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:

$$y_{1t} = \gamma y_{2t} + u_{1t} \tag{19.1.1}$$

$$y_{2t} = y_{2,t-1} + u_{2t}, \tag{19.1.2}$$

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

$$\Delta y_{2t} = u_{2t}, \tag{19.1.3}$$

while differencing [19.1.1] results in

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

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

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

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

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

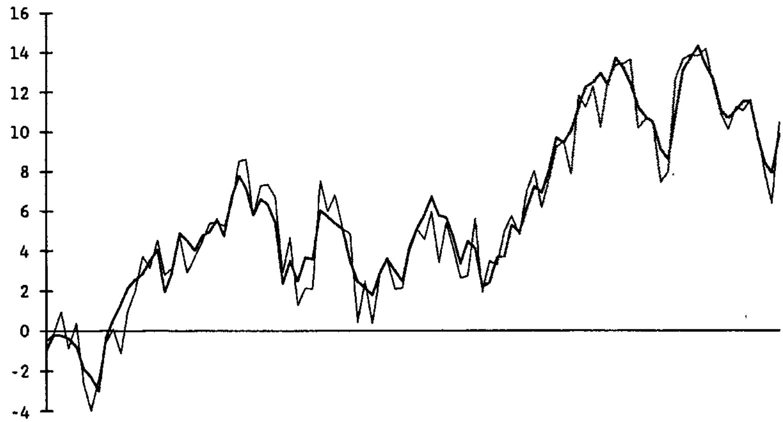

FIGURE 19.1 Sample realization of cointegrated series.

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 $(\Delta y_{1t}, \Delta y_{2t})'$,

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

where

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

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

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

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

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

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 $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and

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

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

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

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

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

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

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

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

$$\mathbf{A}'\mathbf{y}_t = \mathbf{A}'(\mathbf{y}_0 - \boldsymbol{\eta}_0) + \mathbf{A}'\boldsymbol{\delta}\cdot t + \mathbf{A}'\boldsymbol{\Psi}(1)\cdot(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t) + \mathbf{A}'\boldsymbol{\eta}_t. \qquad [19.1.14]$$

If $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$ is nonsingular, then $\mathbf{c}'(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t)$ 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 $\mathbf{A}'\boldsymbol{\Psi}(1) = \mathbf{0}$. 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

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

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 $\boldsymbol{\Psi}(1)$, such as $\mathbf{a}_1'\boldsymbol{\Psi}(1)$, are zero, meaning that the determinant $|\boldsymbol{\Psi}(z)| = 0$ at $z = 1$. This in turn means that the matrix operator $\boldsymbol{\Psi}(L)$ 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

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

and

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

This is a singular matrix with $\mathbf{A}'\boldsymbol{\Psi}(1) = \mathbf{0}$ 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

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

$$= [\mathbf{I}_h \quad -\boldsymbol{\Gamma}'],$$

where $\boldsymbol{\Gamma}'$ 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

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

or

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

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 $\mathbf{b}$ such that

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

or

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

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

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

where

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

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

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

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

where

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

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

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

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

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

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

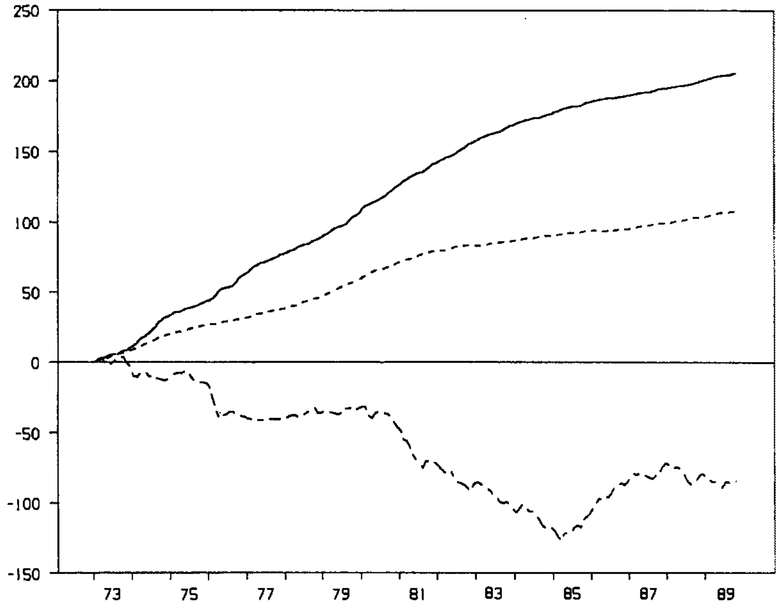

FIGURE 19.2 One hundred times the log of the price level in the United States (p_t), the dollar-lira exchange rate (s_t), and the price level in Italy (p_t^*), monthly, 1973–89. Key: ---- p_t; ——— s_t; ——— p_t^*.

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

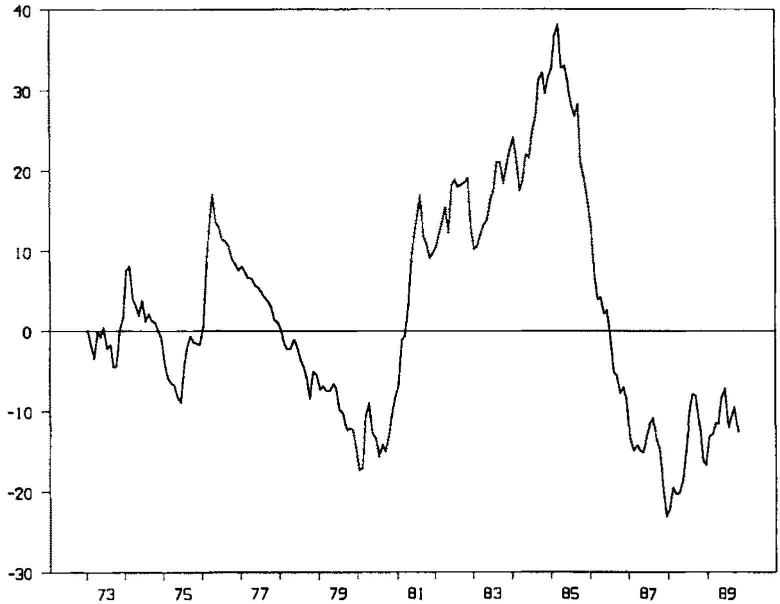

FIGURE 19.3 The real dollar-lira exchange rate, monthly, 1973–89.

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,

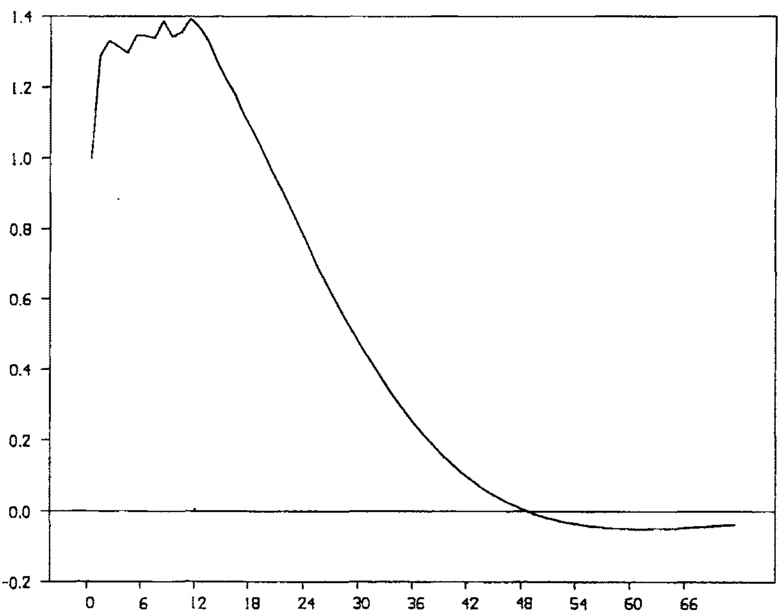

FIGURE 19.4 Impulse-response function for the real dollar-lira exchange rate. Graph shows $\psi_j = \partial(p_{t+j} - s_{t+j} - p^*_{t+j})/\varepsilon_t$ as a function of j.

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 $\mathbf{z}_t^* \equiv (\mathbf{z}_{1t}^*, \mathbf{z}_{2t}^*, \ldots, \mathbf{z}_{ht}^*)'$ 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

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

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]:

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

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]:

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

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

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

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:

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

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*

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

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*

$$Z_{\rho,T} \xrightarrow{L} Z_n, \qquad [19.2.37]$$

where

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

(*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*

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

(*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,

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

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:

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

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:

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

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:

$$y_{1t} = \alpha + \gamma y_{2t} + u_t. \tag{19.2.43}$$

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,

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

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:

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

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

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

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

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

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

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

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 $\{s \cdot \mathbf{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable. The matrix $\mathbf{\Psi}(1)$ 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,

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

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:

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

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

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

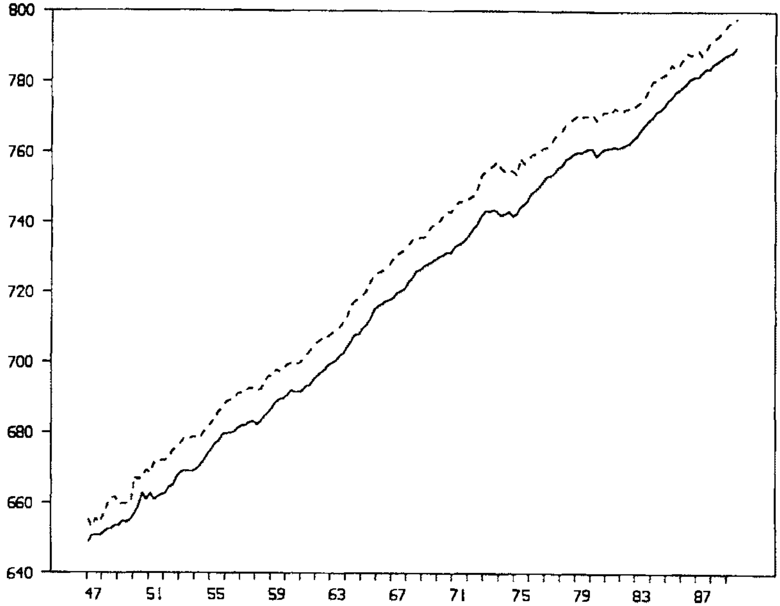

FIGURE 19.5 One hundred times the log of personal consumption expenditures (c_t) and personal disposable income (y_t) for the United States in billions of 1982 dollars, quarterly, 1947–89. Key: ——— c_t; ---- y_t.

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

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

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

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

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

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

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

$$\begin{aligned} &\left[\begin{matrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{matrix} \middle| (\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2T}) \right] \\ &\quad \sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \sigma_1^2 \left\{ \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T \cdot \mathbf{I}_g \end{bmatrix} \begin{bmatrix} T & \Sigma \mathbf{y}_{2t}' \\ \Sigma \mathbf{y}_{2t} & \Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{1/2} & \mathbf{0}' \\ \mathbf{0} & T \cdot \mathbf{I}_g \end{bmatrix} \right\} \right) \qquad [19.3.7] \\ &\quad = N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \sigma_1^2 \begin{bmatrix} 1 & T^{-3/2}\Sigma \mathbf{y}_{2t}' \\ T^{-3/2}\Sigma \mathbf{y}_{2t} & T^{-2}\Sigma \mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \right). \end{aligned}$$

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

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

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

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

$$\left[\begin{array}{c} \nu_1 \\ \boldsymbol{\nu}_2 \end{array}\middle| \mathbf{W}_2(\cdot)\right] \sim N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\, dr\right\}\mathbf{P}_{22}' \\ \mathbf{P}_{22}\int \mathbf{W}_2(r)\, dr & \mathbf{P}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\, dr\right\}\mathbf{P}_{22}' \end{bmatrix}^{-1}\right). \qquad [19.3.12]$$

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

$$\begin{aligned}
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} \\
&\quad \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. \\
&\quad \times \left.\left.\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. \\
&\quad \times \left.\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]
\end{aligned}$$

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

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

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

with $\{s\cdot\boldsymbol{\Psi}_s^*\}_{s=0}^{\infty}$ 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 $\boldsymbol{\Psi}^*(L)$ and $\mathbf{P}$ must be block-diagonal:

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

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

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

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

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 Λ_{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 Λ_{22}^*:

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

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

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

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],

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

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

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

$$\xrightarrow{p} (\tilde{\lambda}_{11}^2/s_T^2)[\mathbf{R}_\gamma\boldsymbol{\nu}_2]'\left\{[0 \quad \mathbf{R}_\gamma] \times \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol{\Lambda}}_{22}' \\ \tilde{\boldsymbol{\Lambda}}_{22}\int \mathbf{W}_2(r)\,dr & \tilde{\boldsymbol{\Lambda}}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\,dr\right\}\tilde{\boldsymbol{\Lambda}}_{22}' \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}_\gamma' \end{bmatrix}\right\}^{-1} [\mathbf{R}_\gamma\boldsymbol{\nu}_2]; \qquad [19.3.25]$$

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:

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

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

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

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:

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

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.

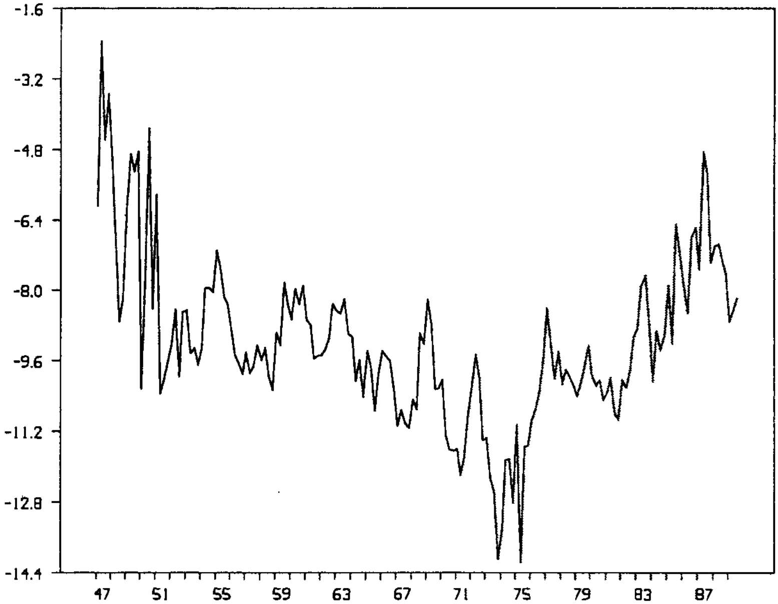

FIGURE 19.6 One hundred times the difference between the log of personal consumption expenditures (c_t) and the log of personal disposable income (y_t) for the United States, quarterly, 1947–89.

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

$$\underset{(h\times 1)}{\mathbf{y}_{1t}} = \underset{(h\times g)}{\Gamma'} \cdot \underset{(g\times 1)}{\mathbf{y}_{2t}} + \underset{(h\times 1)}{\boldsymbol{\mu}_1^*} + \underset{(h\times 1)}{\mathbf{z}_t^*} \qquad [19.3.32]$$

$$\underset{(g\times 1)}{\Delta\mathbf{y}_{2t}} = \underset{(g\times 1)}{\boldsymbol{\delta}_2} + \underset{(g\times 1)}{\mathbf{u}_{2t}} \qquad [19.3.33]$$

with

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

and $\boldsymbol{\Psi}^*(1)$ nonsingular. Here the generalization of the previous approach would be to augment [19.3.32] with leads and lags of $\Delta\mathbf{y}_{2t}$:

$$\mathbf{y}_{1t} = \boldsymbol{\mu}_1^* + \Gamma'\mathbf{y}_{2t} + \sum_{s=-p}^{p} \mathbf{B}_s'\Delta\mathbf{y}_{2,t-s} + \tilde{\mathbf{z}}_t, \qquad [19.3.34]$$

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

$$y_{1t} = \alpha + \boldsymbol{\gamma}'\mathbf{y}_{2t} + z_t^* \qquad [19.3.36]$$

$$\Delta\mathbf{y}_{2t} = \mathbf{u}_{2t} \qquad [19.3.37]$$

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

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{P}\mathbf{P}',$$

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

$$\boldsymbol{\Lambda}^* \equiv \boldsymbol{\Psi}^*(1)\cdot\mathbf{P}$$

$$\underset{(n\times n)}{\boldsymbol{\Sigma}^*} \equiv \boldsymbol{\Lambda}^*\cdot[\boldsymbol{\Lambda}^*]' \equiv \begin{bmatrix} \underset{(1\times 1)}{\Sigma_{11}^*} & \underset{(1\times g)}{\boldsymbol{\Sigma}_{21}^{*\prime}} \\ \underset{(g\times 1)}{\boldsymbol{\Sigma}_{21}^*} & \underset{(g\times g)}{\boldsymbol{\Sigma}_{22}^*} \end{bmatrix}, \qquad [19.3.38]$$

with $\boldsymbol{\Lambda}^*$ 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

$$\begin{aligned} \mathbf{G}(z) &\equiv \sum_{v=-\infty}^{\infty} z^v \begin{bmatrix} E(z_t^* z_{t-v}^*) & E(z_t^* \mathbf{u}_{2,t-v}') \\ E(\mathbf{u}_{2t} z_{t-v}^*) & E(\mathbf{u}_{2t}\mathbf{u}_{2,t-v}') \end{bmatrix} \\ &= [\boldsymbol{\Psi}^*(z)]\cdot\mathbf{P}\mathbf{P}'[\boldsymbol{\Psi}^*(z^{-1})]'. \end{aligned}$$

Thus, $\boldsymbol{\Sigma}^*$ could alternatively be described as the autocovariance-generating function $\mathbf{G}(z)$ evaluated at $z = 1$:

$$\begin{bmatrix} \Sigma_{11}^* & \boldsymbol{\Sigma}_{21}^{*\prime} \\ \boldsymbol{\Sigma}_{21}^* & \boldsymbol{\Sigma}_{22}^* \end{bmatrix} = \sum_{v=-\infty}^{\infty} \begin{bmatrix} E(z_t^* z_{t-v}^*) & E(z_t^* \mathbf{u}_{2,t-v}') \\ E(\mathbf{u}_{2t} z_{t-v}^*) & E(\mathbf{u}_{2t}\mathbf{u}_{2,t-v}') \end{bmatrix}. \qquad [19.3.39]$$

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 $\boldsymbol{\Sigma}_{21}^*$, 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

$$\boldsymbol{\aleph} \equiv \sum_{v=0}^{\infty} E(\mathbf{u}_{2t} z_{t+v}^*). \qquad [19.3.40]$$

The first issue can be addressed by subtracting $\boldsymbol{\Sigma}_{21}^{*\prime}(\boldsymbol{\Sigma}_{22}^*)^{-1}\Delta\mathbf{y}_{2t}$ from both sides of [19.3.36], arriving at

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

where

$$\begin{aligned} y_{1t}^\dagger &\equiv y_{1t} - \boldsymbol{\Sigma}_{21}^{*\prime}(\boldsymbol{\Sigma}_{22}^*)^{-1}\Delta\mathbf{y}_{2t} \\ z_t^\dagger &\equiv z_t^* - \boldsymbol{\Sigma}_{21}^{*\prime}(\boldsymbol{\Sigma}_{22}^*)^{-1}\Delta\mathbf{y}_{2t}. \end{aligned} \qquad [19.3.41]$$

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

$$\mathbf{L}' \equiv \begin{bmatrix} 1 & -\boldsymbol{\Sigma}_{21}^{*\prime}(\boldsymbol{\Sigma}_{22}^*)^{-1} \\ \mathbf{0} & \mathbf{I}_g \end{bmatrix} \equiv \begin{bmatrix} \underset{(1\times n)}{\boldsymbol{\ell}_1'} \\ \underset{(g\times n)}{\mathbf{L}_2'} \end{bmatrix}. \qquad [19.3.43]$$

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

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

for $\hat{y}_{1t}^{\dagger} \equiv y_{1t} - \hat{\boldsymbol{\Sigma}}_{21}^{*\prime}(\hat{\boldsymbol{\Sigma}}_{22}^{*})^{-1}\Delta \mathbf{y}_{2t}$. This analysis implies that

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

$$\xrightarrow{L} \sigma_1^{\dagger} \begin{bmatrix} \nu_1 \\ \boldsymbol{\nu}_2 \end{bmatrix},$$

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

$$(\hat{\sigma}_1^{\dagger})^2 = \hat{\Sigma}_{11}^{*} - \hat{\boldsymbol{\Sigma}}_{21}^{*\prime}(\hat{\boldsymbol{\Sigma}}_{22}^{*})^{-1}\hat{\boldsymbol{\Sigma}}_{21}^{*},$$

with $\hat{\boldsymbol{\Sigma}}_{ij}^{*}$ 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:

$$\begin{aligned} (s_T/\hat{\sigma}_1^{\dagger})^2\cdot\chi_T^2 &= \{\mathbf{R}\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \mathbf{r}\}'\left\{(\hat{\sigma}_1^{\dagger})^2[\mathbf{0}\quad \mathbf{R}]\begin{bmatrix} T & \Sigma \mathbf{y}'_{2t} \\ \Sigma \mathbf{y}_{2t} & \Sigma \mathbf{y}_{2t}\mathbf{y}'_{2t} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}' \end{bmatrix}\right\}^{-1}\{\mathbf{R}\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \mathbf{r}\} \\ &= \{\mathbf{R}\cdot T(\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \boldsymbol{\gamma})\}'\left\{(\hat{\sigma}_1^{\dagger})^2[\mathbf{0}\quad \mathbf{R}]\right. \\ &\quad \left.\times \begin{bmatrix} 1 & T^{-3/2}\Sigma \mathbf{y}'_{2t} \\ T^{-3/2}\Sigma \mathbf{y}_{2t} & T^{-2}\Sigma \mathbf{y}_{2t}\mathbf{y}'_{2t} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}' \end{bmatrix}\right\}^{-1}\{\mathbf{R}\cdot T(\hat{\boldsymbol{\gamma}}_T^{\dagger\dagger} - \boldsymbol{\gamma})\} \\ &\xrightarrow{L} (\sigma_1^{\dagger})^2(\mathbf{R}\boldsymbol{\nu}_2)'\left\{(\sigma_1^{\dagger})^2[\mathbf{0}\quad \mathbf{R}]\mathbf{H}^{-1}\begin{bmatrix} \mathbf{0}' \\ \mathbf{R}' \end{bmatrix}\right\}^{-1}(\mathbf{R}\boldsymbol{\nu}_2) \\ &\sim \chi^2(m). \end{aligned}$$

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

$$T^{-1}\sum_{t=1}^{T}\begin{bmatrix}\bar{y}_{1t} \\ \mathbf{y}_{2t}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}'] = T^{-1}\sum_{t=1}^{T}\begin{bmatrix}\bar{y}_{1,t-1} \\ \mathbf{y}_{2,t-1}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}'] + T^{-1}\sum_{t=1}^{T}\begin{bmatrix}z_t^* \\ \mathbf{u}_{2t}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}']$$
$$\xrightarrow{L} \Lambda^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\Lambda^{*\prime} + \sum_{v=0}^{\infty}\Gamma_v^{*\prime}. \qquad [19.A.2]$$

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

$$\begin{bmatrix}T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\gamma}_T - \gamma)\end{bmatrix} = \left\{\begin{bmatrix}T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & T^{-1}\cdot\mathbf{I}_g\end{bmatrix}\begin{bmatrix} T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}'\end{bmatrix}\right.$$
$$\left.\times\begin{bmatrix}T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & T^{-1}\cdot\mathbf{I}_g\end{bmatrix}\right\}^{-1}\left\{\begin{bmatrix}T^{-1/2} & \mathbf{0}' \\ \mathbf{0} & T^{-1}\cdot\mathbf{I}_g\end{bmatrix}\begin{bmatrix}\Sigma z_t^* \\ \Sigma\mathbf{y}_{2t}z_t^*\end{bmatrix}\right\} \qquad [19.A.6]$$
$$= \begin{bmatrix}1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}'\end{bmatrix}^{-1}\begin{bmatrix}T^{-1/2}\Sigma z_t^* \\ T^{-1}\Sigma\mathbf{y}_{2t}z_t^*\end{bmatrix}.$$

But from [19.A.2],

$$T^{-1}\Sigma\mathbf{y}_{2t}z_t^* = [\mathbf{0} \quad \mathbf{I}_g]T^{-1}\sum_{t=1}^{T}\begin{bmatrix}\bar{y}_{1t} \\ \mathbf{y}_{2t}\end{bmatrix}[z_t^* \quad \mathbf{u}_{2t}']\begin{bmatrix}1 \\ \mathbf{0}\end{bmatrix}$$
$$\xrightarrow{L} [\mathbf{0} \quad \mathbf{I}_g]\Lambda^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\Lambda^{*\prime}\begin{bmatrix}1 \\ \mathbf{0}\end{bmatrix}$$
$$+ [\mathbf{0} \quad \mathbf{I}_g]\sum_{v=0}^{\infty}\Gamma_v^{*\prime}\begin{bmatrix}1 \\ \mathbf{0}\end{bmatrix} \qquad [19.A.7]$$
$$= \Lambda_2^*\cdot\left\{\int_0^1 [\mathbf{W}(r)]\,[d\mathbf{W}(r)]'\right\}\cdot\boldsymbol{\lambda}_1^* + \sum_{v=0}^{\infty}E(\mathbf{u}_{2t}z_{t+v}^*).$$

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 $\mathbf{z}_{2t}^* \equiv (z_{2t}^*, z_{3t}^*, \ldots, z_{ht}^*)'$, 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 $\boldsymbol{\beta} \equiv (\beta_2, \beta_3, \ldots, \beta_h)'$ 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

$$\begin{aligned} \begin{bmatrix} \hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} \\ \hat{\alpha}_T^* \\ T^{1/2}\hat{\boldsymbol{\aleph}}_T^* \end{bmatrix} &= \begin{bmatrix} \mathbf{I}_{h-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & T^{1/2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} \Sigma\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime} & \Sigma\mathbf{z}_{2t}^* & \Sigma\mathbf{z}_{2t}^*\mathbf{y}_{2t}' \\ \Sigma\mathbf{z}_{2t}^{*\prime} & T & \Sigma\mathbf{y}_{2t}' \\ \Sigma\mathbf{y}_{2t}\mathbf{z}_{2t}^{*\prime} & \Sigma\mathbf{y}_{2t} & \Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} T\cdot\mathbf{I}_{h-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & T & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & T^{3/2}\mathbf{I}_g \end{bmatrix} \begin{bmatrix} T\cdot\mathbf{I}_{h-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & T & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & T^{3/2}\mathbf{I}_g \end{bmatrix}^{-1} \begin{bmatrix} \Sigma\mathbf{z}_{2t}^* u_t \\ \Sigma u_t \\ \Sigma\mathbf{y}_{2t} u_t \end{bmatrix} \\ &= \begin{bmatrix} T^{-1}\Sigma\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime} & T^{-1}\Sigma\mathbf{z}_{2t}^* & T^{-3/2}\Sigma\mathbf{z}_{2t}^*\mathbf{y}_{2t}' \\ T^{-1}\Sigma\mathbf{z}_{2t}^{*\prime} & 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t}\mathbf{z}_{2t}^{*\prime} & T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\Sigma\mathbf{z}_{2t}^* u_t \\ T^{-1}\Sigma u_t \\ T^{-3/2}\Sigma\mathbf{y}_{2t} u_t \end{bmatrix}. \end{aligned}$$

[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,

$$\begin{aligned} &\begin{bmatrix} T^{-1}\Sigma\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime} & T^{-1}\Sigma\mathbf{z}_{2t}^* & T^{-3/2}\Sigma\mathbf{z}_{2t}^*\mathbf{y}_{2t}' \\ T^{-1}\Sigma\mathbf{z}_{2t}^{*\prime} & 1 & T^{-3/2}\Sigma\mathbf{y}_{2t}' \\ T^{-3/2}\Sigma\mathbf{y}_{2t}\mathbf{z}_{2t}^{*\prime} & T^{-3/2}\Sigma\mathbf{y}_{2t} & T^{-2}\Sigma\mathbf{y}_{2t}\mathbf{y}_{2t}' \end{bmatrix} \\ &\quad \xrightarrow{L} \begin{bmatrix} E(\mathbf{z}_{2t}^*\mathbf{z}_{2t}^{*\prime}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 1 & \left\{\int [\mathbf{W}(r)]'\,dr\right\}\boldsymbol{\Lambda}_2^{*\prime} \\ \mathbf{0} & \boldsymbol{\Lambda}_2^*\int \mathbf{W}(r)\,dr & \boldsymbol{\Lambda}_2^*\left\{\int [\mathbf{W}(r)]\cdot[\mathbf{W}(r)]'\,dr\right\}\boldsymbol{\Lambda}_2^{*\prime} \end{bmatrix}, \end{aligned} \qquad [19.A.11]$$

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

$$\underset{(n\times n)}{\boldsymbol{\Lambda\Lambda}'} \equiv \begin{bmatrix} \underset{(1\times 1)}{\Sigma_{11}} & \underset{(1\times g)}{\boldsymbol{\Sigma}_{21}'} \\ \underset{(g\times 1)}{\boldsymbol{\Sigma}_{21}} & \underset{(g\times g)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}, \qquad [19.A.14]$$

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

$$\hat{u}_t^* = \sigma_1^* \cdot \{(y_{1t}^*/\sigma_1^*) - (1/\sigma_1^*) \cdot \hat{\gamma}_T^{*\prime} \mathbf{y}_{2t}^* - (\hat{\alpha}_T^*/\sigma_1^*)\} \\ \equiv \sigma_1^* \cdot \{[1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}_t^* - (\hat{\alpha}_T^*/\sigma_1^*)\} \tag{19.A.22}$$

for

$$\boldsymbol{\xi}_t^* \equiv \begin{bmatrix} y_{1t}^*/\sigma_1^* \\ \mathbf{y}_{2t}^* \end{bmatrix} = \mathbf{L}'\mathbf{y}_t. \tag{19.A.23}$$

Differencing [19.A.22] results in

$$(\hat{u}_t^* - \hat{u}_{t-1}^*) = \sigma_1^* \cdot [1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*]\Delta\boldsymbol{\xi}_t^*. \tag{19.A.24}$$

Using [19.A.22] and [19.A.24], the numerator of [19.A.21] can be written

$$\begin{aligned} &(T^*)^{-1} \sum_{t=2}^{T} \hat{u}_{t-1}^*(\hat{u}_t^* - \hat{u}_{t-1}^*) \\ &= (\sigma_1^*)^2 \cdot (T^*)^{-1} \sum_{t=2}^{T} \left\{[1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}_{t-1}^* - (\hat{\alpha}_T^*/\sigma_1^*)\right\}\left\{(\Delta\boldsymbol{\xi}_t^{*\prime})\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix}\right\} \\ &= (\sigma_1^*)^2 \cdot [1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix} \\ &\quad - (\sigma_1^*)^2 \cdot (T^*)^{-1/2}(\hat{\alpha}_T^*/\sigma_1^*) \cdot \left\{(T^*)^{-1/2} \sum_{t=2}^{T} (\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix}. \end{aligned} \tag{19.A.25}$$

Notice that the expression

$$[1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix}$$

is a scalar and accordingly equals its own transpose:

$$\begin{aligned} &[1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix} \\ &= (1/2)\left\{[1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix}\right. \\ &\quad \left. + [1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*] \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} (\Delta\boldsymbol{\xi}_t^*)(\boldsymbol{\xi}_{t-1}^{*\prime})\right\}\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix}\right\} \\ &= (1/2)\left\{[1 \quad -\hat{\gamma}_T^{*\prime}/\sigma_1^*]\left\{(T^*)^{-1} \sum_{t=2}^{T} \left(\boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime}) + (\Delta\boldsymbol{\xi}_t^*)(\boldsymbol{\xi}_{t-1}^{*\prime})\right)\right\}\begin{bmatrix} 1 \\ -\hat{\gamma}_T^*/\sigma_1^* \end{bmatrix}\right\}. \end{aligned} \tag{19.A.26}$$

But from result (d) of Proposition 18.1,

$$\begin{aligned} &(T^*)^{-1} \sum_{t=2}^{T} \left(\boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_t^{*\prime}) + (\Delta\boldsymbol{\xi}_t^*)(\boldsymbol{\xi}_{t-1}^{*\prime})\right) \\ &= \mathbf{L}' \cdot \left\{(T^*)^{-1} \sum_{t=2}^{T} \left(\mathbf{y}_{t-1}(\Delta\mathbf{y}_t') + (\Delta\mathbf{y}_t)(\mathbf{y}_{t-1}')\right)\right\} \cdot \mathbf{L} \qquad [19.A.27] \\ &\xrightarrow{L} \mathbf{L}' \cdot \{\boldsymbol{\Lambda} \cdot [\mathbf{W}(1)] \cdot [\mathbf{W}(1)]' \cdot \boldsymbol{\Lambda}' - E[(\Delta\mathbf{y}_t)(\Delta\mathbf{y}_t')]\} \cdot \mathbf{L} \\ &= [\mathbf{W}^*(1)] \cdot [\mathbf{W}^*(1)]' - E[(\Delta\boldsymbol{\xi}_t^*)(\Delta\boldsymbol{\xi}_t^{*\prime})] \end{aligned}$$

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

$$\begin{aligned}&(T^*)^{-1}\sum_{t=2}^{T}\hat{u}_{t-1}^*(\hat{u}_t^* - \hat{u}_{t-1}^*)\\ &\xrightarrow{L} (\sigma_1^*)^2\cdot\left\{\frac{1}{2}\left\{[1 \quad -\mathbf{h}_2']\cdot[\mathbf{W}^*(1)]\cdot[\mathbf{W}^*(1)]'\cdot\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}\right\} - h_1\cdot[\mathbf{W}^*(1)]'\cdot\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}\right.\\ &\qquad \left. - (1/2)\cdot[1 \quad -\mathbf{h}_2']\cdot\{E[(\Delta\boldsymbol{\xi}_t^*)(\Delta\boldsymbol{\xi}_t^{*\prime})]\}\cdot\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}\right\}.\end{aligned} \qquad [19.A.30]$$

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

$$\begin{aligned}&(T^*)^{-1}\sum_{t=j+2}^{T}(\hat{\rho}_T - 1)\hat{u}_{t-1}^*\Delta\hat{u}_{t-j}^*\\ &= (\sigma_1^*)^2\cdot(\hat{\rho}_T - 1)\cdot(T^*)^{-1}\sum_{t=j+2}^{T}\left\{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}_{t-1}^* - (\hat{\alpha}_T^*/\sigma_1^*)\right\}(\Delta\boldsymbol{\xi}_{t-j}^{*\prime})\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\\ &= \left\{(\sigma_1^*)^2\cdot[(T^*)^{1/2}(\hat{\rho}_T - 1)]\cdot[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\cdot(T^*)^{-3/2}\sum_{t=j+2}^{T}\boldsymbol{\xi}_{t-1}^*(\Delta\boldsymbol{\xi}_{t-j}^{*\prime})\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\right\}\\ &\quad - \left\{(\sigma_1^*)^2\cdot[(T^*)^{1/2}(\hat{\rho}_T - 1)]\cdot[(T^*)^{-1/2}(\hat{\alpha}_T^*/\sigma_1^*)](T^*)^{-1}\sum_{t=j+2}^{T}(\Delta\boldsymbol{\xi}_{t-j}^{*\prime})\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\right\}.\end{aligned} \qquad [19.A.33]$$

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,

$$\begin{aligned}
&(T^*)^{-1}\sum_{t=j+2}^{T}(\hat{\rho}_T - 1)^2\hat{u}^*_{t-1}\hat{u}^*_{t-j-1}\\
&= (\sigma_1^*)^2\cdot(T^*)^{-1}\sum_{t=j+2}^{T}(\hat{\rho}_T - 1)^2\left\{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}^*_{t-1} - (\hat{\alpha}_T^*/\sigma_1^*)\right\}\\
&\quad\times\left\{[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*]\boldsymbol{\xi}^*_{t-j-1} - (\hat{\alpha}_T^*/\sigma_1^*)\right\}\\
&= (\sigma_1^*)^2\cdot(T^*)^{-1}\sum_{t=j+2}^{T}(\hat{\rho}_T - 1)^2\begin{bmatrix}1 & -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^* & -(T^*)^{-1/2}\hat{\alpha}_T^*/\sigma_1^*\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}^*_{t-1}\\ (T^*)^{1/2}\end{bmatrix}\\
&\quad\times[\boldsymbol{\xi}^{*\prime}_{t-j-1} \quad (T^*)^{1/2}]\,[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^* \quad -(T^*)^{-1/2}\hat{\alpha}_T^*/\sigma_1^*]' \qquad [19.A.35]\\
&= (\sigma_1^*)^2\cdot[(T^*)^{1/2}(\hat{\rho}_T - 1)]^2\cdot[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^* \quad -(T^*)^{-1/2}\hat{\alpha}_T^*/\sigma_1^*]\\
&\quad\times\left\{(T^*)^{-2}\sum_{t=j+2}^{T}\begin{bmatrix}\boldsymbol{\xi}^*_{t-1}\boldsymbol{\xi}^{*\prime}_{t-j-1} & (T^*)^{1/2}\boldsymbol{\xi}^*_{t-1}\\ (T^*)^{1/2}\boldsymbol{\xi}^{*\prime}_{t-j-1} & T^*\end{bmatrix}\right\}\\
&\quad\times[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^* \quad -(T^*)^{-1/2}\hat{\alpha}_T^*/\sigma_1^*]'\\
&\xrightarrow{p} 0,
\end{aligned}$$

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

$$\begin{aligned}
\hat{c}_{j,T} &\xrightarrow{p} (T^*)^{-1}\sum_{t=j+2}^{T}(\Delta\hat{u}^*_t)\cdot(\Delta\hat{u}^*_{t-j})\\
&= (\sigma_1^*)^2\cdot[1 \quad -\hat{\boldsymbol{\gamma}}_T^{*\prime}/\sigma_1^*](T^*)^{-1}\sum_{t=j+2}^{T}(\Delta\boldsymbol{\xi}^*_t)\cdot(\Delta\boldsymbol{\xi}^{*\prime}_{t-j})\begin{bmatrix}1\\ -\hat{\boldsymbol{\gamma}}_T^*/\sigma_1^*\end{bmatrix}\\
&\xrightarrow{L} (\sigma_1^*)^2\cdot[1 \quad -\mathbf{h}_2']\cdot E\{(\Delta\boldsymbol{\xi}^*_t)\cdot(\Delta\boldsymbol{\xi}^{*\prime}_{t-j})\}\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix} \qquad [19.A.36]\\
&= (\sigma_1^*)^2\cdot[1 \quad -\mathbf{h}_2']\cdot\mathbf{L}'\cdot E\{(\Delta\mathbf{y}_t)\cdot(\Delta\mathbf{y}_{t-j}')\}\cdot\mathbf{L}\begin{bmatrix}1\\ -\mathbf{h}_2\end{bmatrix}.
\end{aligned}$$

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],

$$(T^*)^2 \cdot \hat{\sigma}^2_{\hat{\rho}_T} \div s_T^2 = \frac{1}{(T^*)^{-2} \sum_{t=2}^{T} \hat{u}^2_{t-1}} \xrightarrow{L} \frac{1}{(\sigma_1^*)^2 \cdot H_n}. \qquad [19.A.38]$$

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

$$(\hat{c}_{0,T}/s_T^2) = (T-2)/(T-1) \to 1,$$

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 $\Delta\mathbf{y}_{2t} = \mathbf{n}_{2t}$, where

$$\begin{bmatrix} u_t \\ \mathbf{u}_{2t} \end{bmatrix} = \tilde{\boldsymbol{\Psi}}(L)\boldsymbol{\varepsilon}_t = \begin{bmatrix} \tilde{\psi}_{11}(L) & \mathbf{0}' \\ \mathbf{0} & \tilde{\boldsymbol{\Psi}}_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix}$$

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 $\{s\cdot\tilde{\boldsymbol{\Psi}}_s\}_{s=0}^{\infty}$ is absolutely summable, $\tilde{\lambda}_{11} \equiv \sigma_1\cdot\tilde{\psi}_{11}(1) \neq 0$, and $\tilde{\boldsymbol{\Lambda}}_{22} \equiv \tilde{\boldsymbol{\Psi}}_{22}(1)\cdot\mathbf{P}_{22}$ is nonsingular. Show that the *OLS* estimates satisfy

$$\begin{bmatrix} T^{1/2}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \\ T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \mathbf{Q}^{-1}\mathbf{h}_1 \\ \tilde{\lambda}_{11}\cdot\nu_1 \\ \tilde{\lambda}_{11}\cdot\boldsymbol{\nu}_2 \\ \tilde{\lambda}_{11}\cdot\nu_3 \end{bmatrix},$$

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

$$\begin{bmatrix} \nu_1 \\ \boldsymbol{\nu}_2 \\ \nu_3 \end{bmatrix} \equiv \mathbf{H}^{-1} \begin{bmatrix} W_1(1) \\ \tilde{\boldsymbol{\Lambda}}_{22}\cdot\left\{\int [\mathbf{W}_2(r)]\, dW_1(r)\right\} \\ \left\{W_1(1) - \int W_1(r)\, dr\right\} \end{bmatrix}$$

$$\mathbf{H} \equiv \begin{bmatrix} 1 & \left\{\int [\mathbf{W}_2(r)]'\, dr\right\}\tilde{\boldsymbol{\Lambda}}'_{22} & 1/2 \\ \tilde{\boldsymbol{\Lambda}}_{22}\int \mathbf{W}_2(r)\, dr & \tilde{\boldsymbol{\Lambda}}_{22}\left\{\int [\mathbf{W}_2(r)]\cdot[\mathbf{W}_2(r)]'\, dr\right\}\tilde{\boldsymbol{\Lambda}}'_{22} & \tilde{\boldsymbol{\Lambda}}_{22}\int r\mathbf{W}_2(r)\, dr \\ 1/2 & \left\{\int r[\mathbf{W}_2(r)]'\, dr\right\}\tilde{\boldsymbol{\Lambda}}'_{22} & 1/3 \end{bmatrix}.$$

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 $(\tilde{\lambda}_{11}^2/s_T^2)$ times a $\chi^2(m)$ variable.

19.5. Consider the regression model

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