \times (nq)$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}.$$

The following properties of the Kronecker product are readily verified. For any matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$,

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}' \qquad [A.4.34]$$

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}). \qquad [A.4.35]$$

Also, for $\mathbf{A}$ and $\mathbf{B}$ both $(m \times n)$ matrices and $\mathbf{C}$ any matrix,

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C}) \qquad [A.4.36]$$

$$\mathbf{C} \otimes (\mathbf{A} + \mathbf{B}) = (\mathbf{C} \otimes \mathbf{A}) + (\mathbf{C} \otimes \mathbf{B}). \qquad [A.4.37]$$

Let $\mathbf{A}$ be $(m \times n)$, $\mathbf{B}$ be $(p \times q)$, $\mathbf{C}$ be $(n \times k)$, and $\mathbf{D}$ be $(q \times r)$. Then

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}); \qquad [A.4.38]$$

that is,

$$\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \begin{bmatrix} c_{11}\mathbf{D} & c_{12}\mathbf{D} & \cdots & c_{1k}\mathbf{D} \\ c_{21}\mathbf{D} & c_{22}\mathbf{D} & \cdots & c_{2k}\mathbf{D} \\ \vdots & \vdots & \cdots & \vdots \\ c_{n1}\mathbf{D} & c_{n2}\mathbf{D} & \cdots & c_{nk}\mathbf{D} \end{bmatrix}$$

$$= \begin{bmatrix} \sum a_{1j}c_{j1}\mathbf{BD} & \sum a_{1j}c_{j2}\mathbf{BD} & \cdots & \sum a_{1j}c_{jk}\mathbf{BD} \\ \sum a_{2j}c_{j1}\mathbf{BD} & \sum a_{2j}c_{j2}\mathbf{BD} & \cdots & \sum a_{2j}c_{jk}\mathbf{BD} \\ \vdots & \vdots & \cdots & \vdots \\ \sum a_{mj}c_{j1}\mathbf{BD} & \sum a_{mj}c_{j2}\mathbf{BD} & \cdots & \sum a_{mj}c_{jk}\mathbf{BD} \end{bmatrix}.$$

For $\mathbf{A}$ $(n \times n)$ and $\mathbf{B}$ $(p \times p)$ both nonsingular matrices we can set $\mathbf{C} = \mathbf{A}^{-1}$ and $\mathbf{D} = \mathbf{B}^{-1}$ in [A.4.38] to deduce that

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) = (\mathbf{A}\mathbf{A}^{-1}) \otimes (\mathbf{B}\mathbf{B}^{-1}) = \mathbf{I}_n \otimes \mathbf{I}_p = \mathbf{I}_{np}.$$

Thus,

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}). \qquad \text{[A.4.39]}$$

Eigenvalues of a Kronecker Product

For $\mathbf{A}$ an $(n \times n)$ matrix with (possibly nondistinct) eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $\mathbf{B}$ $(p \times p)$ with eigenvalues $(\mu_1, \mu_2, \ldots, \mu_p)$, then the (np) eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are given by $\lambda_i\mu_j$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$. To see this, write $\mathbf{A}$ and $\mathbf{B}$ in Jordan form as

$$\mathbf{A} = \mathbf{M}_A\mathbf{J}_A\mathbf{M}_A^{-1}$$
$$\mathbf{B} = \mathbf{M}_B\mathbf{J}_B\mathbf{M}_B^{-1}.$$

Then $(\mathbf{M}_A \otimes \mathbf{M}_B)$ has inverse given by $(\mathbf{M}_A^{-1} \otimes \mathbf{M}_B^{-1})$. Moreover, we know from [A.4.28] that the eigenvalues of $(\mathbf{A} \otimes \mathbf{B})$ are the same as the eigenvalues of

$$\begin{aligned}(\mathbf{M}_A^{-1} \otimes \mathbf{M}_B^{-1})(\mathbf{A} \otimes \mathbf{B})(\mathbf{M}_A \otimes \mathbf{M}_B) &= (\mathbf{M}_A^{-1}\mathbf{A}\mathbf{M}_A) \otimes (\mathbf{M}_B^{-1}\mathbf{B}\mathbf{M}_B) \\ &= \mathbf{J}_A \otimes \mathbf{J}_B.\end{aligned}$$

But $\mathbf{J}_A$ and $\mathbf{J}_B$ are both upper triangular, meaning that $(\mathbf{J}_A \otimes \mathbf{J}_B)$ is upper triangular as well. The eigenvalues of $(\mathbf{A} \otimes \mathbf{B})$ are thus just the terms on the principal diagonal of $(\mathbf{J}_A \otimes \mathbf{J}_B)$, which are given by $\lambda_i\mu_j$.

Positive Definite Matrices

An $(n \times n)$ real symmetric matrix $\mathbf{A}$ is said to be *positive semidefinite* if for any real $(n \times 1)$ vector $\mathbf{x}$,

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0.$$

We make the stronger statement that a real symmetric matrix $\mathbf{A}$ is *positive definite* if for any real nonzero $(n \times 1)$ vector $\mathbf{x}$,

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0;$$

hence, any positive definite matrix could also be said to be positive semidefinite.

Let λ be an eigenvalue of $\mathbf{A}$ associated with the eigenvector $\mathbf{x}$:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$

Premultiplying this equation by $\mathbf{x}'$ results in

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda\mathbf{x}'\mathbf{x}.$$

Since an eigenvector $\mathbf{x}$ cannot be the zero vector, $\mathbf{x}'\mathbf{x} > 0$. Thus, for a positive semidefinite matrix $\mathbf{A}$, any eigenvalue λ of $\mathbf{A}$ must be greater than or equal to zero. For $\mathbf{A}$ positive definite, all eigenvalues are strictly greater than zero. Since the determinant of $\mathbf{A}$ is the product of the eigenvalues, the determinant of a positive definite matrix $\mathbf{A}$ is strictly positive.

Let $\mathbf{A}$ be a positive definite $(n \times n)$ matrix and let $\mathbf{B}$ denote a nonsingular $(n \times n)$ matrix. Then $\mathbf{B}'\mathbf{A}\mathbf{B}$ is positive definite. To see this, let $\mathbf{x}$ be any nonzero vector. Define

$$\tilde{\mathbf{x}} \equiv \mathbf{B}\mathbf{x}.$$

Then $\tilde{\mathbf{x}}$ cannot be the zero vector, for if it were, this equation would state that there exists a nonzero vector $\mathbf{x}$ such that

$$\mathbf{B}\mathbf{x} = 0 \cdot \mathbf{x},$$

in which case zero would be an eigenvalue of $\mathbf{B}$ associated with the eigenvector $\mathbf{x}$. But since $\mathbf{B}$ is nonsingular, none of its eigenvalues can be zero. Thus, $\tilde{\mathbf{x}} = \mathbf{B}\mathbf{x}$ cannot be the zero vector, and

$$\mathbf{x}'\mathbf{B}'\mathbf{A}\mathbf{B}\mathbf{x} = \tilde{\mathbf{x}}'\mathbf{A}\tilde{\mathbf{x}} > 0,$$

establishing that the matrix $\mathbf{B}'\mathbf{A}\mathbf{B}$ is positive definite.

A special case of this result is obtained by letting $\mathbf{A}$ be the identity matrix. Then the result implies that any matrix that can be written as $\mathbf{B}'\mathbf{B}$ for some nonsingular matrix $\mathbf{B}$ is positive definite. More generally, any matrix that can be written as $\mathbf{B}'\mathbf{B}$ for an arbitrary matrix $\mathbf{B}$ must be positive semidefinite:

$$\mathbf{x}'\mathbf{B}'\mathbf{B}\mathbf{x} = \tilde{\mathbf{x}}'\tilde{\mathbf{x}} = \tilde{x}_1^2 + \tilde{x}_2^2 + \cdots + \tilde{x}_n^2 \geq 0, \qquad [\text{A.4.40}]$$

where $\tilde{\mathbf{x}} \equiv \mathbf{B}\mathbf{x}$.

The converse propositions are also true: if $\mathbf{A}$ is positive semidefinite, then there exists a matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{B}'\mathbf{B}$; if $\mathbf{A}$ is positive definite, then there exists a nonsingular matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{B}'\mathbf{B}$. A proof of this claim and an algorithm for calculating $\mathbf{B}$ are provided in Section 4.4.

Conjugate Transposes

Let $\mathbf{A}$ denote an $(m \times n)$ matrix of (possibly) complex numbers:

$$\mathbf{A} = \begin{bmatrix} a_{11} + b_{11}i & \cdots & a_{1n} + b_{1n}i \\ a_{21} + b_{21}i & \cdots & a_{2n} + b_{2n}i \\ \vdots & \cdots & \vdots \\ a_{m1} + b_{m1}i & \cdots & a_{mn} + b_{mn}i \end{bmatrix}.$$

The *conjugate transpose* of $\mathbf{A}$, denoted $\mathbf{A}^H$, is formed by transposing $\mathbf{A}$ and replacing each element with its complex conjugate:

$$\mathbf{A}^H = \begin{bmatrix} a_{11} - b_{11}i & \cdots & a_{m1} - b_{m1}i \\ a_{12} - b_{12}i & \cdots & a_{m2} - b_{m2}i \\ \vdots & \cdots & \vdots \\ a_{1n} - b_{1n}i & \cdots & a_{mn} - b_{mn}i \end{bmatrix}.$$

Thus, if $\mathbf{A}$ is real, then $\mathbf{A}^H$ and $\mathbf{A}'$ would denote the same matrix.

Notice that if an $(n \times 1)$ complex vector is premultiplied by its conjugate transpose, the result is a nonnegative real scalar:

$$\mathbf{x}^H\mathbf{x} = [(a_1 - b_1i) \quad (a_2 - b_2i) \quad \cdots \quad (a_n - b_ni)] \begin{bmatrix} a_1 + b_1i \\ a_2 + b_2i \\ \vdots \\ a_n + b_ni \end{bmatrix}$$

$$= \sum_{i=1}^{n} (a_i^2 + b_i^2) \geq 0.$$

For $\mathbf{B}$ a real $(m \times n)$ matrix and $\mathbf{x}$ a complex $(n \times 1)$ vector,

$$(\mathbf{Bx})^H = \mathbf{x}^H\mathbf{B}'.$$

More generally, if both $\mathbf{B}$ and $\mathbf{x}$ are complex,

$$(\mathbf{Bx})^H = \mathbf{x}^H\mathbf{B}^H.$$

Notice that if $\mathbf{A}$ is positive semidefinite, then

$$\mathbf{x}^H\mathbf{A}\mathbf{x} = \mathbf{x}^H\mathbf{B}'\mathbf{B}\mathbf{x} = \tilde{\mathbf{x}}^H\tilde{\mathbf{x}},$$

with $\tilde{\mathbf{x}} \equiv \mathbf{Bx}$. Thus, $\mathbf{x}^H\mathbf{A}\mathbf{x}$ is a nonnegative real scalar for any $\mathbf{x}$ when $\mathbf{A}$ is positive semidefinite. It is a positive real scalar for $\mathbf{A}$ positive definite.

Continuity of Functions of Vectors

A function of more than one argument, such as

$$y = f(x_1, x_2, \ldots, x_n), \qquad \text{[A.4.41]}$$

is said to be *continuous* at $(c_1, c_2, \ldots, c_n)$ if $f(c_1, c_2, \ldots, c_n)$ is finite and for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$|f(x_1, x_2, \ldots, x_n) - f(c_1, c_2, \ldots, c_n)| < \varepsilon$$

whenever

$$(x_1 - c_1)^2 + (x_2 - c_2)^2 + \cdots + (x_n - c_n)^2 < \delta^2.$$

Partial Derivatives

The *partial derivative* of f with respect to x_i is defined by

$$\begin{aligned} \frac{\partial f}{\partial x_i} = \lim_{\Delta \to 0} \Delta^{-1} \cdot \{ & f(x_1, x_2, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_n) \\ & - f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)\}. \end{aligned} \qquad \text{[A.4.42]}$$

Gradient

If we collect the n partial derivatives in [A.4.42] in a vector, we obtain the *gradient* of the function f, denoted ∇:

$$\underset{(n \times 1)}{\nabla} \equiv \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f \partial x_n \end{bmatrix}. \qquad \text{[A.4.43]}$$

For example, suppose f is a linear function:

$$f(x_1, x_2, \ldots, x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n. \qquad \text{[A.4.44]}$$

Define $\mathbf{a}$ and $\mathbf{x}$ to be the following $(n \times 1)$ vectors:

$$\mathbf{a} \equiv \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad \text{[A.4.45]}$$

$$\mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \qquad \text{[A.4.46]}$$

Then [A.4.44] can be written

$$f(\mathbf{x}) = \mathbf{a}'\mathbf{x}.$$

The partial derivative of $f(\cdot)$ with respect to the ith argument is

$$\frac{\partial f}{\partial x_i} = a_i,$$

and the gradient is

$$\nabla = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{a}.$$

Second-Order Derivatives

A *second-order derivative* of [A.4.41] is given by

$$\frac{\partial^2 f(x_1, \ldots, x_n)}{\partial x_i \, \partial x_j} = \frac{\partial}{\partial x_i}\left[\frac{\partial f(x_1, \ldots, x_n)}{\partial x_j}\right].$$

Where second-order derivatives exist and are continuous for all i and j, the order of differentiation is irrelevant:

$$\frac{\partial}{\partial x_i}\left[\frac{\partial f(x_1, \ldots, x_n)}{\partial x_j}\right] = \frac{\partial}{\partial x_j}\left[\frac{\partial f(x_1, \ldots, x_n)}{\partial x_i}\right].$$

Sometimes these second-order derivatives are collected in an $n \times n$ matrix $\mathbf{H}$ called the *Hessian matrix*:

$$\mathbf{H} = \left[\frac{\partial^2 f}{\partial x_i \, \partial x_j}\right].$$

We will also use the notation

$$\frac{\partial^2 f}{\partial \mathbf{x} \, \partial \mathbf{x}'}$$

to represent the matrix $\mathbf{H}$.

Derivatives of Vector-Valued Functions

Suppose we have a set of m functions $f_1(\cdot), f_2(\cdot), \ldots, f_m(\cdot)$, each of which depends on the n variables $(x_1, x_2, \ldots, x_n)$. We can collect the m functions into a single vector-valued function:

$$\underset{(m\times 1)}{\mathbf{f}(\mathbf{x})} = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}.$$

We sometimes write

$$\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

to indicate that the function takes n different real numbers (summarized by the vector $\mathbf{x}$, an element of $\mathbb{R}^n$) and calculates m different new numbers (summarized by the value of $\mathbf{f}$, an element of $\mathbb{R}^m$). Suppose that each of the functions $f_1(\cdot)$, $f_2(\cdot), \ldots, f_m(\cdot)$ has derivatives with respect to each of the arguments $x_1, x_2, \ldots, x_n$. We can summarize these derivatives in an $(m \times n)$ matrix, called the *Jacobian matrix* of $\mathbf{f}$ and indicated by $\partial\mathbf{f}/\partial\mathbf{x}'$:

$$\underset{(m\times n)}{\frac{\partial \mathbf{f}}{\partial \mathbf{x}'}} \equiv \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots & \partial f_2/\partial x_n \\ \vdots & \vdots & \cdots & \vdots \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \cdots & \partial f_m/\partial x_n \end{bmatrix}.$$

For example, suppose that each of the functions $f_i(\mathbf{x})$ is linear:

$$\begin{aligned} f_1(\mathbf{x}) &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ f_2(\mathbf{x}) &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\vdots \\ f_m(\mathbf{x}) &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n. \end{aligned}$$

We could write this system in matrix form as

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x},$$

where

$$\underset{(m\times n)}{\mathbf{A}} \equiv \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

and $\mathbf{x}$ is the $(n \times 1)$ vector defined in [A.4.46]. Then

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}'} = \mathbf{A}.$$

Taylor's Theorem with Multiple Arguments

Let f: $\mathbb{R}^n \rightarrow \mathbb{R}^1$ as in [A.4.41], with continuous second derivatives. A first-order Taylor series expansion of $f(\mathbf{x})$ around $\mathbf{c}$ is given by

$$f(\mathbf{x}) = f(\mathbf{c}) + \left.\frac{\partial f}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}} \cdot (\mathbf{x} - \mathbf{c}) + R_1(\mathbf{c}, \mathbf{x}). \qquad \text{[A.4.47]}$$

Here $\partial f/\partial \mathbf{x}'$ denotes the $(1 \times n)$ vector that is the transpose of the gradient, and the remainder $R_1(\cdot)$ satisfies

$$R_1(\mathbf{c}, \mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left.\frac{\partial^2 f}{\partial x_i\, \partial x_j}\right|_{\mathbf{x}=\boldsymbol{\delta}(i,j)} \cdot (x_i - c_i)(x_j - c_j)$$

for $\boldsymbol{\delta}(i, j)$ an $(n \times 1)$ vector, potentially different for each i and j, with each $\boldsymbol{\delta}(i, j)$ between $\mathbf{c}$ and $\mathbf{x}$, that is, $\boldsymbol{\delta}(i, j) = \lambda(i, j)\mathbf{c} + [1 - \lambda(i, j)]\mathbf{x}$ for some $\lambda(i, j)$ between 0 and 1. Furthermore,

$$\lim_{\mathbf{x}\to\mathbf{c}} \frac{R_1(\mathbf{c}, \mathbf{x})}{[(\mathbf{x} - \mathbf{c})'(\mathbf{x} - \mathbf{c})]^{1/2}} = 0.$$

An implication of [A.4.47] is that if we wish to approximate the consequences for f of simultaneously changing x_1 by Δ_1, x_2 by Δ_2, . . . , and x_n by Δ_n, we could use

$$\begin{aligned} f(x_1 + \Delta_1, x_2 + \Delta_2, \ldots, x_n + \Delta_n) &- f(x_1, x_2, \ldots, x_n) \\ &\cong \frac{\partial f}{\partial x_1}\cdot\Delta_1 + \frac{\partial f}{\partial x_2}\cdot\Delta_2 + \cdots + \frac{\partial f}{\partial x_n}\cdot\Delta_n. \end{aligned} \qquad \text{[A.4.48]}$$

If $f(\cdot)$ has continuous third derivatives, a second-order Taylor series expansion of $f(\mathbf{x})$ around $\mathbf{c}$ is given by

$$\begin{aligned} f(\mathbf{x}) = f(\mathbf{c}) &+ \left.\frac{\partial f}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}} \cdot (\mathbf{x} - \mathbf{c}) \\ &+ \frac{1}{2}(\mathbf{x} - \mathbf{c})' \left.\frac{\partial^2 f}{\partial \mathbf{x}\, \partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}} (\mathbf{x} - \mathbf{c}) + R_2(\mathbf{c}, \mathbf{x}), \end{aligned} \qquad \text{[A.4.49]}$$

where

$$R_2(\mathbf{c}, \mathbf{x}) = \frac{1}{3!}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \left.\frac{\partial^3 f}{\partial x_i\, \partial x_j\, \partial x_k}\right|_{\mathbf{x}=\boldsymbol{\delta}(i,j,k)} \cdot (x_i - c_i)(x_j - c_j)(x_k - c_k)$$

with $\boldsymbol{\delta}(i, j, k)$ between $\mathbf{c}$ and $\mathbf{x}$ and

$$\lim_{\mathbf{x}\to\mathbf{c}} \frac{R_2(\mathbf{c}, \mathbf{x})}{(\mathbf{x} - \mathbf{c})'(\mathbf{x} - \mathbf{c})} = 0.$$

Multiple Integrals

The notation

$$\int_a^b\!\!\int_c^d f(x, y)\, dy\, dx$$

indicates the following operation: first integrate

$$\int_c^d f(x, y)\, dy$$

with respect to y, with x held fixed, and then integrate the resulting function with respect to x. For example,

$$\int_0^1\!\!\int_0^2 x^4 y\, dy\, dx = \int_0^1 x^4[(2^2/2) - (0^2/2)]\, dx = 2[1^5/5 - 0^5/5] = 2/5.$$

Provided that $f(x, y)$ is continuous, the order of integration can be reversed. For example,

$$\int_0^2\int_0^1 x^4 y\,dx\,dy = \int_0^2 (1^5/5)y\,dy = (1/5)\cdot(2^2/2) = 2/5.$$

A.5. Probability and Statistics

Densities and Distributions

A *stochastic* or *random* variable X is said to be *discrete-valued* if it can take on only one of K particular values; call these $x_1, x_2, \ldots, x_K$. Its *probability distribution* is a set of numbers that give the probability of each outcome:

$$P\{X = x_k\} \equiv \text{probability that } X \text{ takes on the value } x_k, \qquad k = 1, \ldots, K.$$

The probabilities sum to unity:

$$\sum_{k=1}^{K} P\{X = x_k\} = 1.$$

Assuming that the possible outcomes are ordered $x_1 < x_2 < \cdots < x_K$, the probability that X takes on a value less than or equal to the value x_j is given by

$$P\{X \leq x_j\} = \sum_{k=1}^{j} P\{X = x_k\}.$$

If X is equal to a constant c with probability 1, then X is *nonstochastic*.

The probability law for a *continuous-valued* random variable X can often be described by the *density* function $f_X(x)$ with

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1. \qquad \text{[A.5.1]}$$

The subscript X in $f_X(x)$ indicates that this is the density of the random variable X; the argument x of $f_X(x)$ indexes the integration in [A.5.1]. The *cumulative distribution function* of X (denoted $F_X(a)$) gives the probability that X takes on a value less than or equal to a:

$$\begin{aligned} F_X(a) &\equiv P\{X \leq a\} \\ &= \int_{-\infty}^{a} f_X(x)\,dx. \end{aligned} \qquad \text{[A.5.2]}$$

Population Moments

The *population mean* μ of a continuous-valued random variable X is given by

$$\mu = \int_{-\infty}^{\infty} x\cdot f_X(x)\,dx,$$

provided this integral exists. (In the formulas that follow, we assume for simplicity of exposition that the density functions are continuous and that the indicated integrals all exist.) The *population variance* is

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x)\, dx.$$

The square root of the variance is called the *population standard deviation*.

In general, the rth *population moment* is given by

$$\int_{-\infty}^{\infty} x^r \cdot f_X(x)\, dx.$$

The population mean could thus be described as the first population moment.

Expectation

The population mean μ is also called the *expectation* of X, denoted $E(X)$ or sometimes simply EX. In general, the expectation of a function $g(X)$ is given by

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx, \qquad [A.5.3]$$

where $f_X(x)$ is the density of X. For example, the rth population moment of X is the expectation of X^r.

Consider the random variable $a + bX$ for constants a and b. Its expectation is

$$\begin{aligned} E(a + bX) &= \int_{-\infty}^{\infty} [a + bx] \cdot f_X(x)\, dx \\ &= a \int_{-\infty}^{\infty} f_X(x)\, dx + b \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx \\ &= a + b \cdot E(X). \end{aligned}$$

The variance of $a + bX$ is

$$\begin{aligned} \text{Var}(a + bX) &= \int_{-\infty}^{\infty} [(a + bx) - (a + b\mu)]^2 \cdot f_X(x)\, dx \\ &= b^2 \cdot \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x)\, dx \qquad [A.5.4] \\ &= b^2 \cdot \text{Var}(X). \end{aligned}$$

Another useful result is

$$\begin{aligned} E(X^2) &= E[(X - \mu + \mu)^2] \\ &= E[(X - \mu)^2 + 2\mu(X - \mu) + \mu^2] \\ &= E[(X - \mu)^2] + 2\mu \cdot [E(X) - \mu] + \mu^2 \\ &= \text{Var}(X) + 0 + [E(X)]^2. \end{aligned}$$

To simplify the appearance of expressions, we adopt the convention that exponentiation and multiplication are carried out before the expectation operator. Thus, we will use $E(X - \mu + \mu)^2$ to indicate the same operation as $E[(X - \mu + \mu)^2]$. The square of $E(X - \mu + \mu)$ is indicated by using additional parentheses, as $[E(X - \mu + \mu)]^2$.

Sample Moments

A *sample moment* is a particular estimate of a population moment based on an observed set of data, say, $\{x_1, x_2, \ldots, x_T\}$. The first sample moment is the sample mean,

$$\bar{x} \equiv (1/T)\cdot(x_1 + x_2 + \cdots + x_T),$$

which is a natural estimate of the population mean μ. The sample variance,

$$s^2 \equiv (1/T)\cdot\sum_{t=1}^{T}(x_t - \bar{x})^2,$$

affords an estimate of the population variance σ^2. More generally, the rth sample moment is given by

$$(1/T)\cdot(x_1^r + x_2^r + \cdots + x_T^r),$$

where x_t^r denotes x_t raised to the rth power.

Bias and Efficiency

Let $\hat{\boldsymbol{\theta}}$ be a sample estimate of a vector of population parameters $\boldsymbol{\theta}$. For example, $\hat{\boldsymbol{\theta}}$ could be the sample mean $\bar{x}$ and $\boldsymbol{\theta}$ the population mean μ. The estimate is said to be *unbiased* if $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$.

Suppose that $\hat{\boldsymbol{\theta}}$ is an unbiased estimate of $\boldsymbol{\theta}$. The estimate $\hat{\boldsymbol{\theta}}$ is said to be *efficient* if it is the case that for any other unbiased estimate $\hat{\boldsymbol{\theta}}^*$, the following matrix is positive semidefinite:

$$\mathbf{P} \equiv E[(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})\cdot(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})'] - E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\cdot(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'].$$

Joint Distributions

For two random variables X and Y with the joint density $f_{X,Y}(x, y)$, we calculate the probability of the joint event that both $X \leq a$ and $Y \leq b$ from

$$P\{X \leq a, Y \leq b\} = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x, y)\, dy\, dx.$$

This can be represented in terms of the joint cumulative distribution function:

$$F_{X,Y}(a, b) = P\{X \leq a, Y \leq b\}.$$

The probability that $X \leq a$ by itself can be calculated from

$$P\{X \leq a, Y \text{ any}\} = \int_{-\infty}^{a}\left[\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\right] dx. \qquad [A.5.5]$$

Comparison of [A.5.5] with [A.5.2] reveals that the *marginal* density $f_X(x)$ is obtained by integrating the joint density $f_{X,Y}(x, y)$ with respect to y:

$$f_X(x) = \left[\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\right]. \qquad [A.5.6]$$

Conditional Distributions

The *conditional density* of Y given X is given by

$$f_{Y|X}(y|x) \equiv \begin{cases} \dfrac{f_{X,Y}(x, y)}{f_X(x)} & \text{if } f_X(x) > 0 \\ 0 & \text{otherwise.} \end{cases} \qquad [A.5.7]$$

Notice that this satisfies the requirement of a density [A.5.1]:

$$\begin{aligned}\int_{-\infty}^{\infty} f_{Y|X}(y|x)\,dy &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy \\ &= \frac{1}{f_X(x)}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\ &= \frac{f_X(x)}{f_X(x)} = 1.\end{aligned}$$

A further obvious implication of the definition in [A.5.7] is that a joint density can be written as the product of the marginal density and the conditional density:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)\cdot f_X(x). \qquad [A.5.8]$$

The *conditional expectation* of Y given that the random variable X takes on the particular value x is

$$E(Y|X = x) = \int_{-\infty}^{\infty} y\cdot f_{Y|X}(y|x)\,dy. \qquad [A.5.9]$$

Law of Iterated Expectations

Note that the conditional expectation is a function of the value of the random variable X. For different realizations of X, the conditional expectation will be a different number. Suppose we view $E(Y|X)$ as a random variable and take its expectation with respect to the distribution of X:

$$E_X[E_{Y|X}(Y|X)] = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} y\cdot f_{Y|X}(y|x)\,dy\right] f_X(x)\,dx.$$

Results [A.5.8] and [A.5.6] can be used to express this expectation as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\cdot f_{Y,X}(y, x)\,dy\,dx = \int_{-\infty}^{\infty} y\cdot f_Y(y)\,dy.$$

Thus,

$$E_X[E_{Y|X}(Y|X)] = E_Y(Y). \qquad [A.5.10]$$

In words, the random variable $E(Y|X)$ has the same expectation as the random variable Y. This is known as the *law of iterated expectations*.

Independence

The variables Y and X are said to be *independent* if

$$f_{X,Y}(x, y) = f_X(x)\cdot f_Y(y). \qquad [A.5.11]$$

Comparing [A.5.11] with [A.5.8], if Y and X are independent, then

$$f_{Y|X}(y|x) = f_Y(y). \qquad [A.5.12]$$

Covariance

Let μ_X denote $E(X)$ and μ_Y denote $E(Y)$. The *population covariance* between X and Y is given by

$$\text{Cov}(X, Y) \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)\cdot f_{X,Y}(x, y)\,dy\,dx. \qquad [A.5.13]$$

Correlation

The *population correlation* between X and Y is given by

$$\text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}.$$

If the covariance (or correlation) between X and Y is zero, then X and Y are said to be *uncorrelated.*

Relation Between Correlation and Independence

Note that if X and Y are independent, then they are uncorrelated:

$$\begin{aligned}\text{Cov}(X, Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)\cdot f_X(x)\cdot f_Y(y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} (x - \mu_X)\left[\int_{-\infty}^{\infty} (y - \mu_Y)\cdot f_Y(y)\, dy\right]\cdot f_X(x)\, dx.\end{aligned}$$

Furthermore,

$$\begin{aligned}\left[\int_{-\infty}^{\infty} (y - \mu_Y)\cdot f_Y(y)\, dy\right] &= \int_{-\infty}^{\infty} y\cdot f_Y(y)\, dy - \mu_Y\cdot\int_{-\infty}^{\infty} f_Y(y)\, dy \\ &= \mu_Y - \mu_Y \\ &= 0.\end{aligned}$$

Thus, if X and Y are independent, then $\text{Cov}(X, Y) = 0$, as claimed.

The converse proposition, however, is not true—the fact that X and Y are uncorrelated is not enough to deduce that they are independent. To construct a counterexample, suppose that Z and Y are independent random variables each with mean zero, and let $X \equiv Z \cdot Y$. Then

$$\begin{aligned}E(X - \mu_X)(Y - \mu_Y) &= E[(ZY)\cdot Y] \\ &= E(Z)\cdot E(Y^2) = 0,\end{aligned}$$

and so X and Y are uncorrelated. They are not, however, independent—the value of ZY depends on Y.

Orthogonality

Consider a sample of size T on two random variables, $\{x_1, x_2, \ldots, x_T\}$ and $\{y_1, y_2, \ldots, y_T\}$. The two variables are said to be *orthogonal* if

$$\sum_{t=1}^{T} x_t y_t = 0.$$

Thus, orthogonality is the sample analog of absence of correlation.

For example, let $x_t = 1$ denote a sequence of constants and let $y_t = w_t - \bar{w}$, where $\bar{w} \equiv (1/T)\Sigma_{t=1}^{T} w_t$ is the sample mean of the variable w. Then x and y are orthogonal:

$$\sum_{t=1}^{T} 1\cdot(w_t - \bar{w}) = \sum_{t=1}^{T} w_t - T\cdot\bar{w} = 0.$$

Population Moments of Sums

Consider the random variable $aX + bY$. Its mean is given by

$$E(aX + bY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax + by)\cdot f_{X,Y}(x, y)\, dy\, dx$$

$$= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot f_{X,Y}(x, y)\, dy\, dx + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\cdot f_{X,Y}(x, y)\, dy\, dx$$

$$= a\int_{-\infty}^{\infty} x\cdot f_X(x)\, dx + b\int_{-\infty}^{\infty} y\cdot f_Y(y)\, dy,$$

and so

$$E(aX + bY) = a\cdot E(X) + b\cdot E(Y). \qquad \text{[A.5.14]}$$

The variance of $(aX + bY)$ is

$$\mathrm{Var}(aX + bY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(ax + by) - (a\mu_X + b\mu_Y)]^2\cdot f_{X,Y}(x, y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(ax - a\mu_X)^2 + 2(ax - a\mu_X)(by - b\mu_Y)$$

$$+ (by - b\mu_Y)^2]\cdot f_{X,Y}(x, y)\, dy\, dx$$

$$= a^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^2\cdot f_{X,Y}(x, y)\, dy\, dx$$

$$+ 2ab\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)\cdot f_{X,Y}(x, y)\, dy\, dx$$

$$+ b^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \mu_Y)^2\cdot f_{X,Y}(x, y)\, dy\, dx.$$

Thus,

$$\mathrm{Var}(aX + bY) = a^2\cdot \mathrm{Var}(X) + 2ab\cdot \mathrm{Cov}(X, Y) + b^2\cdot \mathrm{Var}(Y). \qquad \text{[A.5.15]}$$

When X and Y are uncorrelated,

$$\mathrm{Var}(aX + bY) = a^2\cdot \mathrm{Var}(X) + b^2\cdot \mathrm{Var}(Y).$$

It is straightforward to generalize results [A.5.14] and [A.5.15]. If $\{X_1, X_2, \ldots, X_n\}$ denotes a collection of n random variables, then

$$\begin{aligned} &E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\ &\quad = a_1\cdot E(X_1) + a_2\cdot E(X_2) + \cdots + a_n\cdot E(X_n) \end{aligned} \qquad \text{[A.5.16]}$$

$$\begin{aligned} &\mathrm{Var}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\ &\quad = a_1^2\cdot \mathrm{Var}(X_1) + a_2^2\cdot \mathrm{Var}(X_2) + \cdots + a_n^2\cdot \mathrm{Var}(X_n) \\ &\qquad + 2a_1a_2\cdot \mathrm{Cov}(X_1, X_2) + 2a_1a_3\cdot \mathrm{Cov}(X_1, X_3) + \cdots \\ &\qquad + 2a_1a_n\cdot \mathrm{Cov}(X_1, X_n) + 2a_2a_3\cdot \mathrm{Cov}(X_2, X_3) + 2a_2a_4\cdot \mathrm{Cov}(X_2, X_4) \\ &\qquad + \cdots + 2a_{n-1}a_n\cdot \mathrm{Cov}(X_{n-1}, X_n). \end{aligned} \qquad \text{[A.5.17]}$$

If the X's are uncorrelated, then [A.5.17] simplifies to

$$\begin{aligned}&\mathrm{Var}(a_1X_1 + a_2X_2 + \cdots + a_nX_n)\\&\quad = a_1^2\cdot\mathrm{Var}(X_1) + a_2^2\cdot\mathrm{Var}(X_2) + \cdots + a_n^2\cdot\mathrm{Var}(X_n).\end{aligned} \qquad [A.5.18]$$

Cauchy-Schwarz Inequality

The *Cauchy-Schwarz inequality* states that for any random variables X and Y whose variances and covariance exist, the correlation is no greater than unity in absolute value:

$$-1 \le \mathrm{Corr}(X, Y) \le 1. \qquad [A.5.19]$$

To establish the far right inequality in [A.5.19], consider the random variable

$$Z \equiv \frac{X - \mu_X}{\sqrt{\mathrm{Var}(X)}} - \frac{Y - \mu_Y}{\sqrt{\mathrm{Var}(Y)}}.$$

The square of this variable cannot take on negative values, so

$$E\left[\frac{(X - \mu_X)}{\sqrt{\mathrm{Var}(X)}} - \frac{(Y - \mu_Y)}{\sqrt{\mathrm{Var}(Y)}}\right]^2 \ge 0.$$

Recognizing that $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ denote population moments (as opposed to random variables), equation [A.5.15] can be used to deduce

$$\frac{E(X - \mu_X)^2}{\mathrm{Var}(X)} - 2\left[\frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}\right] + \frac{E(Y - \mu_Y)^2}{\mathrm{Var}(Y)} \ge 0.$$

Thus,

$$1 - 2\cdot\mathrm{Corr}(X, Y) + 1 \ge 0,$$

meaning that

$$\mathrm{Corr}(X, Y) \le 1.$$

To establish the far left inequality in [A.5.19], notice that

$$E\left[\frac{(X - \mu_X)}{\sqrt{\mathrm{Var}(X)}} + \frac{(Y - \mu_Y)}{\sqrt{\mathrm{Var}(Y)}}\right]^2 \ge 0,$$

implying that

$$1 + 2\cdot\mathrm{Corr}(X, Y) + 1 \ge 0,$$

so that

$$\mathrm{Corr}(X, Y) \ge -1.$$

The Normal Distribution

The variable Y_t has a *Gaussian*, or *Normal*, distribution with mean μ and variance σ^2 if

$$f_{Y_t}(y_t) = \frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left[\frac{-(y_t - \mu)^2}{2\sigma^2}\right]. \qquad [A.5.20]$$

We write

$$Y_t \sim N(\mu, \sigma^2)$$

to indicate that the density of Y_t is given by [A.5.20].

Centered odd-ordered population moments for a Gaussian variable are zero:

$$E(Y_t - \mu)^r = 0 \qquad \text{for } r = 1, 3, 5, \ldots .$$

The centered fourth moment is

$$E(Y_t - \mu)^4 = 3\sigma^4.$$

Skew and Kurtosis

The skewness of a variable Y_t with mean μ is represented by

$$\frac{E(Y_t - \mu)^3}{[\mathrm{Var}(Y_t)]^{3/2}}.$$

A variable with a negative skew is more likely to be far below the mean than it is to be far above the mean. The *kurtosis* is

$$\frac{E(Y_t - \mu)^4}{[\mathrm{Var}(Y_t)]^2}.$$

A distribution whose kurtosis exceeds 3 has more mass in the tails than a Gaussian distribution with the same variance.

Other Useful Univariate Distributions

Let $(X_1, X_2, \ldots, X_n)$ be independent and identically distributed (i.i.d.) $N(0, 1)$ variables, and consider the sum of their squares:

$$Y = X_1^2 + X_2^2 + \cdots + X_n^2.$$

Then Y is said to have a *chi-square* distribution with n degrees of freedom, denoted

$$Y \sim \chi^2(n).$$

Let $X \sim N(0, 1)$ and $Y \sim \chi^2(n)$ with X and Y independent. Then

$$Z = \frac{X}{\sqrt{Y/n}}$$

is said to have a *t distribution* with n degrees of freedom, denoted

$$Z \sim t(n).$$

Let $Y_1 \sim \chi^2(n_1)$ and $Y_2 \sim \chi^2(n_2)$ with Y_1 and Y_2 independent. Then

$$Z = \frac{Y_1/n_1}{Y_2/n_2}$$

is said to have an *F distribution* with n_1 numerator degrees of freedom and n_2 denominator degrees of freedom, denoted

$$Z \sim F(n_1, n_2).$$

Note that if $Z \sim t(n)$, then $Z^2 \sim F(1, n)$.

Likelihood Function

Suppose we have observed a sample of size T on some random variable Y_t. Let $f_{Y_1,Y_2,\ldots,Y_T}(y_1, y_2, \ldots, y_T; \boldsymbol{\theta})$ denote the joint density of $Y_1, Y_2, \ldots, Y_T$.

The notation emphasizes that this joint density is presumed to depend on a vector of population parameters $\boldsymbol{\theta}$. If we view this joint density as a function of $\boldsymbol{\theta}$ (given the data on Y), the result is called the *sample likelihood function.*

For example, consider a sample of T i.i.d. variables drawn from a $N(\mu, \sigma^2)$ distribution. For this distribution, $\boldsymbol{\theta} = (\mu, \sigma^2)'$, and from [A.5.11] the joint density is the product of individual terms such as [A.5.20]:

$$f_{Y_1,Y_2,\ldots,Y_T}(y_1, y_2, \ldots, y_T; \mu, \sigma^2) = \prod_{t=1}^{T} f_{Y_t}(y_t; \mu, \sigma^2).$$

The log of the joint density is the sum of the logs of these terms:

$$\begin{aligned} &\log f_{Y_1,Y_2,\ldots,Y_T}(y_1, y_2, \ldots, y_T; \mu, \sigma^2) \\ &= \sum_{t=1}^{T} \log f_{Y_t}(y_t; \mu, \sigma^2) \qquad \text{[A.5.21]} \\ &= (-T/2)\log(2\pi) - (T/2)\log(\sigma^2) - \sum_{t=1}^{T} \frac{(y_t - \mu)^2}{2\sigma^2}. \end{aligned}$$

Thus, for a sample of T Gaussian random variables with mean μ and variance σ^2, the sample log likelihood function, denoted $\mathcal{L}(\mu, \sigma^2; y_1, y_2, \ldots, y_T)$, is given by:

$$\mathcal{L}(\mu, \sigma^2; y_1, y_2, \ldots, y_T) = k - (T/2)\log(\sigma^2) - \sum_{t=1}^{T} \frac{(y_t - \mu)^2}{2\sigma^2}. \qquad \text{[A.5.22]}$$

In calculating the sample log likelihood function, any constant term that does not involve the parameter μ or σ^2 can be ignored for most purposes. In [A.5.22], this constant term is

$$k \equiv -(T/2)\log(2\pi).$$

Maximum Likelihood Estimation

For a given sample of observations $(y_1, y_2, \ldots, y_T)$, the value of $\boldsymbol{\theta}$ that makes the sample likelihood as large as possible is called the *maximum likelihood estimate* (*MLE*) of $\boldsymbol{\theta}$. For example, the maximum likelihood estimate of the population mean μ for an i.i.d. sample of size T from a $N(\mu, \sigma^2)$ distribution is found by setting the derivative of [A.5.22] with respect to μ equal to zero:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{t=1}^{T} \frac{y_t - \mu}{\sigma^2} = 0,$$

or

$$\hat{\mu} = (1/T) \sum_{t=1}^{T} y_t. \qquad \text{[A.5.23]}$$

The *MLE* of σ^2 is characterized by

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \sum_{t=1}^{T} \frac{(y_t - \mu)^2}{2\sigma^4} = 0. \qquad \text{[A.5.24]}$$

Substituting [A.5.23] into [A.5.24] and solving for σ^2 gives

$$\hat{\sigma}^2 = (1/T) \sum_{t=1}^{T} (y_t - \hat{\mu})^2. \qquad \text{[A.5.25]}$$

Thus, the sample mean is the *MLE* of the population mean and the sample variance is the *MLE* of the population variance for an i.i.d. sample of Gaussian variables.

Multivariate Gaussian Distribution

Let

$$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$$

be a collection of n random variables. The vector $\mathbf{Y}$ has a *multivariate Normal*, or *multivariate Gaussian*, distribution if its density takes the form

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-n/2}|\mathbf{\Omega}|^{-1/2}\exp[(-1/2)(\mathbf{y} - \boldsymbol{\mu})'\mathbf{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})]. \qquad \text{[A.5.26]}$$

The mean of $\mathbf{Y}$ is given by the vector $\boldsymbol{\mu}$:

$$E(\mathbf{Y}) = \boldsymbol{\mu};$$

and its variance-covariance matrix is $\mathbf{\Omega}$:

$$E(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})' = \mathbf{\Omega}.$$

Note that $(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'$ is symmetric and positive semidefinite for any $\mathbf{Y}$, meaning that any variance-covariance matrix must be symmetric and positive semidefinite; the form of the likelihood in [A.5.26] assumes that $\mathbf{\Omega}$ is positive definite.

Result [A.4.15] is sometimes used to write the multivariate Gaussian density in an equivalent form:

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-n/2}|\mathbf{\Omega}^{-1}|^{1/2}\exp[(-1/2)(\mathbf{y} - \boldsymbol{\mu})'\mathbf{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})].$$

If $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{\Omega})$, then for any nonstochastic $(r \times n)$ matrix $\mathbf{H}'$ and $(r \times 1)$ vector $\mathbf{b}$,

$$\mathbf{H}'\mathbf{Y} + \mathbf{b} \sim N\big((\mathbf{H}'\boldsymbol{\mu} + \mathbf{b}), \mathbf{H}'\mathbf{\Omega}\mathbf{H}\big).$$

Correlation and Independence for Multivariate Gaussian Variates

If $\mathbf{Y}$ has a multivariate Gaussian distribution, then absence of correlation implies independence. To see this, note that if the elements of $\mathbf{Y}$ are uncorrelated, then $E[(Y_i - \mu)(Y_j - \mu)] = 0$ for $i \neq j$ and the off-diagonal elements of $\mathbf{\Omega}$ are zero:

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

For such a diagonal matrix $\mathbf{\Omega}$,

$$|\mathbf{\Omega}| = \sigma_1^2\sigma_2^2 \cdots \sigma_n^2 \qquad \text{[A.5.27]}$$

$$\mathbf{\Omega}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n^2 \end{bmatrix}. \qquad \text{[A.5.28]}$$

Substituting [A.5.27] and [A.5.28] into [A.5.26] produces

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= (2\pi)^{-n/2}[\sigma_1^2\sigma_2^2 \cdots \sigma_n^2]^{-1/2} \\ &\quad \times \exp[(-1/2)\{(y_1 - \mu_1)^2/\sigma_1^2 + (y_2 - \mu_2)^2/\sigma_2^2 + \cdots \\ &\quad + (y_n - \mu_n)^2/\sigma_n^2\}] \\ &= \prod_{i=1}^{n} (2\pi)^{-1/2}[\sigma_i^2]^{-1/2} \exp[(-1/2)\{(y_i - \mu_i)^2/\sigma_i^2\}], \end{aligned}$$

which is the product of n univariate Gaussian densities. Since the joint density is the product of the individual densities, the random variables $(Y_1, Y_2, \ldots, Y_n)$ are independent.

Probability Limit

Let $\{X_1, X_2, \ldots, X_T\}$ denote a sequence of random variables. Often we are interested in what happens to this sequence as T becomes large. For example, X_T might denote the sample mean of T observations:

$$X_T = (1/T)\cdot(Y_1 + Y_2 + \cdots + Y_T), \qquad [A.5.29]$$

in which case we might want to know the properties of the sample mean as the size of the sample T grows large.

The sequence $\{X_1, X_2, \ldots, X_T\}$ is said to *converge in probability* to c if for every $\varepsilon > 0$ and $\delta > 0$ there exists a value N such that, for all $T \geq N$,

$$P\{|X_T - c| > \delta\} < \varepsilon. \qquad [A.5.30]$$

When [A.5.30] is satisfied, the number c is called the *probability limit*, or *plim*, of the sequence $\{X_1, X_2, \ldots, X_T\}$. This is sometimes indicated as

$$X_T \xrightarrow{p} c.$$

Law of Large Numbers

Under fairly general conditions detailed in Chapter 7, the sample mean [A.5.29] converges in probability to the population mean:

$$(1/T)(Y_1 + Y_2 + \cdots + Y_T) \xrightarrow{p} E(Y_t). \qquad [A.5.31]$$

When [A.5.31] holds, we say that the sample mean gives a *consistent* estimate of the population mean.

Convergence in Mean Square

A stronger condition than convergence in probability is *mean square convergence*. The sequence $\{X_1, X_2, \ldots, X_T\}$ is said to converge in mean square if for every $\varepsilon > 0$ there exists a value N such that, for all $T \geq N$,

$$E(X_T - c)^2 < \varepsilon. \qquad [A.5.32]$$

We indicate that the sequence converges to c in mean square as follows:

$$X_T \xrightarrow{m.s.} c.$$

Convergence in mean square implies convergence in probability, but convergence in probability does not imply convergence in mean square.

Appendix A References

Chiang, Alpha C. 1974. *Fundamental Methods of Mathematical Economics*, 2d ed. New York: McGraw-Hill.

Hoel, Paul G., Sidney C. Port, and Charles J. Stone. 1971. *Introduction to Probability Theory*. Boston: Houghton Mifflin.

Johnston, J. 1984. *Econometric Methods*, 3d ed. New York: McGraw-Hill.

Lindgren, Bernard W. 1976. *Statistical Theory*, 3d ed. New York: Macmillan.

Magnus, Jan R., and Heinz Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.

Marsden, Jerrold E. 1974. *Elementary Classical Analysis*. San Francisco: Freeman.

O'Nan, Michael. 1976. *Linear Algebra*, 2d ed. New York: Harcourt Brace Jovanovich.

Strang, Gilbert. 1976. *Linear Algebra and Its Applications*. New York: Academic Press.

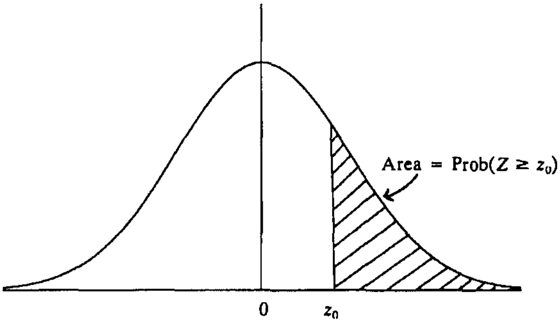
Theil, Henri. 1971. *Principles of Econometrics*. New York: Wiley.

Thomas, George B., Jr. 1972. *Calculus and Analytic Geometry*, alternate ed. Reading, Mass.: Addison-Wesley Publishing Company, Inc.

B

Statistical Tables

TABLE B.1
Standard Normal Distribution



$\rightarrow$ $\downarrow z_0$	Second decimal place of z_0 *.00*	*.01*	*.02*	*.03*	*.04*	*.05*	*.06*	*.07*	*.08*	*.09*
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681

(*continued on next page*)

TABLE B.1 (continued)

$\downarrow z_0$ →	Second decimal place of z_0 .00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

Table entries give the probability that a $N(0, 1)$ variable takes on a value greater than or equal to z_0. For example, if $Z \sim N(0, 1)$, the probability that $Z > 1.96 = 0.0250$. By symmetry, the table entries could also be interpreted as the probability that a $N(0, 1)$ variable takes a value less than or equal to $-z_0$.

Source: Thomas H. Wonnacott and Ronald J. Wonnacott, *Introductory Statistics*, 2d ed., p. 480. Copyright © 1972 by John Wiley & Sons, Inc., New York. Reprinted by permission of John Wiley & Sons, Inc.

TABLE B.2
The χ^2 Distribution

Degrees of freedom	*Probability that $\chi^2(m)$ is greater than entry*						
(m)	*0.995*	*0.990*	*0.975*	*0.950*	*0.900*	*0.750*	*0.500*
1	4×10^{-5}	2×10^{-4}	0.001	0.004	0.016	0.102	0.455
2	0.010	0.020	0.051	0.103	0.211	0.575	1.39
3	0.072	0.115	0.216	0.352	0.584	1.21	2.37
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3

(continued on next page)

TABLE B.2 (continued)

Degrees of freedom (m)	Probability that $\chi^2(m)$ is greater than entry						
	0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	22.7	27.2	30.1	32.9	36.2	38.6	43.8
20	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	28.2	33.2	36.4	39.4	43.0	45.6	51.2
25	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	77.6	85.5	90.5	95.0	100	104	112
80	88.1	96.6	102	107	112	116	125
90	98.6	108	113	118	124	128	137
100	109	118	124	130	136	140	149

The probability shown at the head of the column is the area in the right-hand tail. For example, there is a 10% probability that a χ^2 variable with 2 degrees of freedom would be greater than 4.61.

Source: Adapted from Henri Theil, *Principles of Econometrics*, pp. 718–19. Copyright © 1971 by John Wiley & Sons, Inc., New York. Also Thomas H. Wonnacott and Ronald J. Wonnacott, *Introductory Statistics*, 2d ed., p. 482. Copyright © 1972 by John Wiley & Sons, Inc., New York. Reprinted by permission of John Wiley & Sons, Inc.

TABLE B.3
The *t* Distribution

Degrees of freedom (m)	Probability that $t(m)$ is greater than entry						
	0.25	0.10	0.05	0.025	0.010	0.005	0.001
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	.816	1.886	2.920	4.303	6.965	9.925	22.326
3	.765	1.638	2.353	3.182	4.541	5.841	10.213
4	.741	1.533	2.132	2.776	3.747	4.604	7.173
5	.727	1.476	2.015	2.571	3.365	4.032	5.893
6	.718	1.440	1.943	2.447	3.143	3.707	5.208
7	.711	1.415	1.895	2.365	2.998	3.499	4.785
8	.706	1.397	1.860	2.306	2.896	3.355	4.501
9	.703	1.383	1.833	2.262	2.821	3.250	4.297
10	.700	1.372	1.812	2.228	2.764	3.169	4.144
11	.697	1.363	1.796	2.201	2.718	3.106	4.025
12	.695	1.356	1.782	2.179	2.681	3.055	3.930
13	.694	1.350	1.771	2.160	2.650	3.012	3.852
14	.692	1.345	1.761	2.145	2.624	2.977	3.787
15	.691	1.341	1.753	2.131	2.602	2.947	3.733
16	.690	1.337	1.746	2.120	2.583	2.921	3.686
17	.689	1.333	1.740	2.110	2.567	2.898	3.646
18	.688	1.330	1.734	2.101	2.552	2.878	3.610
19	.688	1.328	1.729	2.093	2.539	2.861	3.579
20	.687	1.325	1.725	2.086	2.528	2.845	3.552
21	.686	1.323	1.721	2.080	2.518	2.831	3.527
22	.686	1.321	1.717	2.074	2.508	2.819	3.505
23	.685	1.319	1.714	2.069	2.500	2.807	3.485
24	.685	1.318	1.711	2.064	2.492	2.797	3.467
25	.684	1.316	1.708	2.060	2.485	2.787	3.450
26	.684	1.315	1.706	2.056	2.479	2.779	3.435
27	.684	1.314	1.703	2.052	2.473	2.771	3.421
28	.683	1.313	1.701	2.048	2.467	2.763	3.408
29	.683	1.311	1.699	2.045	2.462	2.756	3.396
30	.683	1.310	1.697	2.042	2.457	2.750	3.385
40	.681	1.303	1.684	2.021	2.423	2.704	3.307
60	.679	1.296	1.671	2.000	2.390	2.660	3.232
120	.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	.674	1.282	1.645	1.960	2.326	2.576	3.090

The probability shown at the head of the column is the area in the right-hand tail. For example, there is a 10% probability that a t variable with 20 degrees of freedom would be greater than 1.325. By symmetry, there is also a 10% probability that a t variable with 20 degrees of freedom would be less than -1.325.

Source: Thomas H. Wonnacott and Ronald J. Wonnacott, *Introductory Statistics*, 2d ed., p. 481.

TABLE B.4
The F Distribution

Denominator degrees of freedom (m_2)	*Numerator degrees of freedom* (m_1)									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1	161	200	216	225	230	234	237	239	241	242
	4052	**4999**	**5403**	**5625**	**5764**	**5859**	**5928**	**5981**	**6022**	**6056**
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39
	98.49	**99.00**	**99.17**	**99.25**	**99.30**	**99.33**	**99.34**	**99.36**	**99.38**	**99.40**
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.54**
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.27**	**10.15**	**10.05**
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.35**	**5.26**
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.21**	**5.06**	**4.95**	**4.85**
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.63**	**4.54**
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.39**	**4.30**
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.85**	**3.71**	**3.60**	**3.51**
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.52**	**3.43**

(continued on page 758)

11	12	14	16	20	24	30	40	50	75	100	200	500	∞
243	244	245	246	248	249	250	251	252	253	253	254	254	254
6082	**6106**	**6142**	**6169**	**6203**	**6234**	**6258**	**6286**	**6302**	**6323**	**6334**	**6352**	**6361**	**6366**
19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50
99.41	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.49**	**99.49**	**99.49**	**99.50**	**99.50**
8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53
27.13	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.35**	**26.27**	**26.23**	**26.18**	**26.14**	**26.12**
5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63
14.45	**14.37**	**14.24**	**14.15**	**14.02**	**13.93**	**13.83**	**13.74**	**13.69**	**13.61**	**13.57**	**13.52**	**13.48**	**13.46**
4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36
9.96	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.07**	**9.04**	**9.02**
4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
7.79	**7.72**	**7.60**	**7.52**	**7.39**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.94**	**6.90**	**6.88**
3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
6.54	**6.47**	**6.35**	**6.27**	**6.15**	**6.07**	**5.98**	**5.90**	**5.85**	**5.78**	**5.75**	**5.70**	**5.67**	**5.65**
3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93
5.74	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.11**	**5.06**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**
3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71
5.18	**5.11**	**5.00**	**4.92**	**4.80**	**4.73**	**4.64**	**4.56**	**4.51**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**
2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54
4.78	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**
2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40
4.46	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.80**	**3.74**	**3.70**	**3.66**	**3.62**	**3.60**
2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30
4.22	**4.16**	**4.05**	**3.93**	**3.86**	**3.78**	**3.70**	**3.61**	**3.56**	**3.49**	**3.46**	**3.41**	**3.38**	**3.36**
2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21
4.02	**3.96**	**3.85**	**3.78**	**3.67**	**3.59**	**3.51**	**3.42**	**3.37**	**3.30**	**3.27**	**3.21**	**3.18**	**3.16**
2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
3.86	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.34**	**3.26**	**3.21**	**3.14**	**3.11**	**3.06**	**3.02**	**3.00**
2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07
3.73	**3.67**	**3.56**	**3.48**	**3.36**	**3.29**	**3.20**	**3.12**	**3.07**	**3.00**	**2.97**	**2.92**	**2.89**	**2.87**
2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01
3.61	**3.55**	**3.45**	**3.37**	**3.25**	**3.18**	**3.10**	**3.01**	**2.96**	**2.89**	**2.86**	**2.80**	**2.77**	**2.75**
2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96
3.52	**3.45**	**3.35**	**3.27**	**3.16**	**3.08**	**3.00**	**2.92**	**2.86**	**2.79**	**2.76**	**2.70**	**2.67**	**2.65**
2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92
3.44	**3.37**	**3.27**	**3.19**	**3.07**	**3.00**	**2.91**	**2.83**	**2.78**	**2.71**	**2.68**	**2.62**	**2.59**	**2.57**
2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88
3.36	**3.30**	**3.19**	**3.12**	**3.00**	**2.92**	**2.84**	**2.76**	**2.70**	**2.63**	**2.60**	**2.54**	**2.51**	**2.49**

TABLE B.4 (continued)

Denominator degrees of freedom (m_2)	Numerator degrees of freedom (m_1)									
	1	2	3	4	5	6	7	8	9	10
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.71**	**3.56**	**3.45**	**3.37**
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.65**	**3.51**	**3.40**	**3.31**
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.30**	**3.21**
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.25**	**3.17**
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24
	7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.46**	**3.32**	**3.21**	**3.13**
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.17**	**3.09**
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.14**	**3.06**
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.11**	**3.03**
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**3.01**	**2.94**
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.97**	**2.89**
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10
	7.39	**5.25**	**4.38**	**3.89**	**3.58**	**3.35**	**3.18**	**3.04**	**2.94**	**2.86**
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.86**	**2.77**
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.84**	**2.75**
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.82**	**2.73**
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.80**	**2.71**
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.78**	**2.70**
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**

(continued on page 760)

11	*12*	*14*	*16*	*20*	*24*	*30*	*40*	*50*	*75*	*100*	*200*	*500*	∞
2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84
3.30	**3.23**	**3.13**	**3.05**	**2.94**	**2.86**	**2.77**	**2.69**	**2.63**	**2.56**	**2.53**	**2.47**	**2.44**	**2.42**
2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81
3.24	**3.17**	**3.07**	**2.99**	**2.88**	**2.80**	**2.72**	**2.63**	**2.58**	**2.51**	**2.47**	**2.42**	**2.38**	**2.36**
2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78
3.18	**3.12**	**3.02**	**2.94**	**2.83**	**2.75**	**2.67**	**2.58**	**2.53**	**2.46**	**2.42**	**2.37**	**2.33**	**2.31**
2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
3.14	**3.07**	**2.97**	**2.89**	**2.78**	**2.70**	**2.62**	**2.53**	**2.48**	**2.41**	**2.37**	**2.32**	**2.28**	**2.26**
2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73
3.09	**3.03**	**2.93**	**2.85**	**2.74**	**2.66**	**2.58**	**2.49**	**2.44**	**2.36**	**2.33**	**2.27**	**2.23**	**2.21**
2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71
3.05	**2.99**	**2.89**	**2.81**	**2.70**	**2.62**	**2.54**	**2.45**	**2.40**	**2.32**	**2.29**	**2.23**	**2.19**	**2.17**
2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69
3.02	**2.96**	**2.86**	**2.77**	**2.66**	**2.58**	**2.50**	**2.41**	**2.36**	**2.28**	**2.25**	**2.19**	**2.15**	**2.13**
2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
2.98	**2.93**	**2.83**	**2.74**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.25**	**2.21**	**2.16**	**2.12**	**2.10**
2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
2.95	**2.90**	**2.80**	**2.71**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.22**	**2.18**	**2.13**	**2.09**	**2.06**
2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
2.92	**2.87**	**2.77**	**2.68**	**2.57**	**2.49**	**2.41**	**2.32**	**2.27**	**2.19**	**2.15**	**2.10**	**2.06**	**2.03**
2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
2.90	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.38**	**2.29**	**2.24**	**2.16**	**2.13**	**2.07**	**2.03**	**2.01**
2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
2.86	**2.80**	**2.70**	**2.62**	**2.51**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**	**2.08**	**2.02**	**1.98**	**1.96**
2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
2.82	**2.76**	**2.66**	**2.58**	**2.47**	**2.38**	**2.30**	**2.21**	**2.15**	**2.08**	**2.04**	**1.98**	**1.94**	**1.91**
2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
2.78	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.17**	**2.12**	**2.04**	**2.00**	**1.94**	**1.90**	**1.87**
2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
2.75	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.22**	**2.14**	**2.00**	**2.00**	**1.97**	**1.90**	**1.86**	**1.84**
2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
2.73	**2.66**	**2.56**	**2.49**	**2.37**	**2.29**	**2.20**	**2.11**	**2.05**	**1.97**	**1.94**	**1.88**	**1.84**	**1.81**
2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
2.70	**2.64**	**2.54**	**2.46**	**2.35**	**2.26**	**2.17**	**2.08**	**2.02**	**1.94**	**1.91**	**1.85**	**1.80**	**1.78**
2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
2.68	**2.62**	**2.52**	**2.44**	**2.32**	**2.24**	**2.15**	**2.06**	**2.00**	**1.92**	**1.88**	**1.82**	**1.78**	**1.75**
2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
2.66	**2.60**	**2.50**	**2.42**	**2.30**	**2.22**	**2.13**	**2.04**	**1.98**	**1.90**	**1.86**	**1.80**	**1.76**	**1.72**
1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
2.64	**2.58**	**2.48**	**2.40**	**2.28**	**2.20**	**2.11**	**2.02**	**1.96**	**1.88**	**1.84**	**1.78**	**1.73**	**1.70**
1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
2.62	**2.56**	**2.46**	**2.39**	**2.26**	**2.18**	**2.10**	**2.00**	**1.94**	**1.86**	**1.82**	**1.76**	**1.71**	**1.68**
1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41
2.59	**2.53**	**2.43**	**2.35**	**2.23**	**2.15**	**2.06**	**1.96**	**1.90**	**1.82**	**1.78**	**1.71**	**1.66**	**1.64**

TABLE B.4 (continued)

Denominator degrees of freedom (m_2)	*Numerator degrees of freedom* (m_1)									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.70**	**2.61**
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.56**	**2.47**
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.53**	**2.44**
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.50**	**2.41**
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.46**	**2.37**
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84
	6.66	**4.62**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**

The table describes the distribution of an F variable with m_1 numerator and m_2 denominator degrees of freedom. Entries in the standard typeface give the 5% critical value, and boldface entries give the 1% critical value for the distribution. For example, there is a 5% probability that an F variable with 2 numerator and 50 denominator degrees of freedom would exceed 3.18; there is only a 1% probability that it would exceed 5.06.

Source: George W. Snedecor and William G. Cochran, *Statistical Methods*, 8th ed. Copyright 1989 by Iowa State University Press. Reprinted by permission of Iowa State University Press.

11	*12*	*14*	*16*	*20*	*24*	*30*	*40*	*50*	*75*	*100*	*200*	*500*	∞
1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39
2.56	**2.50**	**2.40**	**2.32**	**2.20**	**2.12**	**2.03**	**1.93**	**1.87**	**1.79**	**1.74**	**1.68**	**1.63**	**1.60**
1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37
2.54	**2.47**	**2.37**	**2.30**	**2.18**	**2.09**	**2.00**	**1.90**	**1.84**	**1.76**	**1.71**	**1.64**	**1.60**	**1.56**
1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35
2.51	**2.45**	**2.35**	**2.28**	**2.15**	**2.07**	**1.98**	**1.88**	**1.82**	**1.74**	**1.69**	**1.62**	**1.56**	**1.53**
1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32
2.48	**2.41**	**2.32**	**2.24**	**2.11**	**2.03**	**1.94**	**1.84**	**1.78**	**1.70**	**1.65**	**1.57**	**1.52**	**1.49**
1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28
2.43	**2.36**	**2.26**	**2.19**	**2.06**	**1.98**	**1.89**	**1.79**	**1.73**	**1.64**	**1.59**	**1.51**	**1.46**	**1.43**
1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
2.40	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.75**	**1.68**	**1.59**	**1.54**	**1.46**	**1.40**	**1.37**
1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
2.37	**2.30**	**2.20**	**2.12**	**2.00**	**1.91**	**1.83**	**1.72**	**1.66**	**1.56**	**1.51**	**1.43**	**1.37**	**1.33**
1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
2.34	**2.28**	**2.17**	**2.09**	**1.97**	**1.88**	**1.79**	**1.69**	**1.62**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**
1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
2.29	**2.23**	**2.12**	**2.04**	**1.92**	**1.84**	**1.74**	**1.64**	**1.57**	**1.47**	**1.42**	**1.32**	**1.24**	**1.19**
1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
2.26	**2.20**	**2.09**	**2.01**	**1.89**	**1.81**	**1.71**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.11**
1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00
2.24	**2.18**	**2.07**	**1.99**	**1.87**	**1.79**	**1.69**	**1.59**	**1.52**	**1.41**	**1.36**	**1.25**	**1.15**	**1.00**

TABLE B.5
Critical Values for the Phillips-Perron Z_ρ Test and for the Dickey-Fuller Test Based on Estimated *OLS* Autoregressive Coefficient

Sample size T	*Probability that $T(\hat{\rho} - 1)$ is less than entry*							
	0.01	*0.025*	*0.05*	*0.10*	*0.90*	*0.95*	*0.975*	*0.99*
				Case 1				
25	−11.9	−9.3	−7.3	−5.3	1.01	1.40	1.79	2.28
50	−12.9	−9.9	−7.7	−5.5	0.97	1.35	1.70	2.16
100	−13.3	−10.2	−7.9	−5.6	0.95	1.31	1.65	2.09
250	−13.6	−10.3	−8.0	−5.7	0.93	1.28	1.62	2.04
500	−13.7	−10.4	−8.0	−5.7	0.93	1.28	1.61	2.04
∞	−13.8	−10.5	−8.1	−5.7	0.93	1.28	1.60	2.03
				Case 2				
25	−17.2	−14.6	−12.5	−10.2	−0.76	0.01	0.65	1.40
50	−18.9	−15.7	−13.3	−10.7	−0.81	−0.07	0.53	1.22
100	−19.8	−16.3	−13.7	−11.0	−0.83	−0.10	0.47	1.14
250	−20.3	−16.6	−14.0	−11.2	−0.84	−0.12	0.43	1.09
500	−20.5	−16.8	−14.0	−11.2	−0.84	−0.13	0.42	1.06
∞	−20.7	−16.9	−14.1	−11.3	−0.85	−0.13	0.41	1.04
				Case 4				
25	−22.5	−19.9	−17.9	−15.6	−3.66	−2.51	−1.53	−0.43
50	−25.7	−22.4	−19.8	−16.8	−3.71	−2.60	−1.66	−0.65
100	−27.4	−23.6	−20.7	−17.5	−3.74	−2.62	−1.73	−0.75
250	−28.4	−24.4	−21.3	−18.0	−3.75	−2.64	−1.78	−0.82
500	−28.9	−24.8	−21.5	−18.1	−3.76	−2.65	−1.78	−0.84
∞	−29.5	−25.1	−21.8	−18.3	−3.77	−2.66	−1.79	−0.87

The probability shown at the head of the column is the area in the left-hand tail.

Source: Wayne A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 1976, p. 371.

TABLE B.6
Critical Values for the Phillips-Perron Z_t Test and for the Dickey-Fuller Test Based on Estimated *OLS t* Statistic

Sample size T	Probability that $(\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}}$ is less than entry							
	0.01	*0.025*	*0.05*	*0.10*	*0.90*	*0.95*	*0.975*	*0.99*
				Case 1				
25	−2.66	−2.26	−1.95	−1.60	0.92	1.33	1.70	2.16
50	−2.62	−2.25	−1.95	−1.61	0.91	1.31	1.66	2.08
100	−2.60	−2.24	−1.95	−1.61	0.90	1.29	1.64	2.03
250	−2.58	−2.23	−1.95	−1.62	0.89	1.29	1.63	2.01
500	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
∞	−2.58	−2.23	−1.95	−1.62	0.89	1.28	1.62	2.00
				Case 2				
25	−3.75	−3.33	−3.00	−2.63	−0.37	0.00	0.34	0.72
50	−3.58	−3.22	−2.93	−2.60	−0.40	−0.03	0.29	0.66
100	−3.51	−3.17	−2.89	−2.58	−0.42	−0.05	0.26	0.63
250	−3.46	−3.14	−2.88	−2.57	−0.42	−0.06	0.24	0.62
500	−3.44	−3.13	−2.87	−2.57	−0.43	−0.07	0.24	0.61
∞	−3.43	−3.12	−2.86	−2.57	−0.44	−0.07	0.23	0.60
				Case 4				
25	−4.38	−3.95	−3.60	−3.24	−1.14	−0.80	−0.50	−0.15
50	−4.15	−3.80	−3.50	−3.18	−1.19	−0.87	−0.58	−0.24
100	−4.04	−3.73	−3.45	−3.15	−1.22	−0.90	−0.62	−0.28
250	−3.99	−3.69	−3.43	−3.13	−1.23	−0.92	−0.64	−0.31
500	−3.98	−3.68	−3.42	−3.13	−1.24	−0.93	−0.65	−0.32
∞	−3.96	−3.66	−3.41	−3.12	−1.25	−0.94	−0.66	−0.33

The probability shown at the head of the column is the area in the left-hand tail.

Source: Wayne A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 1976, p. 373.

TABLE B.7
Critical Values for the Dickey-Fuller Test Based on the *OLS F* Statistic

Sample size T	*Probability that F test is greater than entry*							
	0.99	*0.975*	*0.95*	*0.90*	*0.10*	*0.05*	*0.025*	*0.01*
				Case 2				
	(F test of $\alpha = 0$, $\rho = 1$ in regression $y_t = \alpha + \rho y_{t-1} + u_t$)							
25	0.29	0.38	0.49	0.65	4.12	5.18	6.30	7.88
50	0.29	0.39	0.50	0.66	3.94	4.86	5.80	7.06
100	0.29	0.39	0.50	0.67	3.86	4.71	5.57	6.70
250	0.30	0.39	0.51	0.67	3.81	4.63	5.45	6.52
500	0.30	0.39	0.51	0.67	3.79	4.61	5.41	6.47
∞	0.30	0.40	0.51	0.67	3.78	4.59	5.38	6.43
				Case 4				
	(F test of $\delta = 0$, $\rho = 1$ in regression $y_t = \alpha + \delta t + \rho y_{t-1} + u_t$)							
25	0.74	0.90	1.08	1.33	5.91	7.24	8.65	10.61
50	0.76	0.93	1.11	1.37	5.61	6.73	7.81	9.31
100	0.76	0.94	1.12	1.38	5.47	6.49	7.44	8.73
250	0.76	0.94	1.13	1.39	5.39	6.34	7.25	8.43
500	0.76	0.94	1.13	1.39	5.36	6.30	7.20	8.34
∞	0.77	0.94	1.13	1.39	5.34	6.25	7.16	8.27

The probability shown at the head of the column is the area in the right-hand tail.

Source: David A. Dickey and Wayne A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica* 49 (1981), p. 1063.

TABLE B.8
Critical Values for the Phillips Z_ρ Statistic When Applied to Residuals from Spurious Cointegrating Regression

Number of right-hand variables in regression, excluding trend or constant $(n-1)$	*Sample size* (T)	*Probability that* $(T-1)(\hat{\rho}-1)$ *is less than entry*						
		0.010	*0.025*	*0.050*	*0.075*	*0.100*	*0.125*	*0.150*
					Case 1			
1	500	−22.8	−18.9	−15.6	−13.8	−12.5	−11.6	−10.7
2	500	−29.3	−25.2	−21.5	−19.6	−18.2	−17.0	−16.0
3	500	−36.2	−31.5	−27.9	−25.5	−23.9	−22.6	−21.5
4	500	−42.9	−37.5	−33.5	−30.9	−28.9	−27.4	−26.2
5	500	−48.5	−42.5	−38.1	−35.5	−33.8	−32.3	−30.9
					Case 2			
1	500	−28.3	−23.8	−20.5	−18.5	−17.0	−15.9	−14.9
2	500	−34.2	−29.7	−26.1	−23.9	−22.2	−21.0	−19.9
3	500	−41.1	−35.7	−32.1	−29.5	−27.6	−26.2	−25.1
4	500	−47.5	−41.6	−37.2	−34.7	−32.7	−31.2	−29.9
5	500	−52.2	−46.5	−41.9	−39.1	−37.0	−35.5	−34.2
					Case 3			
1	500	−28.9	−24.8	−21.5	—	−18.1	—	—
2	500	−35.4	−30.8	−27.1	−24.8	−23.2	−21.8	−20.8
3	500	−40.3	−36.1	−32.2	−29.7	−27.8	−26.5	−25.3
4	500	−47.4	−42.6	−37.7	−35.0	−33.2	−31.7	−30.3
5	500	−53.6	−47.1	−42.5	−39.7	−37.7	−36.0	−34.6

The probability shown at the head of the column is the area in the left-hand tail.

Source: P. C. B. Phillips and S. Ouliaris, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica* 58 (1990), pp. 189–90. Also Wayne A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 1976, p. 371.

TABLE B.9
Critical Values for the Phillips Z_t Statistic or the Dickey-Fuller t Statistic When Applied to Residuals from Spurious Cointegrating Regression

Number of right-hand variables in regression, excluding trend or constant $(n-1)$	*Sample size* (T)	*Probability that* $(\hat{\rho}-1)/\hat{\sigma}_{\hat{\rho}}$ *is less than entry*						
		0.010	*0.025*	*0.050*	*0.075*	*0.100*	*0.125*	*0.150*
					Case 1			
1	500	−3.39	−3.05	−2.76	−2.58	−2.45	−2.35	−2.26
2	500	−3.84	−3.55	−3.27	−3.11	−2.99	−2.88	−2.79
3	500	−4.30	−3.99	−3.74	−3.57	−3.44	−3.35	−3.26
4	500	−4.67	−4.38	−4.13	−3.95	−3.81	−3.71	−3.61
5	500	−4.99	−4.67	−4.40	−4.25	−4.14	−4.04	−3.94
					Case 2			
1	500	−3.96	−3.64	−3.37	−3.20	−3.07	−2.96	−2.86
2	500	−4.31	−4.02	−3.77	−3.58	−3.45	−3.35	−3.26
3	500	−4.73	−4.37	−4.11	−3.96	−3.83	−3.73	−3.65
4	500	−5.07	−4.71	−4.45	−4.29	−4.16	−4.05	−3.96
5	500	−5.28	−4.98	−4.71	−4.56	−4.43	−4.33	−4.24
					Case 3			
1	500	−3.98	−3.68	−3.42	—	−3.13	—	—
2	500	−4.36	−4.07	−3.80	−3.65	−3.52	−3.42	−3.33
3	500	−4.65	−4.39	−4.16	−3.98	−3.84	−3.74	−3.66
4	500	−5.04	−4.77	−4.49	−4.32	−4.20	−4.08	−4.00
5	500	−5.36	−5.02	−4.74	−4.58	−4.46	−4.36	−4.28

The probability shown at the head of the column is the area in the left-hand tail.

Source: P. C. B. Phillips and S. Ouliaris, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica* 58 (1990), p. 190. Also Wayne A. Fuller, *Introduction to Statistical Time Series*, Wiley, New York, 1976, p. 373.

TABLE B.10
Critical Values for Johansen's Likelihood Ratio Test of the Null Hypothesis of h Cointegrating Relations Against the Alternative of No Restrictions

Number of random walks ($g = n - h$)	*Sample size*	*Probability that $2(\mathcal{L}_A - \mathcal{L}_0)$ is greater than entry*					
(g)	*(T)*	*0.500*	*0.200*	*0.100*	*0.050*	*0.025*	*0.001*
			Case 1				
1	400	0.58	1.82	2.86	3.84	4.93	6.51
2	400	5.42	8.45	10.47	12.53	14.43	16.31
3	400	14.30	18.83	21.63	24.31	26.64	29.75
4	400	27.10	33.16	36.58	39.89	42.30	45.58
5	400	43.79	51.13	55.44	59.46	62.91	66.52
			Case 2				
1	400	2.415	4.905	6.691	8.083	9.658	11.576
2	400	9.335	13.038	15.583	17.844	19.611	21.962
3	400	20.188	25.445	28.436	31.256	34.062	37.291
4	400	34.873	41.623	45.248	48.419	51.801	55.551
5	400	53.373	61.566	65.956	69.977	73.031	77.911
			Case 3				
1	400	0.447	1.699	2.816	3.962	5.332	6.936
2	400	7.638	11.164	13.338	15.197	17.299	19.310
3	400	18.759	23.868	26.791	29.509	32.313	35.397
4	400	33.672	40.250	43.964	47.181	50.424	53.792
5	400	52.588	60.215	65.063	68.905	72.140	76.955

The probability shown at the head of the column is the area in the right-hand tail. The number of random walks under the null hypothesis (g) is given by the number of variables described by the vector autoregression (n) minus the number of cointegrating relations under the null hypothesis (h). In each case the alternative is that $g = 0$.

Source: Michael Osterwald-Lenum, "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics* 54 (1992), p. 462; and Søren Johansen and Katarina Juselius, "Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics* 52 (1990), p. 208.

TABLE B.11
Critical Values for Johansen's Likelihood Ratio Test of the Null Hypothesis of h Cointegrating Relations Against the Alternative of $h + 1$ Relations

Number of random walks $(g = n - h)$ (g)	*Sample size* (T)	*Probability that* $2(\mathcal{L}_A - \mathcal{L}_0)$ *is greater than entry*					
		0.500	*0.200*	*0.100*	*0.050*	*0.025*	*0.001*
			Case 1				
1	400	0.58	1.82	2.86	3.84	4.93	6.51
2	400	4.83	7.58	9.52	11.44	13.27	15.69
3	400	9.71	13.31	15.59	17.89	20.02	22.99
4	400	14.94	18.97	21.58	23.80	26.14	28.82
5	400	20.16	24.83	27.62	30.04	32.51	35.17
			Case 2				
1	400	2.415	4.905	6.691	8.083	9.658	11.576
2	400	7.474	10.666	12.783	14.595	16.403	18.782
3	400	12.707	16.521	18.959	21.279	23.362	26.154
4	400	17.875	22.341	24.917	27.341	29.599	32.616
5	400	23.132	27.953	30.818	33.262	35.700	38.858
			Case 3				
1	400	0.447	1.699	2.816	3.962	5.332	6.936
2	400	6.852	10.125	12.099	14.036	15.810	17.936
3	400	12.381	16.324	18.697	20.778	23.002	25.521
4	400	17.719	22.113	24.712	27.169	29.335	31.943
5	400	23.211	27.899	30.774	33.178	35.546	38.341

The probability shown at the head of the column is the area in the right-hand tail. The number of random walks under the null hypothesis (g) is given by the number of variables described by the vector autoregression (n) minus the number of cointegrating relations under the null hypothesis (h). In each case the alternative is that there are $h + 1$ cointegrating relations.

Source: Michael Osterwald-Lenum, "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics* 54 (1992), p. 462; and Søren Johansen and Katarina Juselius, "Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics* 52 (1990), p. 208.

C

Answers to Selected Exercises

Chapter 3. Stationary ARMA *Processes*

3.1. Yes, any *MA* process is covariance-stationary. Autocovariances:

$$\begin{aligned}\gamma_0 &= 7.4\\ \gamma_{\pm 1} &= 4.32\\ \gamma_{\pm 2} &= 0.8\\ \gamma_j &= 0 \qquad \text{for } |j| > 2.\end{aligned}$$

3.2. Yes, the process is covariance-stationary, since

$$(1 - 1.1z + 0.18z^2) = (1 - 0.9z)(1 - 0.2z);$$

the eigenvalues (0.9 and 0.2) are both inside the unit circle. The autocovariances are as follows:

$$\begin{aligned}\gamma_0 &= 7.89\\ \gamma_1 &= 7.35\\ \gamma_j &= 1.1\gamma_{j-1} - 0.18\gamma_{j-2} \qquad \text{for } j = 2, 3, \ldots\\ \gamma_{-j} &= \gamma_j.\end{aligned}$$

3.3. Equating coefficients on:

$$\begin{aligned}&L^0 \text{ gives } \psi_0 = 1\\ &L^1 \text{ gives } -\phi_1\psi_0 + \psi_1 = 0\\ &L^2 \text{ gives } -\phi_2\psi_0 - \phi_1\psi_1 + \psi_2 = 0\\ &\quad\vdots\\ &L^j \text{ gives } -\phi_p\psi_{j-p} - \phi_{p-1}\psi_{j-p+1} - \cdots - \phi_1\psi_{j-1} + \psi_j = 0,\\ &\qquad\qquad \text{for } j = p, p+1, \ldots.\end{aligned}$$

These imply

$$\begin{aligned}\psi_0 &= 1\\ \psi_1 &= \phi_1\\ \psi_2 &= \phi_1^2 + \phi_2\\ &\vdots\\ \psi_j &= \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + \cdots + \phi_p\psi_{j-p} \qquad \text{for } j = p, p+1, \ldots.\end{aligned}$$

Thus the values of ψ_j are the solution to a pth-order difference equation with starting values $\psi_0 = 1$ and $\psi_{-1} = \psi_{-2} = \cdots = \psi_{-p+1} = 0$. Thus, from the results on difference equations,

$$\begin{bmatrix}\psi_j\\ \psi_{j-1}\\ \vdots\\ \psi_{j-p+1}\end{bmatrix} = \mathbf{F}^j\begin{bmatrix}1\\0\\ \vdots\\0\end{bmatrix};$$

that is

$$\psi_j = f_{11}^{(j)}.$$

3.4. From [2.1.6],

$$\psi(L)c = (\psi_0 + \psi_1 + \psi_2 + \psi_3 + \cdots)\cdot c.$$

But the sum $(\psi_0 + \psi_1 + \psi_2 + \psi_3 + \cdots)$ can be viewed as the polynomial $\psi(z)$ evaluated at $z = 1$:

$$\psi(L)c = \psi(1)\cdot c.$$

Moreover, from [3.4.19],

$$\psi(1) = 1/(1 - \phi_1 - \phi_2).$$

3.5. Let λ_1 and λ_2 satisfy $(1 - \phi_1 z - \phi_2 z^2) = (1 - \lambda_1 z)(1 - \lambda_2 z)$, noting that λ_1 and λ_2 are both inside the unit circle for a covariance-stationary $AR(2)$ process.

Consider first the case where λ_1 and λ_2 are real and distinct. Then from [1.2.29],

$$\begin{aligned}
\sum_{j=0}^{\infty} |\psi_j| &= \sum_{j=0}^{\infty} |c_1\lambda_1^j + c_2\lambda_2^j| \\
&< \sum_{j=0}^{\infty} |c_1\lambda_1^j| + \sum_{j=0}^{\infty} |c_2\lambda_2^j| \\
&= |c_1|/(1 - |\lambda_1|) + |c_2|/(1 - |\lambda_2|) \\
&< \infty.
\end{aligned}$$

Consider next the case where λ_1 and λ_2 are distinct complex conjugates. Let $R = |\lambda_i|$ denote the modulus of λ_1 or λ_2. Then $0 \leq R < 1$, and from [1.2.39],

$$\begin{aligned}
\sum_{j=0}^{\infty} |\psi_j| &= \sum_{j=0}^{\infty} |c_1\lambda_1^j + c_2\lambda_2^j| \\
&= \sum_{j=0}^{\infty} |2\alpha R^j \cos(\theta j) - 2\beta R^j \sin(\theta j)| \\
&\leq |2\alpha| \sum_{j=0}^{\infty} R^j|\cos(\theta j)| + |2\beta| \sum_{j=0}^{\infty} R^j|\sin(\theta j)| \\
&\leq |2\alpha| \sum_{j=0}^{\infty} R^j + |2\beta| \sum_{j=0}^{\infty} R^j \\
&= 2(|\alpha| + |\beta|)/(1 - R) \\
&< \infty.
\end{aligned}$$

Finally, for the case of a repeated real root $|\lambda| < 1$,

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |k_1\lambda^j + k_2 j\lambda^{j-1}| \leq |k_1| \sum_{j=0}^{\infty} |\lambda|^j + |k_2| \sum_{j=0}^{\infty} |j\lambda^{j-1}|.$$

But

$$|k_1| \sum_{j=0}^{\infty} |\lambda|^j = |k_1|/(1 - |\lambda|) < \infty$$

and

$$\begin{aligned}
\sum_{j=0}^{\infty} |j\lambda^{j-1}| &= 1 + 2|\lambda| + 3|\lambda|^2 + 4|\lambda|^3 + \cdots \\
&= 1 + (|\lambda| + |\lambda|) + (|\lambda|^2 + |\lambda|^2 + |\lambda|^2) \\
&\quad + (|\lambda|^3 + |\lambda|^3 + |\lambda|^3 + |\lambda|^3) + \cdots \\
&= (1 + |\lambda| + |\lambda|^2 + |\lambda|^3 + \cdots) + (|\lambda| + |\lambda|^2 + |\lambda|^3 + \cdots) \\
&\quad + (|\lambda|^2 + |\lambda|^3 + \cdots) \\
&= 1/(1 - |\lambda|) + |\lambda|/(1 - |\lambda|) + |\lambda|^2/(1 - |\lambda|) + \cdots \\
&= 1/(1 - |\lambda|)^2 \\
&< \infty.
\end{aligned}$$

3.8. $(1 + 2.4z + 0.8z^2) = (1 + 0.4z)(1 + 2z)$.

The invertible operator is

$$(1 + 0.4z)(1 + 0.5z) = (1 + 0.9z + 0.2z^2),$$

so the invertible representation is

$$Y_t = (1 + 0.9L + 0.2L^2)\varepsilon_t$$
$$E(\varepsilon_t^2) = 4.$$

Chapter 4. Forecasting

4.3. $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

4.4. No. The projection of Y_4 on Y_3, Y_2, and Y_1 can be calculated from

$$\hat{P}(Y_4|Y_3, Y_2, Y_1) = a_{41}Y_1 + a_{42}[Y_2 - \hat{P}(Y_2|Y_1)] + a_{43}[Y_3 - \hat{P}(Y_3|Y_2, Y_1)].$$

The projection $\hat{P}(Y_3|Y_2, Y_1)$, in turn, is given by

$$\hat{P}(Y_3|Y_2, Y_1) = a_{31}Y_1 + a_{32}[Y_2 - \hat{P}(Y_2|Y_1)].$$

The coefficient on Y_2 in $\hat{P}(Y_4|Y_3, Y_2, Y_1)$ is therefore given by $a_{42} - a_{43}a_{32}$.

Chapter 5. Maximum Likelihood Estimation

5.2. The negative of the matrix of second derivatives is

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix},$$

so that [5.7.12] implies

$$\boldsymbol{\theta}^{(1)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Chapter 7. Asymptotic Distribution Theory

7.1. By continuity, $|g(X_T, c_T) - g(\xi, c)| > \delta$ only if $|X_T - \xi| + |c_T - c| > \eta$ for some η. But $c_T \to c$ and $X_T \xrightarrow{p} \xi$ means that we can find an N such that $|c_T - c| < \eta/2$ for all $T \geq N$ and such that $P\{|X_T - \xi| > \eta/2\} < \varepsilon$ for all $T \geq N$. Hence $P\{|X_T - \xi| + |c_T - c| > \eta\}$ is less than ε for all $T \geq N$, implying that $P\{|g(X_T, c_T) - g(\xi, c)| > \delta\} < \varepsilon$.

7.2. (a) For an $AR(1)$ process, $\psi(z) = 1/(1 - \phi z)$ and $g_Y(z) = \sigma^2/(1 - \phi z)(1 - \phi z^{-1})$, with

$$g_Y(1) = \frac{\sigma^2}{(1 - \phi)(1 - \phi)} = \frac{1}{(1 - 0.8)^2} = 25.$$

Thus $\lim_{T\to\infty} T\cdot\text{Var}(\overline{Y}_T) = 25$.

(b) $T = 10{,}000$ ($\sqrt{25/10{,}000} = 0.05$).

7.3. No, the variance can be a function of time.

7.4. Yes, ε_t has variance σ^2 for all t. Since ε_t is a martingale difference sequence, it has mean zero and must be serially uncorrelated. Thus $\{\varepsilon_t\}$ is white noise and this is a covariance-stationary $MA(\infty)$ process.

7.7. From the results of Chapter 3, Y_t can be written as $Y_t = \mu + \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$ with $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$. Then (a) follows immediately from Proposition 7.5 and result [3.3.19]. For (b), notice that $E|\varepsilon_t|^r < \infty$ for $r = 4$, so that result [7.2.14] establishes that

$$[1/(T - k)] \sum_{t=k+1}^{T} \tilde{Y}_t\tilde{Y}_{t-k} \xrightarrow{p} E(\tilde{Y}_t\tilde{Y}_{t-k}),$$

where $\tilde{Y}_t \equiv Y_t - \mu$. But

$$\begin{aligned}
[1/(T-k)] \sum_{t=k+1}^{T} Y_t Y_{t-k} &= [1/(T-k)] \sum_{t=k+1}^{T} (\tilde{Y}_t + \mu)(\tilde{Y}_{t-k} + \mu) \\
&= [1/(T-k)] \sum_{t=k+1}^{T} \tilde{Y}_t \tilde{Y}_{t-k} + \mu[1/(T-k)] \sum_{t=k+1}^{T} \tilde{Y}_{t-k} \\
&\quad + \mu[1/(T-k)] \sum_{t=k+1}^{T} \tilde{Y}_t + \mu^2 \\
&\xrightarrow{p} E(\tilde{Y}_t \tilde{Y}_{t-k}) + 0 + 0 + \mu^2 \\
&= E(\tilde{Y}_t + \mu)(\tilde{Y}_{t-k} + \mu) \\
&= E(Y_t Y_{t-k}).
\end{aligned}$$

Chapter 8. Linear Regression Models

8.1. $$\begin{aligned}
R_u^2 &= \frac{\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}} \\
&= \frac{\mathbf{y}'\mathbf{y} - \mathbf{y}'[\mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}}{\mathbf{y}'\mathbf{y}} \\
&= 1 - [(\mathbf{y}'\mathbf{M_X}\mathbf{M_X}\mathbf{y})/(\mathbf{y}'\mathbf{y})] \\
&= 1 - [(\hat{\mathbf{u}}'\hat{\mathbf{u}})/(\mathbf{y}'\mathbf{y})]. \\
R_c^2 &= \frac{\mathbf{y}'\mathbf{y}_t - \mathbf{y}'\mathbf{M_X}\mathbf{y} - T\bar{y}^2}{\mathbf{y}'\mathbf{y} - T\bar{y}^2} \\
&= 1 - [(\hat{\mathbf{u}}'\hat{\mathbf{u}})/(\mathbf{y}'\mathbf{y} - T\bar{y}^2)]
\end{aligned}$$

and

$$\mathbf{y}'\mathbf{y} - T\bar{y}^2 = \sum_{t=1}^{T} y_t^2 - T\bar{y}^2 = \sum_{t=1}^{T} (y_t - \bar{y})^2.$$

8.2. The 5% critical value for a $\chi^2(2)$ variable is 5.99. An $F(2, N)$ variable will thus have a critical value that approaches $5.99/2 = 3.00$ as $N \to \infty$. One needs N of around 300 observations before the critical value of an $F(2, N)$ variable reaches 3.03, or within 1% of the limiting value.

8.3. Fourth moments of $\mathbf{x}_t u_t$ are of the form $E(\varepsilon_t^4) \cdot E(y_{t-i} y_{t-j} y_{t-l} y_{t-m})$. The first term is bounded under Assumption 8.4, and the second term is bounded as in Example 7.14. Moreover, a typical element of $(1/T)\Sigma_{t=1}^{T} u_t^2 \mathbf{x}_t \mathbf{x}_t'$ is of the form

$$\begin{aligned}
(1/T) \sum_{t=1}^{T} \varepsilon_t^2 y_{t-i} y_{t-j} &= (1/T) \sum_{t=1}^{T} (\varepsilon_t^2 - \sigma^2) y_{t-i} y_{t-j} + \sigma^2 \cdot (1/T) \sum_{t=1}^{T} y_{t-i} y_{t-j} \\
&\xrightarrow{p} 0 + \sigma^2 \cdot E(y_{t-i} y_{t-j}).
\end{aligned}$$

Hence, the conditions of Proposition 7.9 are satisfied.

8.4. Proposition 7.5 and result [7.2.14] establish

$$\begin{bmatrix} \hat{c}_T \\ \hat{\phi}_{1,T} \\ \vdots \\ \hat{\phi}_{p,T} \end{bmatrix} = \begin{bmatrix} 1 & (1/T)\Sigma y_{t-1} & \cdots & (1/T)\Sigma y_{t-p} \\ (1/T)\Sigma y_{t-1} & (1/T)\Sigma y_{t-1}^2 & \cdots & (1/T)\Sigma y_{t-1} y_{t-p} \\ \vdots & \vdots & \cdots & \vdots \\ (1/T)\Sigma y_{t-p} & (1/T)\Sigma y_{t-p} y_{t-1} & \cdots & (1/T)\Sigma y_{t-p}^2 \end{bmatrix}^{-1} \begin{bmatrix} (1/T)\Sigma y_t \\ (1/T)\Sigma y_{t-1} y_t \\ \vdots \\ (1/T)\Sigma y_{t-p} y_t \end{bmatrix}$$

$$\xrightarrow{p} \begin{bmatrix} 1 & \mu & \cdots & \mu \\ \mu & \gamma_0 + \mu^2 & \cdots & \gamma_{p-1} + \mu^2 \\ \vdots & \vdots & \cdots & \vdots \\ \mu & \gamma_{p-1} + \mu^2 & \cdots & \gamma_0 + \mu^2 \end{bmatrix}^{-1} \begin{bmatrix} \mu \\ \gamma_1 + \mu^2 \\ \vdots \\ \gamma_p + \mu^2 \end{bmatrix},$$

which equals $\boldsymbol{\alpha}^{(p)}$ given in [4.3.6].

10.2. (a) $$\Gamma_0 = \begin{bmatrix} (1+\theta^2)\sigma_\varepsilon^2 & h_1\theta\sigma_\varepsilon^2 \\ h_1\theta\sigma_\varepsilon^2 & \{h_1^2(1+\theta^2)\sigma_\varepsilon^2 + \sigma_u^2\} \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} \theta\sigma_\varepsilon^2 & 0 \\ h_1(1+\theta^2)\sigma_\varepsilon^2 & h_1^2\theta\sigma_\varepsilon^2 \end{bmatrix}$$

$$\Gamma_2 = \begin{bmatrix} 0 & 0 \\ h_1\theta\sigma_\varepsilon^2 & 0 \end{bmatrix}$$

$$\Gamma_{-1} = \Gamma_1' \qquad \Gamma_{-2} = \Gamma_2'$$

$$\Gamma_k = \mathbf{0} \qquad \text{for } k = \pm 3, \pm 4, \ldots .$$

(b) $$\mathbf{s}_Y(\omega) = (2\pi)^{-1}\begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$

$$s_{11} = (1+\theta^2)\sigma_\varepsilon^2 + \theta\sigma_\varepsilon^2 e^{-i\omega} + \theta\sigma_\varepsilon^2 e^{i\omega}$$

$$s_{12} = h_1\theta\sigma_\varepsilon^2 e^{2i\omega} + h_1(1+\theta^2)\sigma_\varepsilon^2 e^{i\omega} + h_1\theta\sigma_\varepsilon^2$$

$$s_{21} = h_1\theta\sigma_\varepsilon^2 e^{-2i\omega} + h_1(1+\theta^2)\sigma_\varepsilon^2 e^{-i\omega} + h_1\theta\sigma_\varepsilon^2$$

$$s_{22} = h_1^2(1+\theta^2)\sigma_\varepsilon^2 + \sigma_u^2 + h_1^2\theta\sigma_\varepsilon^2 e^{-i\omega} + h_1^2\theta\sigma_\varepsilon^2 e^{i\omega}$$

$$c_{YX}(\omega) = (2\pi)^{-1}h_1\sigma_\varepsilon^2\{\theta\cdot\cos(2\omega) + (1+\theta^2)\cdot\cos(\omega) + \theta\}$$

$$q_{YX}(\omega) = -(2\pi)^{-1}h_1\sigma_\varepsilon^2\{\theta\cdot\sin(2\omega) + (1+\theta^2)\cdot\sin(\omega)\}.$$

(c) The variable X_t follows an $MA(1)$ process, for which the spectrum is indeed s_{11}. The term s_{21} is s_{11} times $h(e^{-i\omega}) = h_1\cdot e^{-i\omega}$. Multiplying s_{21} in turn by $h(e^{i\omega}) = h_1\cdot e^{i\omega}$ and adding σ_u^2 produces s_{22}.

(d) $$(2\pi)^{-1}\int_{-\pi}^{\pi}\frac{s_{YX}(\omega)}{s_{XX}(\omega)}e^{i\omega k}\,d\omega = (2\pi)^{-1}\int_{-\pi}^{\pi} h_1\cdot e^{-i\omega}e^{i\omega k}\,d\omega.$$

When $k = 1$, this is simply

$$(2\pi)^{-1}\int_{-\pi}^{\pi} h_1\,d\omega = h_1,$$

as desired. When $k \neq 1$, the integral is

$$(2\pi)^{-1}\int_{-\pi}^{\pi} h_1\cdot e^{(k-1)i\omega}\,d\omega$$

$$= (2\pi)^{-1}\int_{-\pi}^{\pi} h_1\cdot\cos[(k-1)\omega]\,d\omega + i\cdot(2\pi)^{-1}\int_{-\pi}^{\pi} h_1\cdot\sin[(k-1)\omega]\,d\omega$$

$$= [(k-1)2\pi]^{-1}h_1\Big[\sin[(k-1)\omega]\Big]_{\omega=-\pi}^{\pi} - [(k-1)2\pi]^{-1}h_1\Big[\cos[(k-1)\omega]\Big]_{\omega=-\pi}^{\pi}$$

$$= 0.$$

Chapter 11. Vector Autoregressions

11.1. A typical element of [11.A.2] states that

$$(1/T)\sum_{t=1}^{T}\varepsilon_{i_1,t}y_{j_1,t-l_1}\varepsilon_{i_2,t}y_{j_2,t-l_2} \xrightarrow{p} E(\varepsilon_{i_1,t}\varepsilon_{i_2,t})\cdot E(y_{j_1,t-l_1}y_{j_2,t-l_2}).$$

But

$$(1/T)\sum_{t=1}^{T}\varepsilon_{i_1,t}y_{j_1,t-l_1}\varepsilon_{i_2,t}y_{j_2,t-l_2} = (1/T)\sum_{t=1}^{T}z_t + E(\varepsilon_{i_1,t}\varepsilon_{i_2,t})\cdot(1/T)\sum_{t=1}^{T}y_{j_1,t-l_1}y_{j_2,t-l_2},$$

where

$$z_t \equiv \{\varepsilon_{i_1,t}\varepsilon_{i_2,t} - E(\varepsilon_{i_1,t}\varepsilon_{i_2,t})\}y_{j_1,t-l_1}y_{j_2,t-l_2}.$$

Notice that z_t is a martingale difference sequence whose variance is finite by virtue of

Proposition 7.10. Hence, $(1/T)\Sigma_{t=1}^{T} z_t \xrightarrow{p} 0$. Moreover,

$$(1/T) \sum_{t=1}^{T} y_{j_1,t-l_1} y_{j_2,t-l_2} \xrightarrow{p} E(y_{j_1,t-l_1} y_{j_2,t-l_2}),$$

by virtue of Proposition 10.2(d).

11.2. (a) No. (b) Yes. (c) No.

11.3.
$$\begin{aligned}
\alpha_j &= \zeta_j && \text{for } j = 1, 2, \ldots, p \\
\beta_j &= \eta_j && \text{for } j = 1, 2, \ldots, p \\
\lambda_0 &= \Omega_{21}\Omega_{11}^{-1} && \\
\lambda_j &= \gamma_j - \Omega_{21}\Omega_{11}^{-1}\alpha_j && \text{for } j = 1, 2, \ldots, p \\
\xi_j &= \delta_j - \Omega_{21}\Omega_{11}^{-1}\beta_j && \text{for } j = 1, 2, \ldots, p \\
\sigma_1^2 &= \Omega_{11} && \\
\sigma_2^2 &= \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} && \\
u_{1t} &= \varepsilon_{1t} && \\
u_{2t} &= \varepsilon_{2t} - \Omega_{21}\Omega_{11}^{-1}\varepsilon_{1t} &&
\end{aligned}$$

11.4. Premultiplying by $\mathbf{A}^*(L)$ results in

$$\begin{aligned}
\begin{bmatrix} |\mathbf{A}(L)| & 0 \\ 0 & |\mathbf{A}(L)| \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} &= \begin{bmatrix} 1 - \xi(L) & \eta(L) \\ \lambda_0 + \lambda(L) & 1 - \zeta(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \\
&= \begin{bmatrix} [1 - \xi(L)]u_{1t} + \eta(L)u_{2t} \\ [\lambda_0 + \lambda(L)]u_{1t} + [1 - \zeta(L)]u_{2t} \end{bmatrix} \\
&= \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}.
\end{aligned}$$

Thus,

$$\begin{aligned}
|\mathbf{A}(L)|y_{1t} &= v_{1t} \\
|\mathbf{A}(L)|y_{2t} &= v_{2t}.
\end{aligned}$$

Now the determinant $|\mathbf{A}(L)|$ is the following polynomial in the lag operator:

$$|\mathbf{A}(L)| = [1 - \xi(L)][1 - \zeta(L)] - [\eta(L)][\lambda_0 + \lambda(L)].$$

The coefficient on L^0 in this polynomial is unity, and the highest power of L is L^{2p}, which has coefficient $(\xi_p\zeta_p - \eta_p\lambda_p)$:

$$|\mathbf{A}(L)| = 1 + a_1L + a_2L^2 + \cdots + a_{2p}L^{2p}.$$

Furthermore, v_{1t} is the sum of two mutually nncorrelated $MA(p)$ processes, and so v_{1t} is itself $MA(p)$. Hence, y_{1t} follows an $ARMA(2p, p)$ process; a similar argument shows that y_{2t} follows an $ARMA(2p, p)$ process with the same autoregressive coefficients but different moving average coefficients.

In general, consider an n-variable VAR of the form

$$\mathbf{\Phi}(L)\mathbf{y}_t = \boldsymbol{\varepsilon}_t$$

with

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_\tau') = \begin{cases} \mathbf{\Omega} & \text{if } t = \tau \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Find the triangular factorization of $\mathbf{\Omega} = \mathbf{ADA}'$ and premultiply the system by $\mathbf{A}^{-1}$, yielding

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{u}_t,$$

where

$$\begin{aligned}
\mathbf{A}(L) &= \mathbf{A}^{-1}\mathbf{\Phi}(L) \\
\mathbf{u}_t &= \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t \\
E(\mathbf{u}_t\mathbf{u}_t') &= \mathbf{D}.
\end{aligned}$$

Thus, the elements of $\mathbf{u}_t$ are mutually uncorrelated and $\mathbf{A}(0)$ has 1s along its principal diagonal. The adjoint matrix $\mathbf{A}^*(L)$ has the property

$$\mathbf{A}^*(L) \cdot \mathbf{A}(L) = |\mathbf{A}(L)| \cdot \mathbf{I}_n.$$

Premultiplying the system by $\mathbf{A}^*(L)$,

$$|\mathbf{A}(L)| \cdot \mathbf{y}_t = \mathbf{A}^*(L)\mathbf{u}_t.$$

The determinant $|\mathbf{A}(L)|$ is a scalar polynomial containing terms up to order L^{np}, while elements of $\mathbf{A}^*(L)$ contain terms up to order $L^{(n-1)p}$. Hence, the ith row of the system takes the form

$$|\mathbf{A}(L)| \cdot y_{it} = v_{it},$$

where v_{it} is the sum of n mutually uncorrelated $MA[(n - 1)p]$ processes and is therefore itself $MA[(n - 1)p]$. Hence, $y_{it} \sim ARMA[np, (n - 1)p]$.

11.5. (a) $|\mathbf{I}_2 - \mathbf{\Phi}_1 z| = (1 - 0.3z)(1 - 0.4z) - (0.8z)(0.9z)$

$$= 1 - 0.7z - 0.6z^2$$
$$= (1 - 1.2z)(1 + 0.5z).$$

Since $z^* = 1/1.2$ is inside the unit circle, the system is nonstationary.

(b) $\mathbf{\Psi}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{\Psi}_1 = \begin{bmatrix} 0.3 & 0.8 \\ 0.9 & 0.4 \end{bmatrix} \quad \mathbf{\Psi}_2 = \begin{bmatrix} 0.81 & 0.56 \\ 0.63 & 0.88 \end{bmatrix}$

$\mathbf{\Psi}_s$ diverges as $s \to \infty$.

(c) $y_{1,t+2} - \hat{E}(y_{1,t+2}|\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots) = \varepsilon_{1,t+2} + 0.3\varepsilon_{1,t+1} + 0.8\varepsilon_{2,t+1}$

$$MSE = 1 + (0.3)^2 + (0.8)^2(2) = 2.37.$$

The fraction due to $\varepsilon_1 = 1.09/2.37 = 0.46$.

Chapter 12. Bayesian Analysis

12.1. Take $k = 1$, $\mathbf{X} = \mathbf{1}$, $\boldsymbol{\beta} = \mu$, and $\mathbf{M} = 1/\nu$, and notice that $\mathbf{1}'\mathbf{1} = T$ and $\mathbf{1}'\mathbf{y} = T\bar{y}$.

Chapter 13. The Kalman Filter

13.3. No, because $\mathbf{v}_t$ is not white noise.

13.5. Notice that

$$\begin{aligned}
\tilde{\sigma}^2 + \tilde{\theta}^2\tilde{p}_{t+1} &= \frac{\tilde{\sigma}^2(1 + \tilde{\theta}^2 + \tilde{\theta}^4 + \cdots + \tilde{\theta}^{2[t+1]})}{1 + \tilde{\theta}^2 + \tilde{\theta}^4 + \cdots + \tilde{\theta}^{2t}} \\
&= \frac{\tilde{\sigma}^2(1 - \tilde{\theta}^{2[t+2]})}{1 - \tilde{\theta}^{2[t+1]}} \\
&= \frac{\theta^2\sigma^2(1 - \theta^{-2[t+2]})}{1 - \theta^{-2[t+1]}} \\
&= \frac{\theta^2\sigma^2(\theta^{2[t+2]} - 1)}{\theta^{2[t+2]} - \theta^2} \\
&= \frac{\sigma^2(1 - \theta^{2[t+2]})}{1 - \theta^{2[t+1]}} \\
&= \sigma^2 + \theta^2 p_{t+1}.
\end{aligned}$$

Furthermore, from [13.3.19],

$$\begin{aligned}\bar{\theta}\hat{\bar{\varepsilon}}_{t|t} &= \{\bar{\theta}\bar{\sigma}^2/[\bar{\sigma}^2 + \bar{\theta}^2\tilde{p}_t]\}\cdot\{y_t - \mu - \bar{\theta}\hat{\bar{\varepsilon}}_{t-1|t-1}\}\\ &= \{\theta^{-1}\theta^2\sigma^2/[\sigma^2 + \theta^2 p_t]\}\cdot\{y_t - \mu - \bar{\theta}\hat{\bar{\varepsilon}}_{t-1|t-1}\}\\ &= \{\theta\sigma^2/[\sigma^2 + \theta^2 p_t]\}\cdot\{y_t - \mu - \bar{\theta}\hat{\bar{\varepsilon}}_{t-1|t-1}\},\end{aligned}$$

which is the same difference equation that generates $\{\theta\hat{\varepsilon}_{t|t}\}$, with both sequences, of course, beginning with $\theta\hat{\varepsilon}_{0|0} = \bar{\theta}\hat{\bar{\varepsilon}}_{0|0} = 0$. With the sequences $(\mathbf{H}'\mathbf{P}_{t+1|t}\mathbf{H} + \mathbf{R})$ and $\mathbf{A}'\mathbf{x}_{t+1} + \mathbf{H}'\hat{\boldsymbol{\xi}}_{t+1|t}$ identical for the two representations, the likelihood in [13.4.1] to [13.4.3] must be identical.

13.6. The innovation ε_t in [13.5.22] will be fundamental when $|\phi - K| < 1$. From [13.5.25], we see that

$$\phi - K = \phi\sigma_W^2/(\sigma_W^2 + P).$$

Since P is a variance, it follows that $P \geq 0$, and so $|\phi - K| \leq |\phi|$, which is specified to be less than unity. This arises as a consequence of the general result in Proposition 13.2 that the eigenvalues of $\mathbf{F} - \mathbf{K}\mathbf{H}'$ lie inside the unit circle.

From [13.5.23] and the preceding expression for $\phi - K$,

$$-(\phi - K)E(\varepsilon_t^2) = -(\phi - K)(\sigma_W^2 + P) = -\phi\sigma_W^2,$$

as claimed. Furthermore,

$$\begin{aligned}[1 + (\phi - K)^2]E(\varepsilon_t^2) &= (\sigma_W^2 + P) + (\phi - K)\phi\sigma_W^2\\ &= (1 + \phi^2)\sigma_W^2 + P - K\phi\sigma_W^2.\end{aligned}$$

But from [13.5.24] and [13.5.25],

$$P = K\phi\sigma_W^2 + \sigma_V^2,$$

and so

$$[1 + (\phi - K)^2]E(\varepsilon_t^2) = (1 + \phi^2)\sigma_W^2 + \sigma_V^2.$$

To understand these formulas from the perspective of the formulas in Chapter 4, note that the model adds an $AR(1)$ process to white noise, producing an $ARMA(1, 1)$ process:

$$(1 - \phi L)y_{t+1} = v_{t+1} + (1 - \phi L)w_{t+1}.$$

The first autocovariance of the $MA(1)$ process on the right side of this expression is $-\phi\sigma_W^2$, while the variance is $(1 + \phi^2)\sigma_W^2 + \sigma_V^2$.

Chapter 16. Processes with Deterministic Time Trends

16.1. $$E\left((1/T)\sum_{t=1}^{T}[\lambda_1 + \lambda_2(t/T)]^2\varepsilon_t^2 - (1/T)\sum_{t=1}^{T}\sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2]\right)^2$$
$$= (1/T^2)\sum_{t=1}^{T}[\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2]^2\cdot E(\varepsilon_t^2 - \sigma^2)^2.$$

But

$$(1/T)\sum_{t=1}^{T}[\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2]^2 \to M < \infty,$$

and thus

$$T\cdot E\left((1/T)\sum_{t=1}^{T}[\lambda_1 + \lambda_2(t/T)]^2\varepsilon_t^2 - (1/T)\sum_{t=1}^{T}\sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2]\right)^2$$
$$\to M\cdot E(\varepsilon_t^2 - \sigma^2)^2 < \infty.$$

Thus

$$\begin{aligned}&(1/T)\sum_{t=1}^{T}[\lambda_1 + \lambda_2(t/T)]^2\varepsilon_t^2\\ &\xrightarrow{m.s.} (1/T)\sum_{t=1}^{T}\sigma^2[\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2]\\ &\to \sigma^2\,\boldsymbol{\lambda}'\mathbf{Q}\boldsymbol{\lambda}.\end{aligned}$$

16.2. Recall that the variance of $\mathbf{b}_T$ is given by

$$E(\mathbf{b}_T - \boldsymbol{\beta})(\mathbf{b}_T - \boldsymbol{\beta})' = \sigma^2\left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}$$
$$= \sigma^2\begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix}^{-1}.$$

Pre- and postmultiplying by $\mathbf{Y}_T$ results in

$$E[\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})(\mathbf{b}_T - \boldsymbol{\beta})'\mathbf{Y}_T]$$
$$= \sigma^2\mathbf{Y}_T\begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix}^{-1}\mathbf{Y}_T$$
$$= \sigma^2\cdot\left\{\mathbf{Y}_T^{-1}\begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix}\mathbf{Y}_T^{-1}\right\}^{-1}$$
$$\rightarrow \sigma^2\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1}.$$

The (2, 2) element of this matrix expression holds that

$$E[T^{3/2}(\hat{\delta}_T - \delta)]^2 \rightarrow 12\sigma^2,$$

and so

$$T(\hat{\delta}_T - \delta) \xrightarrow{m.s.} 0.$$

16.3. Notice that

$$\left[T^{-1}\sum_{t=1}^{T}(t/T)y_t\right]^2 = T^{-2}[(1/T)y_1 + (2/T)y_2 + \cdots + (T/T)y_T]$$
$$\times [(1/T)y_1 + (2/T)y_2 + \cdots + (T/T)y_T],$$

which has expectation

$$E\left[T^{-1}\sum_{t=1}^{T}(t/T)y_t\right]^2$$
$$= T^{-2}\{[(1/T)^2 + (2/T)^2 + \cdots + (T/T)^2]\gamma_0$$
$$+ [(1/T)(2/T) + (2/T)(3/T) + \cdots + [(T-1)/T](T/T)]2\gamma_1$$
$$+ [(1/T)(3/T) + (2/T)(4/T) + \cdots + [(T-2)/T](T/T)]2\gamma_2$$
$$+ \cdots + [(1/T)(T/T)]2\gamma_{T-1}\}$$
$$\leq T^{-1}\{|\gamma_0| + 2|\gamma_1| + 2|\gamma_2| + \cdots + 2|\gamma_{T-1}|\}$$
$$\rightarrow 0.$$

Chapter 17. Univariate Processes with Unit Roots

17.2. (a) $T(\hat{\rho}_T - 1) = \dfrac{T^{-1}\Sigma y_{t-1}u_t}{T^{-2}\Sigma y_{t-1}^2} \xrightarrow{L} \dfrac{\frac{1}{2}\{\lambda^2\cdot[W(1)]^2 - \gamma_0\}}{\lambda^2\cdot\int [W(r)]^2\,dr}$

from Proposition 17.3(e) and (h).

(b) $$T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T} = T^2\cdot s_T^2 \div (\Sigma y_{t-1}^2)$$
$$= s_T^2 \div (T^{-2}\Sigma y_{t-1}^2)$$
$$\xrightarrow{L} \gamma_0 \div \left(\lambda^2\cdot\int [W(r)]^2\,dr\right),$$

from Proposition 17.3(h) and [17.6.10].

(c) $$t_T = T(\hat{\rho}_T - 1) \div (T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T})^{1/2}$$
$$\xrightarrow{L} \frac{\frac{1}{2}\{\lambda^2\cdot[W(1)]^2 - \gamma_0\}}{\lambda^2\cdot\int [W(r)]^2\,dr} \times \left(\lambda^2\cdot\int [W(r)]^2\,dr\right)^{1/2} \div (\gamma_0)^{1/2},$$

from answers (a) and (b). This, in turn, can be written

$$(\lambda^2/\gamma_0)^{1/2}\frac{\frac{1}{2}\{\lambda^2\cdot[W(1)]^2-\gamma_0\}}{\lambda^2\left\{\int[W(r)]^2\,dr\right\}^{1/2}}=(\lambda^2/\gamma_0)^{1/2}\left\{\frac{\frac{1}{2}\{[W(1)]^2-1\}}{\left\{\int[W(r)]^2\,dr\right\}^{1/2}}+\frac{\frac{1}{2}(\lambda^2-\gamma_0)}{\lambda^2\left\{\int[W(r)]^2\,dr\right\}^{1/2}}\right\}.$$

$$\text{(d)}\quad (T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T}\div s_T^2)=1/(T^{-2}\Sigma y_{t-1}^2)\xrightarrow{L}1/\left(\lambda^2\cdot\int[W(r)]^2\,dr\right),$$

from Proposition 17.3(h). Thus,

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

with the next-to-last line following from answer (a).

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

from answers (c) and (b). Adding these terms produces the desired result.

To estimate γ_0 and λ, one could use

$$\hat{\gamma}_j=T^{-1}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}\qquad\text{for } j=0,1,\ldots,q$$

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

where $\hat{u}_t$ is the *OLS* sample residual and q is the number of autocovariances used to represent the serial correlation of $\psi(L)\varepsilon_t$. The statistic in (d) can then be compared with the case 1 entries of Table B.5, while the statistic in (e) can be compared with the case 1 entries of Table B.6.

17.3. (a)
$$\begin{bmatrix}1 & T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma t\\ T^{-3/2}\Sigma\xi_{t-1} & T^{-2}\Sigma\xi_{t-1}^2 & T^{-5/2}\Sigma\xi_{t-1}t\\ T^{-2}\Sigma t & T^{-5/2}\Sigma t\xi_{t-1} & T^{-3}\Sigma t^2\end{bmatrix}$$

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

(b) $$\begin{bmatrix} T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma \xi_{t-1}u_t \\ T^{-3/2}\Sigma t u_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \lambda\cdot W(1) \\ (1/2)\{\lambda^2\cdot[W(1)]^2 - \gamma_0\} \\ \lambda\cdot\{W(1) - \int W(r)\,dr\} \end{bmatrix}.$$

(c) This follows from expression [17.4.52] and answers (a) and (b).

(d) The calculations are virtually identical to those in [17.4.54].

(e) $t_T = T(\hat{\rho}_T - 1) \div \{T^2\cdot\hat{\sigma}^2_{\hat{\rho}_T}\}^{1/2} \xrightarrow{p} T(\hat{\rho}_T - 1) \div \{(s_T^2/\lambda^2)\cdot Q\}^{1/2}$.

(f) Answer (c) establishes that

$$T(\hat{\rho}_T - 1)$$
$$\xrightarrow{L} \left\{ [0 \quad 1 \quad 0] \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \frac{1}{2}\{[W(1)]^2 - 1\} \\ W(1) - \int W(r)\,dr \end{bmatrix} \right\}$$
$$+ \tfrac{1}{2}\{1 - (\gamma_0/\lambda^2)\}[0 \quad 1 \quad 0] \begin{bmatrix} 1 & \int W(r)\,dr & 1/2 \\ \int W(r)\,dr & \int [W(r)]^2\,dr & \int rW(r)\,dr \\ 1/2 & \int rW(r)\,dr & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$
$$= V + \tfrac{1}{2}\{1 - (\gamma_0/\lambda^2)\}Q.$$

Moreover, answer (d) implies that

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

(g) From answers (d) and (e),

$$(\gamma_0/\lambda^2)^{1/2}\cdot t_T - \{\tfrac{1}{2}(\lambda^2 - \gamma_0)\lambda\} \times \{T\cdot\hat{\sigma}_{\hat{\rho}_T} \div s_T\}$$
$$\xrightarrow{p} T(\hat{\rho}_T - 1)/\sqrt{Q} - \{\tfrac{1}{2}(\lambda^2 - \gamma_0)/\lambda\} \times \sqrt{Q}/\lambda$$
$$= \{T(\hat{\rho}_T - 1) - \tfrac{1}{2}(Q/\lambda^2)(\lambda^2 - \gamma_0)\} \div \sqrt{Q}$$
$$\xrightarrow{L} V \div \sqrt{Q},$$

from the analysis of (f).

To estimate γ_0 and λ, one could use

$$\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T} \hat{u}_t\hat{u}_{t-j} \qquad \text{for } j = 0, 1, \ldots, q$$
$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{q} [1 - j/(q+1)]\hat{\gamma}_j,$$

where $\hat{u}_t$ is the *OLS* sample residual and q is the number of autocovariances used to approximate the dynamics of $\psi(L)\varepsilon_t$. The statistic in (f) can then be compared with the case 4 entries of Table B.5, while the statistic in (g) can be compared with the case 4 entries of Table B.6.

17.4. (b) Case 1 of Table B.5 is appropriate asymptotically. (c) Case 1 of Table B.6 is appropriate asymptotically.

Chapter 18. Unit Roots in Multivariate Time Series

18.1. Under the null hypothesis $\mathbf{R\beta} = \mathbf{r}$, we have

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

For $\mathbf{Y}_T$ the $(k \times k)$ matrix defined in [18.2.18] and $\mathbf{R}$ of the specified form, observe that $\sqrt{T}\cdot\mathbf{R} = \mathbf{R}\mathbf{Y}_T$. Thus,

$$\begin{aligned}
\chi_T^2 &= [\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})]'\Big[s_T^2\mathbf{R}\mathbf{Y}_T(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{Y}_T\mathbf{R}'\Big]^{-1}[\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})] \\
&= [\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})]'\Big[s_T^2\mathbf{R}(\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t'\mathbf{Y}_T^{-1})^{-1}\mathbf{R}'\Big]^{-1}[\mathbf{R}\mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta})] \\
&\xrightarrow{L} \left(\mathbf{R}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2\end{bmatrix}\right)'\left(\sigma_{ii}\mathbf{R}\begin{bmatrix}\mathbf{V}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1}\end{bmatrix}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2\end{bmatrix}\right) \\
&= (\mathbf{R}_1\mathbf{V}^{-1}\mathbf{h}_1)'(\sigma_{ii}\mathbf{R}_1\mathbf{V}^{-1}\mathbf{R}_1')^{-1}(\mathbf{R}_1\mathbf{V}^{-1}\mathbf{h}_1),
\end{aligned}$$

where the indicated convergence follows from [18.2.25], [18.2.20], and consistency of s_T^2. Since $\mathbf{h}_1 \sim N(\mathbf{0}, \sigma_{ii}\mathbf{V})$, it follows that

$$\mathbf{R}_1\mathbf{V}^{-1}\mathbf{h}_1 \sim N(\mathbf{0}, \sigma_{ii}\mathbf{R}_1\mathbf{V}^{-1}\mathbf{R}_1').$$

Hence, from Proposition 8.1, the asymptotic distribution of χ_T^2 is $\chi^2(m)$.

18.2. Here

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

where $\mathbf{x}_t$ is as defined in Exercise 18.1 and

$$\underset{(n_2p \times k)}{\mathbf{R}} = \begin{bmatrix}\underset{[n_2(p-1)\times n(p-1)]}{(\mathbf{I}_{p-1} \otimes \mathbf{R}_1)} & \underset{[n_2(p-1)\times(n+1)]}{\mathbf{0}} \\ \underset{[n_2 \times n(p-1)]}{\mathbf{0}} & \underset{[n_2\times(n+1)]}{\mathbf{R}_2}\end{bmatrix}$$

$$\underset{(n_2 \times n)}{\mathbf{R}_1} = \begin{bmatrix}\underset{(n_2\times n_1)}{\mathbf{0}} & \underset{(n_2\times n_2)}{\mathbf{I}_{n_2}}\end{bmatrix}$$

$$\underset{[n_2\times(n+1)]}{\mathbf{R}_2} = \begin{bmatrix}\underset{(n_2\times 1)}{\mathbf{0}} & \underset{(n_2\times n)}{\mathbf{R}_1}\end{bmatrix}.$$

From the results of Exercise 18.1,

$$\begin{aligned}
\chi_T^2 &\xrightarrow{L} \left(\mathbf{R}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2\end{bmatrix}\right)'\left(\sigma_{ii}\mathbf{R}\begin{bmatrix}\mathbf{V}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{-1}\end{bmatrix}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\begin{bmatrix}\mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2\end{bmatrix}\right) \\
&= \begin{bmatrix}(\mathbf{I}_{p-1} \otimes \mathbf{R}_1)\mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{R}_2\mathbf{Q}^{-1}\mathbf{h}_2\end{bmatrix}'\sigma_{ii}^{-1}\cdot\begin{bmatrix}(\mathbf{I}_{p-1} \otimes \mathbf{R}_1)\mathbf{V}^{-1}(\mathbf{I}_{p-1} \otimes \mathbf{R}_1') & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2\mathbf{Q}^{-1}\mathbf{R}_2'\end{bmatrix}^{-1} \\
&\quad\times\begin{bmatrix}(\mathbf{I}_{p-1} \otimes \mathbf{R}_1)\mathbf{V}^{-1}\mathbf{h}_1 \\ \mathbf{R}_2\mathbf{Q}^{-1}\mathbf{h}_2\end{bmatrix}.
\end{aligned}$$

18.3. (a) The null hypothesis is that $\phi = 1$ and $\gamma = \alpha = \eta = 0$, in which case $\Delta y_{2t} = \varepsilon_{2t}$ and $u_t = \varepsilon_{1t}$. Let $\mathbf{x}_t \equiv (\varepsilon_{2t}, 1, y_{1,t-1}, y_{2,t-1})'$ and

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

Then

$$\begin{aligned}
&\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t'\mathbf{Y}_T^{-1} \\
&= \begin{bmatrix}T^{-1}\Sigma\varepsilon_{2t}^2 & T^{-1}\Sigma\varepsilon_{2t} & T^{-3/2}\Sigma\varepsilon_{2t}y_{1,t-1} & T^{-3/2}\Sigma\varepsilon_{2t}y_{2,t-1} \\ T^{-1}\Sigma\varepsilon_{2t} & 1 & T^{-3/2}\Sigma y_{1,t-1} & T^{-3/2}\Sigma y_{2,t-1} \\ T^{-3/2}\Sigma y_{1,t-1}\varepsilon_{2t} & T^{-3/2}\Sigma y_{1,t-1} & T^{-2}\Sigma y_{1,t-1}^2 & T^{-2}\Sigma y_{1,t-1}y_{2,t-1} \\ T^{-3/2}\Sigma y_{2,t-1}\varepsilon_{2t} & T^{-3/2}\Sigma y_{2,t-1} & T^{-2}\Sigma y_{2,t-1}y_{1,t-1} & T^{-2}\Sigma y_{2,t-1}^2\end{bmatrix} \\
&\xrightarrow{L} \begin{bmatrix}\sigma_2^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{Q}\end{bmatrix},
\end{aligned}$$

where

$$\mathbf{Q} \equiv \begin{bmatrix} 1 & \sigma_1\cdot\int W_1(r)\,dr & \sigma_2\cdot\int W_2(r)\,dr \\ \sigma_1\cdot\int W_1(r)\,dr & \sigma_1^2\cdot\int [W_1(r)]^2\,dr & \sigma_1\sigma_2\cdot\int [W_1(r)]\cdot[W_2(r)]\,dr \\ \sigma_2\cdot\int W_2(r)\,dr & \sigma_2\sigma_1\cdot\int [W_2(r)]\cdot[W_1(r)]\,dr & \sigma_2^2\cdot\int [W_2(r)]^2\,dr \end{bmatrix}$$

and

$$\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t u_t = \begin{bmatrix} T^{-1/2}\Sigma\varepsilon_{2t}\varepsilon_{1t} \\ T^{-1/2}\Sigma\varepsilon_{1t} \\ T^{-1}\Sigma y_{1,t-1}\varepsilon_{1t} \\ T^{-1}\Sigma y_{2,t-1}\varepsilon_{1t} \end{bmatrix} \xrightarrow{L} \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix},$$

and where $h_1 \sim N(0, \sigma_1^2\sigma_2^2)$ and the second and third elements of the (3×1) vector $\mathbf{h}_2$ have a nonstandard distribution. Hence,

$$\begin{aligned} \mathbf{Y}_T(\mathbf{b}_T - \boldsymbol{\beta}) &= (\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t'\mathbf{Y}_T^{-1})^{-1}(\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t u_t) \\ &\xrightarrow{L} \begin{bmatrix} \sigma_2^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_2^{-2}h_1 \\ \mathbf{Q}^{-1}\mathbf{h}_2 \end{bmatrix}. \end{aligned}$$

(b) Let $\mathbf{e}_1$ denote the first column of the (4×4) identity matrix. Then

$$\begin{aligned} t_T &= \hat{\gamma}_T \div \left\{s_T^2\,\mathbf{e}_1'(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{e}_1\right\}^{1/2} \\ &= T^{1/2}\hat{\gamma}_T \div \left\{s_T^2\mathbf{e}_1'\mathbf{Y}_T(\Sigma\mathbf{x}_t\mathbf{x}_t')^{-1}\mathbf{Y}_T\mathbf{e}_1\right\}^{1/2} \\ &= T^{1/2}\hat{\gamma}_T \div \left\{s_T^2\mathbf{e}_1'(\mathbf{Y}_T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t'\mathbf{Y}_T^{-1})^{-1}\mathbf{e}_1\right\}^{1/2} \\ &\xrightarrow{L} \sigma_2^{-2}h_1 \div \left\{\sigma_1^2\mathbf{e}_1'\begin{bmatrix} \sigma_2^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}^{-1}\mathbf{e}_1\right\}^{1/2} \\ &= h_1/(\sigma_1\sigma_2) \sim N(0, 1). \end{aligned}$$

(c) Recall that $\hat{\delta}_T = \hat{\eta}_T - \hat{\gamma}_T$, where $\hat{\eta}_T$ is $O_p(T^{-1})$ and $\hat{\gamma}_T$ is $O_p(T^{-1/2})$. Under the null hypothesis, all three values are zero; hence,

$$T^{1/2}\hat{\delta}_T \xrightarrow{p} -T^{1/2}\hat{\gamma}_T,$$

which is asymptotically Gaussian. The t test of $\delta = 0$ is asymptotically equivalent to the t test of $\gamma = 0$.

Chapter 19. Cointegration

19.1. (a) The *OLS* estimates are given by

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\gamma}_T \end{bmatrix} = \begin{bmatrix} T & \Sigma y_{2t} \\ \Sigma y_{2t} & \Sigma y_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_{1t} \\ \Sigma y_{2t}y_{1t} \end{bmatrix},$$

from which

$$\begin{aligned} \begin{bmatrix} \hat{\alpha}_T \\ \hat{\gamma}_T - \gamma_0 \end{bmatrix} &= \begin{bmatrix} T & \Sigma y_{2t} \\ \Sigma y_{2t} & \Sigma y_{2t}^2 \end{bmatrix}^{-1} \left\{\begin{bmatrix} \Sigma y_{1t} \\ \Sigma y_{2t}y_{1t} \end{bmatrix} - \gamma_0\begin{bmatrix} \Sigma y_{2t} \\ \Sigma y_{2t}^2 \end{bmatrix}\right\} \\ &= \begin{bmatrix} T & \Sigma y_{2t} \\ \Sigma y_{2t} & \Sigma y_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma(y_{1t} - \gamma_0 y_{2t}) \\ \Sigma y_{2t}(y_{1t} - \gamma_0 y_{2t}) \end{bmatrix} \end{aligned}$$

and

$$\begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{1/2} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_T \\ \hat{\gamma}_T - \gamma_0 \end{bmatrix} = \begin{bmatrix} T^{-1/2} & 0 \\ 0 & T^{1/2} \end{bmatrix} \begin{bmatrix} T & \Sigma y_{2t} \\ \Sigma y_{2t} & \Sigma y_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2} & 0 \\ 0 & T^{-5/2} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} T^{-3/2} & 0 \\ 0 & T^{-5/2} \end{bmatrix} \begin{bmatrix} \Sigma(y_{1t} - \gamma_0 y_{2t}) \\ \Sigma y_{2t}(y_{1t} - \gamma_0 y_{2t}) \end{bmatrix}$$

$$= \left\{ \begin{bmatrix} T^{-3/2} & 0 \\ 0 & T^{-5/2} \end{bmatrix} \begin{bmatrix} T & \Sigma y_{2t} \\ \Sigma y_{2t} & \Sigma y_{2t}^2 \end{bmatrix} \begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{-1/2} \end{bmatrix} \right\}^{-1}$$

$$\times \begin{bmatrix} T^{-3/2} & 0 \\ 0 & T^{-5/2} \end{bmatrix} \begin{bmatrix} \Sigma(y_{1t} - \gamma_0 y_{2t}) \\ \Sigma y_{2t}(y_{1t} - \gamma_0 y_{2t}) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & T^{-2}\Sigma y_{2t} \\ T^{-2}\Sigma y_{2t} & T^{-3}\Sigma y_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-3/2}\Sigma(y_{1t} - \gamma_0 y_{2t}) \\ T^{-5/2}\Sigma y_{2t}(y_{1t} - \gamma_0 y_{2t}) \end{bmatrix}.$$

But

$$\Sigma y_{2t} = \underbrace{T y_{2,0}}_{O_p(T)} + \underbrace{\delta_2 \cdot \Sigma t}_{O_p(T^2)} + \underbrace{\Sigma \xi_{2t}}_{O_p(T^{3/2})},$$

and thus $T^{-2}\Sigma y_{2t} \xrightarrow{p} T^{-2}\delta_2 \cdot \Sigma t \to \delta_2/2$. Similarly, $T^{-3}\Sigma y_{2t}^2 \xrightarrow{p} T^{-3}\delta_2^2 \cdot \Sigma t^2 \to \delta_2^2/3$. Furthermore,

$$\Sigma(y_{1t} - \gamma_0 y_{2t}) = \underbrace{T(y_{1,0} - \gamma_0 y_{2,0})}_{O_p(T)} + \underbrace{\Sigma(\xi_{1t} - \gamma_0 \xi_{2t})}_{O_p(T^{3/2})},$$

establishing that $T^{-3/2}\Sigma(y_{1t} - \gamma_0 y_{2t}) \xrightarrow{p} T^{-3/2}\Sigma(\xi_{1t} - \gamma_0 \xi_{2t})$. Similarly,

$$\Sigma y_{2t}(y_{1t} - \gamma_0 y_{2t}) = \Sigma(y_{2,0} + \delta_2 t + \xi_{2t})(y_{1,0} + \xi_{1t} - \gamma_0 y_{2,0} - \gamma_0 \xi_{2t})$$

and $T^{-5/2}\Sigma y_{2t}(y_{1t} - \gamma_0 y_{2t}) \xrightarrow{p} T^{-5/2}\Sigma \delta_2 t(\xi_{1t} - \gamma_0 \xi_{2t})$.

(b) $$\begin{aligned} \Delta \hat{u}_t &= (y_{1t} - \hat{\alpha}_T - \hat{\gamma}_T y_{2t}) - (y_{1,t-1} - \hat{\alpha}_T - \hat{\gamma}_T y_{2,t-1}) \\ &= \Delta y_{1t} - \hat{\gamma}_T \Delta y_{2t} \\ &\xrightarrow{p} \Delta y_{1t} - \gamma_0 \Delta y_{2t}, \end{aligned}$$

since $\hat{\gamma}_T \xrightarrow{p} \gamma_0$.

19.2. Proposition 18.1 is used to show that

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \\ T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} T^{-1}\Sigma \mathbf{w}_t \mathbf{w}_t' & T^{-1}\Sigma \mathbf{w}_t & T^{-3/2}\Sigma \mathbf{w}_t \mathbf{y}_{2t}' \\ T^{-1}\Sigma \mathbf{w}_t' & 1 & T^{-3/2}\Sigma \mathbf{y}_{2t}' \\ T^{-3/2}\Sigma \mathbf{y}_{2t}\mathbf{w}_t' & T^{-3/2}\Sigma \mathbf{y}_{2t} & T^{-2}\Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma \mathbf{w}_t \tilde{z}_t \\ T^{-1/2}\Sigma \tilde{z}_t \\ T^{-1}\Sigma \mathbf{y}_{2t}\tilde{z}_t \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} \mathbf{Q} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \left\{ \int [\mathbf{W}_2(r)]' \, dr \right\} \tilde{\boldsymbol{\Lambda}}_{22}' \\ \mathbf{0} & \tilde{\boldsymbol{\Lambda}}_{22} \int \mathbf{W}_2(r) \, dr & \tilde{\boldsymbol{\Lambda}}_{22} \left\{ \int [\mathbf{W}_2(r)] \cdot [\mathbf{W}_2(r)]' \, dr \right\} \tilde{\boldsymbol{\Lambda}}_{22}' \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \mathbf{h}_1 \\ \tilde{\lambda}_{11} \cdot W_1(1) \\ \tilde{\boldsymbol{\Lambda}}_{22} \cdot \left\{ \int [\mathbf{W}_2(r)] \, dW_1(r) \right\} \cdot \tilde{\lambda}_{11} \end{bmatrix},$$

as claimed.

19.3. Notice as in [19.3.13] that under the null hypothesis,

$$\chi_T^2 = \{\mathbf{R}_\gamma \cdot T(\hat{\boldsymbol\gamma}_T - \boldsymbol\gamma)\}'\left\{s_T^2[\mathbf{0}\quad 0\quad \mathbf{R}_\gamma]\right.$$

$$\times \left. \begin{bmatrix} T^{-1}\Sigma\mathbf{w}_t\mathbf{w}_t' & T^{-1}\Sigma\mathbf{w}_t & T^{-3/2}\Sigma\mathbf{w}_t\mathbf{y}_{2t}' \\ T^{-1}\Sigma\mathbf{w}_t' & 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t}\mathbf{w}_t' & T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1} \{\mathbf{R}_\gamma \cdot T(\hat{\boldsymbol\gamma}_T - \boldsymbol\gamma)\}$$

$$\xrightarrow{p} [\mathbf{R}_\gamma \lambda_{11}^* \boldsymbol\nu_2]'\left\{s_T^2[\mathbf{0}\quad 0\quad \mathbf{R}_\gamma]\right.$$

$$\times \left. \begin{bmatrix} \mathbf{Q} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\boldsymbol\Lambda_{22}^{*\prime} \\ \mathbf{0} & \boldsymbol\Lambda_{22}^*\int \mathbf{W}_2(r)\,dr & \boldsymbol\Lambda_{22}^*\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\boldsymbol\Lambda_{22}^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1} [\mathbf{R}_\gamma \lambda_{11}^* \boldsymbol\nu_2],$$

from which [19.3.25] follows immediately.

19.4.

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol\beta}_T - \boldsymbol\beta) \\ T^{1/2}(\hat\alpha_T - \alpha) \\ T(\hat{\boldsymbol\gamma}_T - \boldsymbol\gamma) \\ T^{3/2}(\hat\delta_T - \delta) \end{bmatrix} = \begin{bmatrix} T^{-1}\Sigma\mathbf{w}_t\mathbf{w}_t' & T^{-1}\Sigma\mathbf{w}_t & T^{-3/2}\Sigma\mathbf{w}_t\mathbf{y}_{2t}' & T^{-2}\Sigma\mathbf{w}_t t \\ T^{-1}\Sigma\mathbf{w}_t' & 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' & T^{-2}\Sigma t \\ T^{-3/2}\Sigma\mathbf{y}_{2t}\mathbf{w}_t' & T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' & T^{-5/2}\Sigma\mathbf{y}_{2t}t \\ T^{-2}\Sigma t\mathbf{w}_t' & T^{-2}\Sigma t & T^{-5/2}\Sigma t\mathbf{y}_{2t}' & T^{-3}\Sigma t^2 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} T^{-1/2}\Sigma\mathbf{w}_t u_t \\ T^{-1/2}\Sigma u_t \\ T^{-1}\Sigma\mathbf{y}_{2t}u_t \\ T^{-3/2}\Sigma t u_t \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} \mathbf{Q} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\boldsymbol\Lambda_{22}^{*\prime} & 1/2 \\ \mathbf{0} & \boldsymbol\Lambda_{22}^*\int \mathbf{W}_2(r)\,dr & \boldsymbol\Lambda_{22}^*\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\boldsymbol\Lambda_{22}^{*\prime} & \boldsymbol\Lambda_{22}^*\int r\mathbf{W}_2(r)\,dr \\ \mathbf{0}' & 1/2 & \left\{\int r[\mathbf{W}_2(r)]'\,dr\right\}\boldsymbol\Lambda_{22}^{*\prime} & 1/3 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \mathbf{h}_1 \\ \lambda_{11}^* W_1(1) \\ \boldsymbol\Lambda_{22}^*\left\{\int [\mathbf{W}_2(r)]\,dW_1(r)\right\}\lambda_{11}^* \\ \lambda_{11}^*\left\{W_1(1) - \int W_1(r)\,dr\right\} \end{bmatrix},$$

as claimed.

Chapter 20. Full-Information Maximum Likelihood Analysis of Cointegrated Systems

20.1. Form the Lagrangean

$$\mathbf{k}_1'\boldsymbol\Sigma_{\mathbf{YX}}\mathbf{a}_1 + \mu_k(1 - \mathbf{k}_1'\boldsymbol\Sigma_{\mathbf{YY}}\mathbf{k}_1) + \mu_a(1 - \mathbf{a}_1'\boldsymbol\Sigma_{\mathbf{XX}}\mathbf{a}_1),$$

with μ_k and μ_a Lagrange multipliers. First-order conditions are

(a) $\Sigma_{YX}\mathbf{a}_1 = 2\mu_k\Sigma_{YY}\mathbf{k}_1$

(b) $\Sigma_{XY}\mathbf{k}_1 = 2\mu_a\Sigma_{XX}\mathbf{a}_1$.

Premultiply (a) by $\mathbf{k}_1'$ and (b) by $\mathbf{a}_1'$ to deduce that

$$2\mu_k = 2\mu_a \equiv r_1.$$

Next, premultiply (a) by $r_1^{-1}\Sigma_{YY}^{-1}$ and substitute the result into (b):

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a}_1 = r_1^2\Sigma_{XX}\mathbf{a}_1$$

or

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a}_1 = r_1^2\mathbf{a}_1.$$

Thus, r_1^2 is an eigenvalue of $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$ with $\mathbf{a}_1$ the associated eigenvector, as claimed.

Similarly, premultiplying (b) by $r_1^{-1}\Sigma_{XX}^{-1}$ and substituting the result into (a) reveals that

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_1 = r_1^2\mathbf{k}_1.$$

20.2. The restriction when $h = 0$ is that $\boldsymbol{\zeta}_0 = \mathbf{0}$. In this case, [20.3.2] would be

$$\mathcal{L}_0^* = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\Sigma}_{UU}|,$$

where $\hat{\Sigma}_{UU}$ is the variance-covariance matrix for the residuals of [20.2.4]. This will be recognized from expression [11.1.32] as the maximum value attained for the log likelihood for the model

$$\Delta\mathbf{y}_t = \boldsymbol{\pi}_0 + \mathbf{\Pi}_1\Delta\mathbf{y}_{t-1} + \mathbf{\Pi}_2\Delta\mathbf{y}_{t-2} + \cdots + \mathbf{\Pi}_{p-1}\Delta\mathbf{y}_{t-p+1} + \mathbf{u}_t,$$

as claimed.

20.3. The residuals $\hat{\mathbf{g}}_t$ are the same as the residuals from an unrestricted regression of $\hat{\mathbf{u}}_t$ on $\hat{\mathbf{v}}_t$. The MSE matrix for the latter regression is given by $\hat{\Sigma}_{UU} - \hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}$. Thus,

$$\begin{aligned}|\hat{\Sigma}_{GG}| &= |\hat{\Sigma}_{UU} - \hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}| \\ &= |\hat{\Sigma}_{UU}|\cdot|\mathbf{I}_n - \hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}| \\ &= |\hat{\Sigma}_{UU}|\cdot\prod_{i=1}^{n}\theta_i,\end{aligned}$$

where θ_i denotes the ith eigenvalue of $\mathbf{I}_n - \hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}$. Recalling that λ_i is an eigenvalue of $\hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}$ associated with the eigenvector $\mathbf{k}_i$, we have that

$$\left[\mathbf{I}_n - \hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}\right]\mathbf{k}_i = (1 - \lambda_i)\mathbf{k}_i,$$

so that $\theta_i = (1 - \lambda_i)$ is an eigenvalue of $\mathbf{I}_n - \hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}$ and

$$|\hat{\Sigma}_{GG}| = |\hat{\Sigma}_{UU}|\cdot\prod_{i=1}^{n}(1 - \lambda_i).$$

Hence, the two expressions are equivalent.

20.4. Here, $\hat{\lambda}_1$ is the scalar

$$\hat{\lambda}_1 = \hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}\hat{\Sigma}_{UU}^{-1}\hat{\Sigma}_{UV}$$

and the test statistic is

$$-T\log(1 - \hat{\lambda}_1) = -T\log[(\hat{\Sigma}_{UU}^{-1})\cdot(\hat{\Sigma}_{UU} - \hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU})].$$

But $\hat{u}_t$ is the residual from a regression of Δy_t on a constant and $\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1}$, meaning that $\hat{\Sigma}_{UU} = \hat{\sigma}_0^2$. Likewise, $\hat{v}_t$ is the residual from a regression of y_{t-1} on $\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1}$. The residual from a regression of $\hat{u}_t$ on $\hat{v}_t$, whose average squared value is given by $(\hat{\Sigma}_{UU} - \hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU})$, is the same as the residual from a regression of y_t on a constant, y_{t-1}, and $\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1}$, whose average squared value is denoted $\hat{\sigma}_1^2$:

$$(\hat{\Sigma}_{UU} - \hat{\Sigma}_{UV}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{VU}) = \hat{\sigma}_1^2.$$

Hence, the test statistic is equivalent to $T[\log(\hat{\sigma}_0^2) - \log(\hat{\sigma}_1^2)]$, as claimed.

Chapter 22. Modeling Time Series with Changes in Regime

22.1. $$\mathbf{PT} = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix} \times \begin{bmatrix} (1 - p_{22})/(2 - p_{11} - p_{22}) & -1 \\ (1 - p_{11})/(2 - p_{11} - p_{22}) & 1 \end{bmatrix}$$

$$= \begin{bmatrix} (1 - p_{22}) \left(\dfrac{p_{11}}{2 - p_{11} - p_{22}} + \dfrac{1 - p_{11}}{2 - p_{11} - p_{22}} \right) & 1 - p_{11} - p_{22} \\ (1 - p_{11}) \left(\dfrac{1 - p_{22}}{2 - p_{11} - p_{22}} + \dfrac{p_{22}}{2 - p_{11} - p_{22}} \right) & -1 + p_{11} + p_{22} \end{bmatrix}$$

$$= \begin{bmatrix} (1 - p_{22})/(2 - p_{11} - p_{22}) & -\lambda_2 \\ (1 - p_{11})/(2 - p_{11} - p_{22}) & \lambda_2 \end{bmatrix}$$

$$= \mathbf{T\Lambda}.$$

D
Greek Letters and Mathematical Symbols Used in the Text

Greek letters and common interpretation

alpha	α	population linear projection coefficient (page 74)
beta	$\boldsymbol{\beta}$	population regression coefficient (page 200)
gamma	$\boldsymbol{\Gamma}$	autocovariance matrix for vector process (page 261)
	γ	autocovariance for scalar process (page 45)
delta	Δ	change in value of variable (page 436)
	δ	small number; coefficient on time trend (page 435)
epsilon	ε	a white noise variable (page 47)
zeta	ζ	constant term in $ARCH$ specification (page 658)
eta	η	$AR(\infty)$ coefficient (page 79)
theta	$\boldsymbol{\Theta}$	matrix of moving average coefficients (page 262)
	$\boldsymbol{\theta}$	vector of population parameters (page 747)
	θ	scalar $MA(q)$ coefficient (page 50)
kappa	κ	kernel (page 165)
lambda	$\boldsymbol{\Lambda}$	matrix of eigenvalues (page 730)
	λ	individual eigenvalue (page 729) Lagrange multiplier (page 135)
mu	μ	population mean (page 739)
nu	ν	degrees of freedom (page 409)
xi	$\boldsymbol{\Xi}$	matrix of derivatives (page 339)
	$\boldsymbol{\xi}$	state vector (pages 7 and 372)
pi	Π	product (page 747)
	π	the number 3.14159....
rho	ρ	autocorrelation (page 49) autoregressive coefficient (page 517)
sigma	$\sum$	summation
	$\boldsymbol{\Sigma}$	long-run variance-covariance matrix (page 614)
	σ	population standard deviation (page 745)
tau	τ	time index
upsilon	$\boldsymbol{\Upsilon}$	scaling matrix to calculate asymptotic distributions (page 457)
phi	$\boldsymbol{\Phi}$	matrix of autoregressive coefficients (page 257)
	ϕ	scalar autoregressive coefficient (page 53)
chi	χ	a variable with a chi-square distribution (page 746)
psi	$\boldsymbol{\Psi}$	matrix of moving average coefficients for vector $MA(\infty)$ process (page 262)
	ψ	moving average coefficient for scalar $MA(\infty)$ process (page 52)
omega	$\boldsymbol{\Omega}$	variance-covariance matrix (page 748)
	ω	frequency (page 153)

Common uses of other letters

a	number of elements of unknown parameter vector (page 135)
$\mathbf{b}$ or $\mathbf{b}_T$	estimated *OLS* regression coefficient based on sample of size T (page 75)
$\mathbf{c}$	vector of constant terms for vector autoregression (page 257)
c	constant term in univariate autoregression (page 53)
$\mathbf{e}_i$	the ith column of the identity matrix
e	the base for the natural logarithms (page 715)
$\mathbf{I}_n$	the $(n \times n)$ identity matrix (page 722)
i	the square root of negative one (page 708)
J	value of Lagrangean (page 135)
k	the number of explanatory variables in a regression
L	the lag operator (page 26)
$\mathcal{L}$	value of log likelihood function (page 747)
n	number of variables observed at date t in a vector system (page 257)
$O_p(T)$	order T in probability (page 460)
$\mathbf{P}_{t+1\|t}$	*MSE* matrix for inference about state vector (page 378)
p	the order of an autoregressive process (page 58)
$\mathbf{Q}$	limiting value of $(\mathbf{X}'\mathbf{X}/T)$ for $\mathbf{X}$ the $(T \times k)$ matrix of explanatory variables for an *OLS* regression (page 208); variance-covariance matrix of disturbances in state equation (page 373)
q	the order of a moving average process (page 50); number of autocovariances used in Newey-West estimate (page 281)
$\mathbf{R}$	variance-covariance matrix of disturbances in observation equation (page 373)
$\mathbb{R}^n$	the set consisting of all real n-dimensional vectors (page 737)
r	number of variables in state equation (page 372); index of date for a continuous-time process
s^2 or s_T^2	unbiased estimate of residual variance for an *OLS* regression with sample of size T (page 203)
s_t	state at date t for a Markov chain
T	the number of dates included in a sample
$\mathbf{X}$	$(T \times k)$ matrix of explanatory variables for an *OLS* regression (page 201)
$\mathcal{Y}_t$	history of observations through date t (page 143)
z	argument of autocovariance generating function (page 61)

Mathematical symbols

$\aleph$	aleph (first letter of the Hebrew alphabet), used for matrix of regression coefficients (page 636)
$\exp(x)$	the number e (the base for natural logarithms) raised to the x power (page 715)
$\log(x)$	natural logarithm of x (page 717)
$x!$	x factorial (page 713)
$[x(L)]_+$	annihilation operator (page 78)
$[x]^*$	greatest integer less than or equal to x
$\|x\|$	absolute value of a real scalar or modulus of a complex scalar x (page 709)
$\|\mathbf{X}\|$	determinant of a square matrix $\mathbf{X}$ (page 724)
$\mathbf{X}'$	transpose of the matrix $\mathbf{X}$ (page 723)
$\mathbf{0}_{nm}$	an $(n \times m)$ matrix of zeros
$\mathbf{0}'$	a $(1 \times n)$ row vector of zeros
∇	gradient vector (page 735)
$\otimes$	Kronecker product (page 732)

$\odot$	element-by-element multiplication (page 692)
$y \cong x$	y is approximately equal to x
$y \equiv x$	y is defined to be the value represented by x
$\max\{y,x\}$	the value given by the larger of y or x
$y = \sup_{r\in[0,1]} f(r)$	y is the smallest number such that $y \geq f(r)$ for all r in $[0,1]$ (page 481)
$x \in A$	x is an element of A
$A \subset B$	A is a subset of B (page 189)
$P\{A\}$	probability that event A occurs (page 739)
$f_Y(y)$	probability density of the random variable Y (page 739)
$Y \sim N(\mu,\sigma^2)$	Y has a $N(\mu,\sigma^2)$ distribution (page 745)
$Y \approx N(\mu,\sigma^2$	Y has a distribution that is approximately $N(\mu,\sigma^2)$ (page 210)
$E(X)$	expectation of X (page 740)
$\mathrm{Var}(X)$	variance of X (page 740)
$\mathrm{Cov}(X,Y)$	covariance between X and Y (page 742)
$\mathrm{Corr}(X,Y)$	correlation between X and Y (page 743)
$Y\|X$	Y conditional on X (page 741)
$\hat{P}(Y\|X)$	linear projection of Y on X (pages 74–75)
$\hat{E}(Y\|X)$	linear projection of Y on X and a constant (pages 74–75)
$\hat{y}_{t+s\|t}$	linear projection of y_{t+s} on a constant and a set of variables observed at date t (page 74)
$x_T \to y$	$\lim_{T\to\infty} x_T = y$ (page 180)
$x_T \xrightarrow{p} y$	x_T converges in probability to y (pages 181, 749)
$x_T \xrightarrow{m.s.} y$	x_T converges in mean square to y (pages 182, 749)
$x_T \xrightarrow{L} y$	x_T converges in distribution to y (page 184)
$x_T(\cdot) \xrightarrow{p} x(\cdot)$	the sequence of functions whose value at r is $x_T(r)$ converges in probability to the function whose value at r is $x(r)$ (page 481)
$x_T(\cdot) \xrightarrow{L} x(\cdot)$	the sequence of functions whose value at r is $x_T(r)$ converges in probability law to the function whose value at r is $x(r)$ (page 481)

Author Index

Subject Index

C

D

H

I

J

K

L

M

S

T

U

V

W

Y