Davidson, James E. H., David F. Hendry, Frank Srba, and Stephen Yeo. 1978. "Econometric Modelling of the Aggregate Time-Series Relationship between Consumers' Expenditure and Income in the United Kingdom." *Economic Journal* 88:661–92.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55:251–76.

——— and Byung Sam Yoo. 1987. "Forecasting and Testing in Co-Integrated Systems." *Journal of Econometrics* 35:143–59.

Granger, C. W. J. 1983. "Co-Integrated Variables and Error-Correcting Models." Unpublished University of California, San Diego, Discussion Paper 83-13.

——— and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2:111–20.

Hansen, Bruce E. 1990. "A Powerful, Simple Test for Cointegration Using Cochrane-Orcutt." University of Rochester. Mimeo.

———. 1992. "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends." *Journal of Econometrics* 53:87–121.

Haug, Alfred A. 1992. "Critical Values for the $\hat{Z}_\alpha$-Phillips-Ouliaris Test for Cointegration." *Oxford Bulletin of Economics and Statistics* 54:473–80.

Johansen, Søren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12:231–54.

———. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59:1551–80.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. 1991. "Stochastic Trends and Economic Fluctuations." *American Economic Review* 81:819–40.

Kremers, Jeroen J. M. 1989. "U.S. Federal Indebtedness and the Conduct of Fiscal Policy." *Journal of Monetary Economics* 23:219–38.

Mosconi, Rocco, and Carlo Giannini. 1992. "Non-Causality in Cointegrated Systems: Representation, Estimation and Testing." *Oxford Bulletin of Economics and Statistics* 54:399–417.

Ogaki, Masao. 1992. "Engel's Law and Cointegration." *Journal of Political Economy* 100:1027–46.

——— and Joon Y. Park. 1992. "A Cointegration Approach to Estimating Preference Parameters." Department of Economics, University of Rochester. Mimeo.

Park, Joon Y. 1992. "Canonical Cointegrating Regressions." *Econometrica* 60:119–43.

——— and Masao Ogaki. 1991. "Inference in Cointegrated Models Using VAR Prewhitening to Estimate Shortrun Dynamics." University of Rochester. Mimeo.

———, S. Ouliaris, and B. Choi. 1988. "Spurious Regressions and Tests for Cointegration." Cornell University. Mimeo.

Phillips, Peter C. B. 1987. "Time Series Regression with a Unit Root." *Econometrica* 55:277–301.

———. 1991. "Optimal Inference in Cointegrated Systems." *Econometrica* 59:283–306.

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

——— and Bruce E. Hansen. 1990. "Statistical Inference in Instrumental Variables Regression with I(1) Processes." *Review of Economic Studies* 57:99–125.

——— and Mico Loretan. 1991. "Estimating Long-Run Economic Equilibria." *Review of Economic Studies* 58:407–36.

——— and S. Ouliaris. 1990. "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica* 58:165–93.

Saikkonen, Pentti. 1991. "Asymptotically Efficient Estimation of Cointegration Regressions." *Econometric Theory* 7:1–21.

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

Stock, James H. 1987. "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors." *Econometrica* 55:1035–56.

———. 1990. "A Class of Tests for Integration and Cointegration." Harvard University. Mimeo.

Stock, James H., and Mark W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* 83:1097–1107.

——— and ———. 1993. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica* 61:783–820.

Wooldridge, Jeffrey M. 1991. "Notes on Regression with Difference-Stationary Data." Michigan State University. Mimeo.

20 Full-Information Maximum Likelihood Analysis of Cointegrated Systems

An $(n \times 1)$ vector $\mathbf{y}_t$ was said to exhibit h cointegrating relations if there exist h linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ such that $\mathbf{a}_i'\mathbf{y}_t$ is stationary. If such vectors exist, their values are not uniquely defined, since any linear combinations of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ would also be described as cointegrating vectors. The approaches described in the previous chapter sidestepped this problem by imposing normalization conditions such as $a_{11} = 1$. For this normalization we would put y_{1t} on the left side of a regression and the other elements of $\mathbf{y}_t$ on the right side. We might equally well have normalized $a_{12} = 1$ instead, in which case y_{2t} would be the variable that belongs on the left side of the regression. The results obtained in practice can thus depend on an essentially arbitrary assumption. Furthermore, if the first variable does not appear in the cointegrating relation at all ($a_{11} = 0$), then setting $a_{11} = 1$ is not a harmless normalization but instead results in a fundamentally misspecified model.

For these reasons there is some value in using full-information maximum likelihood (*FIML*) to estimate the linear space spanned by the cointegrating vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$. This chapter describes the solution to this problem developed by Johansen (1988, 1991), whose work is closely related to that of Ahn and Reinsel (1990), and more distantly to that of Stock and Watson (1988). Another advantage of *FIML* is that it allows us to test for the number of cointegrating relations. The approach of Phillips and Ouliaris (1990) described in Chapter 19 tested the null hypothesis that there are no cointegrating relations. This chapter presents more general tests of the null hypothesis that there are h_0 cointegrating relations, where h_0 could be $0, 1, \ldots,$ or $n - 1$.

To develop these ideas, Section 20.1 begins with a discussion of canonical correlation analysis. Section 20.2 then develops the *FIML* estimates, while Section 20.3 describes hypothesis testing in cointegrated systems. Section 20.4 offers a brief overview of unit roots in time series analysis.

20.1. *Canonical Correlation*

Population Canonical Correlations

Let the $(n_1 \times 1)$ vector $\mathbf{y}_t$ and the $(n_2 \times 1)$ vector $\mathbf{x}_t$ denote stationary random variables. Typically, $\mathbf{y}_t$ and $\mathbf{x}_t$ are measured as deviations from their population means, so that $E(\mathbf{y}_t\mathbf{y}_t')$ represents the variance-covariance matrix of $\mathbf{y}_t$. In general, there might be complicated correlations among the elements of $\mathbf{y}_t$ and $\mathbf{x}_t$, summarized by the joint variance-covariance matrix

$$\begin{bmatrix} \underset{(n_1\times n_1)}{E(\mathbf{y}_t\mathbf{y}_t')} & \underset{(n_1\times n_2)}{E(\mathbf{y}_t\mathbf{x}_t')} \\ \underset{(n_2\times n_1)}{E(\mathbf{x}_t\mathbf{y}_t')} & \underset{(n_2\times n_2)}{E(\mathbf{x}_t\mathbf{x}_t')} \end{bmatrix} = \begin{bmatrix} \underset{(n_1\times n_1)}{\Sigma_{\mathbf{YY}}} & \underset{(n_1\times n_2)}{\Sigma_{\mathbf{YX}}} \\ \underset{(n_2\times n_1)}{\Sigma_{\mathbf{XY}}} & \underset{(n_2\times n_2)}{\Sigma_{\mathbf{XX}}} \end{bmatrix}.$$

We can often gain some insight into the nature of these correlations by defining two new $(n \times 1)$ random vectors, $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$, where n is the smaller of n_1 and n_2. These vectors are linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$, respectively:

$$\boldsymbol{\eta}_t \equiv \mathcal{K}'\mathbf{y}_t \tag{20.1.1}$$

$$\boldsymbol{\xi}_t \equiv \mathcal{A}'\mathbf{x}_t. \tag{20.1.2}$$

Here $\mathcal{K}'$and $\mathcal{A}'$ are $(n \times n_1)$ and $(n \times n_2)$ matrices, respectively. The matrices $\mathcal{K}'$ and $\mathcal{A}'$ are chosen so that the following conditions hold.

(1) The individual elements of $\boldsymbol{\eta}_t$ have unit variance and are uncorrelated with one another:

$$E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \mathcal{K}'\Sigma_{\mathbf{YY}}\mathcal{K} = \mathbf{I}_n. \tag{20.1.3}$$

(2) The individual elements of $\boldsymbol{\xi}_t$ have unit variance and are uncorrelated with one another:

$$E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \mathcal{A}'\Sigma_{\mathbf{XX}}\mathcal{A} = \mathbf{I}_n. \tag{20.1.4}$$

(3) The ith element of $\boldsymbol{\eta}_t$ is uncorrelated with the jth element of $\boldsymbol{\xi}_t$ for $i \neq j$; for $i = j$, the correlation is positive and is given by r_i:

$$E(\boldsymbol{\xi}_t\boldsymbol{\eta}_t') = \mathcal{A}'\Sigma_{\mathbf{XY}}\mathcal{K} = \mathbf{R}, \tag{20.1.5}$$

where

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix}. \tag{20.1.6}$$

(4) The elements of $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ are ordered in such a way that

$$(1 \geq r_1 \geq r_2 \geq \cdots \geq r_n \geq 0). \tag{20.1.7}$$

The correlation r_i is known as the ith *population canonical correlation* between $\mathbf{y}_t$ and $\mathbf{x}_t$.

The population canonical correlations and the values of $\mathcal{K}$ and $\mathcal{A}$ can be calculated from $\Sigma_{\mathbf{YY}}$, $\Sigma_{\mathbf{XX}}$, and $\Sigma_{\mathbf{XY}}$ using any computer program that generates eigenvalues and eigenvectors, as we now describe.

Let $(\lambda_1, \lambda_2, \ldots, \lambda_{n_1})$ denote the eigenvalues of the $(n_1 \times n_1)$ matrix

$$\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}, \tag{20.1.8}$$

ordered as

$$(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_1}), \tag{20.1.9}$$

with associated eigenvectors $(\tilde{\mathbf{k}}_1, \tilde{\mathbf{k}}_2, \ldots, \tilde{\mathbf{k}}_{n_1})$. Recall that the eigenvalue-eigenvector pair $(\lambda_i, \tilde{\mathbf{k}}_i)$ satisfies

$$\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\tilde{\mathbf{k}}_i = \lambda_i\tilde{\mathbf{k}}_i. \tag{20.1.10}$$

Notice that if $\tilde{\mathbf{k}}_i$ satisfies [20.1.10], then so does $c\tilde{\mathbf{k}}_i$ for any value of c. The usual normalization convention for choosing c and thus for determining "the" eigenvector $\tilde{\mathbf{k}}_i$ to associate with λ_i is to set $\tilde{\mathbf{k}}_i'\tilde{\mathbf{k}}_i = 1$. For canonical correlation analysis, however, it is more convenient to choose c so as to ensure that

$$\mathbf{k}_i'\Sigma_{\mathbf{YY}}\mathbf{k}_i = 1 \qquad \text{for } i = 1, 2, \ldots, n_1. \qquad [20.1.11]$$

If a computer program has calculated eigenvectors $(\tilde{\mathbf{k}}_1, \tilde{\mathbf{k}}_2, \ldots, \tilde{\mathbf{k}}_{n_1})$ of the matrix in [20.1.8] normalized by $(\tilde{\mathbf{k}}_i'\tilde{\mathbf{k}}_i) = 1$, it is trivial to change these to eigenvectors $(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{n_1})$ normalized by the condition [20.1.11] by setting

$$\mathbf{k}_i = \tilde{\mathbf{k}}_i \div \sqrt{\tilde{\mathbf{k}}_i'\Sigma_{\mathbf{YY}}\tilde{\mathbf{k}}_i}.$$

We further may multiply $\mathbf{k}_i$ by -1 so as to satisfy a certain sign convention to be detailed in the paragraphs following the next proposition.

The canonical correlations $(r_1, r_2, \ldots, r_n)$ turn out to be given by the square roots of the corresponding first n eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ of [20.1.8]. The associated $(n_1 \times 1)$ eigenvectors $\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_n$, when normalized by [20.1.11] and a sign convention, turn out to make up the rows of the $(n \times n_1)$ matrix $\mathcal{K}'$ appearing in [20.1.1]. The matrix $\mathcal{A}'$ in [20.1.2] can be obtained from the normalized eigenvectors of a matrix closely related to [20.1.8]. These results are developed in the following proposition, proved in Appendix 20.A at the end of this chapter.

Proposition 20.1: *Let*

$$\underset{(n_1+n_2)\times(n_1+n_2)}{\Sigma} \equiv \begin{bmatrix} \underset{(n_1\times n_1)}{\Sigma_{\mathbf{YY}}} & \underset{(n_1\times n_2)}{\Sigma_{\mathbf{YX}}} \\ \underset{(n_2\times n_1)}{\Sigma_{\mathbf{XY}}} & \underset{(n_2\times n_2)}{\Sigma_{\mathbf{XX}}} \end{bmatrix}$$

be a positive definite symmetric matrix and let $(\lambda_1, \lambda_2, \ldots, \lambda_{n_1})$ be the eigenvalues of the matrix in [20.1.8], ordered $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_1}$. Let $(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{n_1})$ be the associated $(n_1 \times 1)$ eigenvectors as normalized by [20.1.11]. Let $(\mu_1, \mu_2, \ldots, \mu_{n_2})$ be the eigenvalues of the $(n_2 \times n_2)$ matrix

$$\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}, \qquad [20.1.12]$$

ordered $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{n_2}$. Let $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n_2})$ be the eigenvectors of [20.1.12]:

$$\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\mathbf{a}_i = \mu_i\mathbf{a}_i, \qquad [20.1.13]$$

normalized by

$$\mathbf{a}_i'\Sigma_{\mathbf{XX}}\mathbf{a}_i = 1 \qquad \text{for } i = 1, 2, \ldots, n_2. \qquad [20.1.14]$$

Let n be the smaller of n_1 and n_2, and collect the first n vectors $\mathbf{k}_i$ and the first n vectors $\mathbf{a}_i$ in matrices

$$\underset{(n_1\times n)}{\mathcal{K}} = [\mathbf{k}_1 \;\; \mathbf{k}_2 \;\; \cdots \;\; \mathbf{k}_n]$$

$$\underset{(n_2\times n)}{\mathcal{A}} = [\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \cdots \;\; \mathbf{a}_n].$$

Assuming that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct, then

(a) $0 \leq \lambda_i < 1$ *for* $i = 1, 2, \ldots, n_1$ *and* $0 \leq \mu_j < 1$ *for* $j = 1, 2, \ldots, n_2$;
(b) $\lambda_i = \mu_i$ *for* $i = 1, 2, \ldots, n$;
(c) $\mathcal{K}'\Sigma_{\mathbf{YY}}\mathcal{K} = \mathbf{I}_n$ *and* $\mathcal{A}'\Sigma_{\mathbf{XX}}\mathcal{A} = \mathbf{I}_n$;
(d) $\mathcal{A}'\Sigma_{\mathbf{XY}}\mathcal{K} = \mathbf{R}$,

where $\mathbf{R}$ is a diagonal matrix whose squared diagonal elements correspond to the *eigenvalues of* [20.1.8]:

$$\mathbf{R}^2 = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

If Σ denotes the variance-covariance matrix of the vector $(\mathbf{y}_t', \mathbf{x}_t')'$, then results (c) and (d) are the characterization of the canonical correlations given in [20.1.3] through [20.1.5]. Thus, the proposition establishes that the squares of the canonical correlations $(r_1^2, r_2^2, \ldots, r_n^2)$ can be found from the first n eigenvalues of the matrix in [20.1.8]. Result (b) states that these are the same as the first n eigenvalues of the matrix in [20.1.12]. The matrices $\mathcal{K}$ and $\mathcal{A}$ that characterize the canonical variates in [20.1.1] and [20.1.2] can be found from the normalized eigenvectors of these matrices.

The magnitude $\mathbf{a}_i'\Sigma_{\mathbf{XY}}\mathbf{k}_i$ calculated by the algorithm described in Proposition 20.1 need not be positive—the proposition only ensures that its square is equal to the square of the corresponding canonical correlation. If $\mathbf{a}_i'\Sigma_{\mathbf{XY}}\mathbf{k}_i < 0$ for some i, one can replace $\mathbf{k}_i$ as calculated with $-\mathbf{k}_i$, so that the ith diagonal element of $\mathbf{R}$ will correspond to the positive square root of λ_i.

As an illustration, suppose that $\mathbf{y}_t$ consists of a single variable ($n_1 = n = 1$). In this case, the matrix [20.1.8] is just a scalar, a (1×1) "matrix" that is equal to its own eigenvalue. Thus, the squared population canonical correlation between a scalar y_t and a set of n_2 explanatory variables $\mathbf{x}_t$ is given by

$$r_1^2 = \frac{\Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{X}Y}}{\Sigma_{YY}}.$$

To interpret this expression, recall from equation [4.1.15] that the mean squared error of a linear projection of y_t on $\mathbf{x}_t$ is given by

$$MSE = \Sigma_{YY} - \Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{X}Y},$$

and so

$$1 - r_1^2 = \frac{\Sigma_{YY}}{\Sigma_{YY}} - \frac{\Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{X}Y}}{\Sigma_{YY}} = \frac{MSE}{\Sigma_{YY}}. \qquad [20.1.15]$$

Thus, for this simple case, r_1^2 is the fraction of the population variance that is explained by the linear projection; that is, r_1^2 is the population squared multiple correlation coefficient, commonly denoted R^2.

Another interpretation of canonical correlations is also sometimes helpful. The first canonical variates η_{1t} and ξ_{1t} can be interpreted as those linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$, respectively, such that the correlation between η_{1t} and ξ_{1t} is as large as possible (see Exercise 20.1). The variates η_{2t} and ξ_{2t} give those linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$ that are uncorrelated with η_{1t} and ξ_{1t} and yet yield the largest remaining correlation between η_{2t} and ξ_{2t}, and so on.

Sample Canonical Correlations

The canonical correlations r_i calculated by the procedure just described are population parameters—they are functions of the population moments $\Sigma_{\mathbf{YY}}$, $\Sigma_{\mathbf{YX}}$, and $\Sigma_{\mathbf{XX}}$. Here we describe their sample analogs, to be denoted $\hat{r}_i$.

Suppose we have a sample of T observations on the $(n_1 \times 1)$ vector $\mathbf{y}_t$ and the $(n_2 \times 1)$ vector $\mathbf{x}_t$, whose sample moments are given by

$$\hat{\Sigma}_{\mathbf{YY}} = (1/T) \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' \qquad [20.1.16]$$

$$\hat{\Sigma}_{\mathbf{YX}} = (1/T) \sum_{t=1}^{T} \mathbf{y}_t \mathbf{x}_t' \qquad [20.1.17]$$

$$\hat{\Sigma}_{\mathbf{XX}} = (1/T) \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'. \qquad [20.1.18]$$

Again, in many applications, $\mathbf{y}_t$ and $\mathbf{x}_t$ would be measured in deviations from their sample means.

To calculate sample canonical correlations, the objective is to generate a set of T observations on a new $(n \times 1)$ vector $\hat{\boldsymbol{\eta}}_t$, where n is the smaller of n_1 and n_2. The vector $\hat{\boldsymbol{\eta}}_t$ is a linear combination of the observed value of $\mathbf{y}_t$:

$$\hat{\boldsymbol{\eta}}_t = \hat{\mathcal{K}}' \mathbf{y}_t, \qquad [20.1.19]$$

for $\hat{\mathcal{K}}$ an $(n_1 \times n)$ matrix to be estimated from the data. The task will be to choose $\hat{\mathcal{K}}$ so that the ith generated series $(\hat{\eta}_{it})$ has unit sample variance and is orthogonal to the jth generated series:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_t \hat{\boldsymbol{\eta}}_t' = \mathbf{I}_n. \qquad [20.1.20]$$

Similarly, we will generate an $(n \times 1)$ vector $\hat{\boldsymbol{\xi}}_t$ from the elements of $\mathbf{x}_t$:

$$\hat{\boldsymbol{\xi}}_t = \hat{\mathcal{A}}' \mathbf{x}_t. \qquad [20.1.21]$$

Each of the variables $\hat{\xi}_{it}$ has unit sample variance and is orthogonal to $\hat{\xi}_{jt}$ for $i \neq j$:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t \hat{\boldsymbol{\xi}}_t' = \mathbf{I}_n. \qquad [20.1.22]$$

Finally, $\hat{\eta}_{it}$ is orthogonal to $\hat{\xi}_{jt}$ for $i \neq j$, while the sample correlation between $\hat{\eta}_{it}$ and $\hat{\xi}_{it}$ is called the *sample canonical correlation coefficient*:

$$(1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t \hat{\boldsymbol{\eta}}_t' = \hat{\mathbf{R}} \qquad [20.1.23]$$

for

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}_1 & 0 & \cdots & 0 \\ 0 & \hat{r}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \hat{r}_n \end{bmatrix}. \qquad [20.1.24]$$

Finding matrices $\hat{\mathcal{K}}$, $\hat{\mathcal{A}}$, and $\hat{\mathbf{R}}$ satisfying [20.1.20], [20.1.22], and [20.1.23] involves exactly the same calculations as did finding matrices $\mathcal{K}$, $\mathcal{A}$, and $\mathbf{R}$ satisfying [20.1.3] through [20.1.5]. For example, [20.1.19] allows us to write [20.1.20] as

$$\mathbf{I}_n = (1/T) \sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_t \hat{\boldsymbol{\eta}}_t' = \hat{\mathcal{K}}'(1/T) \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' \hat{\mathcal{K}} = \hat{\mathcal{K}}' \hat{\Sigma}_{\mathbf{YY}} \hat{\mathcal{K}}, \qquad [20.1.25]$$

where the last line follows from [20.1.16]. Expression [20.1.25] is identical to

[20.1.3] with hats placed over the variables. Similarly, substituting [20.1.21] into [20.1.22] gives $\hat{\mathcal{A}}'\hat{\Sigma}_{XX}\hat{\mathcal{A}} = \mathbf{I}_n$, which corresponds to [20.1.4]. Equation [20.1.23] becomes $\hat{\mathcal{A}}'\hat{\Sigma}_{XY}\hat{\mathcal{K}} = \hat{\mathbf{R}}$, as in [20.1.5]. Again, we can replace $\hat{\mathbf{k}}_i$ with $-\hat{\mathbf{k}}_i$ if any of the elements of $\hat{\mathbf{R}}$ should turn out negative.

Thus, to calculate the sample canonical correlations, the procedure described in Proposition 20.1 is simply applied to the sample moments ($\hat{\Sigma}_{YY}$, $\hat{\Sigma}_{YX}$, and $\hat{\Sigma}_{XX}$) rather than to the population moments. In particular, the square of the ith sample canonical correlation ($\hat{r}_i^2$) is given by the ith largest eigenvalue of the matrix

$$\begin{aligned}\hat{\Sigma}_{YY}^{-1}\hat{\Sigma}_{YX}\hat{\Sigma}_{XX}^{-1}\hat{\Sigma}_{XY} &= \left\{(1/T)\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_t'\right\} \\ &\quad\times\left\{(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}\mathbf{x}_t\mathbf{y}_t'\right\}.\end{aligned} \qquad [20.1.26]$$

The ith column of $\hat{\mathcal{K}}$ is given by the eigenvector associated with this ith eigenvalue, normalized so that

$$\hat{\mathbf{k}}_i'\left\{(1/T)\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right\}\hat{\mathbf{k}}_i = 1.$$

The ith column of $\hat{\mathcal{A}}$ is given by the eigenvector associated with the eigenvalue $\hat{\lambda}_i$ for the matrix $\hat{\Sigma}_{XX}^{-1}\hat{\Sigma}_{XY}\hat{\Sigma}_{YY}^{-1}\hat{\Sigma}_{YX}$ normalized by the condition that $\hat{\mathbf{a}}_i'\hat{\Sigma}_{XX}\hat{\mathbf{a}}_i = 1$.

For example, suppose that $\mathbf{y}_t$ is a scalar ($n = n_1 = 1$). Then [20.1.26] is a scalar equal to its own eigenvalue. Hence, the sample squared canonical correlation between the scalar $\mathbf{y}_t$ and a set of n_2 explanatory variables $\mathbf{x}_t$ is given by

$$\begin{aligned}\hat{r}_1^2 &= \frac{\{T^{-1}\Sigma y_t\mathbf{x}_t'\}\{T^{-1}\Sigma\mathbf{x}_t\mathbf{x}_t'\}^{-1}\{T^{-1}\Sigma\mathbf{x}_t y_t\}}{\{T^{-1}\Sigma y_t^2\}} \\ &= \frac{\{\Sigma y_t\mathbf{x}_t'\}\{\Sigma\mathbf{x}_t\mathbf{x}_t'\}^{-1}\{\Sigma\mathbf{x}_t y_t\}}{\{\Sigma y_t^2\}},\end{aligned}$$

which is just the squared sample multiple correlation coefficient R^2.

20.2. *Maximum Likelihood Estimation*

We are now in a position to describe Johansen's approach (1988, 1991) to full-information maximum likelihood estimation of a system characterized by exactly h cointegrating relations.

Let $\mathbf{y}_t$ denote an $(n \times 1)$ vector. The maintained hypothesis is that $\mathbf{y}_t$ follows a *VAR*(p) in levels. Recall from equation [19.1.39] that any pth-order *VAR* can be written in the form

$$\begin{aligned}\Delta\mathbf{y}_t &= \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} \\ &\quad + \boldsymbol{\alpha} + \boldsymbol{\zeta}_0\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,\end{aligned} \qquad [20.2.1]$$

with

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

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

Suppose that each individual variable y_{it} is $I(1)$, although h linear combinations of $\mathbf{y}_t$ are stationary. We saw in equations [19.1.35] and [19.1.40] that this implies that $\boldsymbol{\zeta}_0$ can be written in the form

$$\boldsymbol{\zeta}_0 = -\mathbf{B}\mathbf{A}' \qquad [20.2.2]$$

for $\mathbf{B}$ an $(n \times h)$ matrix and $\mathbf{A}'$ an $(h \times n)$ matrix. That is, under the hypothesis of h cointegrating relations, only h separate linear combinations of the level of $\mathbf{y}_{t-1}$ (the h elements of $\mathbf{z}_{t-1} = \mathbf{A}'\mathbf{y}_{t-1}$) appear in [20.2.1].

Consider a sample of $T + p$ observations on $\mathbf{y}$, denoted $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \ldots, \mathbf{y}_T)$. If the disturbances $\boldsymbol{\varepsilon}_t$ are Gaussian, then the log likelihood of $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$ conditional on $(\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \ldots, \mathbf{y}_0)$ is given by

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\Omega}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1}, \boldsymbol{\alpha}, \boldsymbol{\zeta}_0) \\
&\quad = (-Tn/2)\log(2\pi) - (T/2)\log|\boldsymbol{\Omega}| \\
&\quad\quad - (1/2)\sum_{t=1}^{T}\Big[(\Delta\mathbf{y}_t - \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} - \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} - \cdots - \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} - \boldsymbol{\alpha} - \boldsymbol{\zeta}_0\mathbf{y}_{t-1})' \\
&\quad\quad \times \boldsymbol{\Omega}^{-1}(\Delta\mathbf{y}_t - \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} - \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} - \cdots - \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} - \boldsymbol{\alpha} - \boldsymbol{\zeta}_0\mathbf{y}_{t-1})\Big].
\end{aligned} \qquad [20.2.3]$$

The goal is to chose $(\boldsymbol{\Omega}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1}, \boldsymbol{\alpha}, \boldsymbol{\zeta}_0)$ so as to maximize [20.2.3] subject to the constraint that $\boldsymbol{\zeta}_0$ can be written in the form of [20.2.2].

We will first summarize Johansen's algorithm, and then verify that it indeed calculates the maximum likelihood estimates.

Step 1: Calculate Auxiliary Regressions

The first step is to estimate a $(p-1)$th-order *VAR* for $\Delta\mathbf{y}_t$; that is, regress the scalar Δy_{it} on a constant and all the elements of the vectors $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$ by *OLS*. Collect the $i = 1, 2, \ldots, n$ *OLS* regressions in vector form as

$$\Delta\mathbf{y}_t = \hat{\boldsymbol{\pi}}_0 + \hat{\boldsymbol{\Pi}}_1\Delta\mathbf{y}_{t-1} + \hat{\boldsymbol{\Pi}}_2\Delta\mathbf{y}_{t-2} + \cdots + \hat{\boldsymbol{\Pi}}_{p-1}\Delta\mathbf{y}_{t-p+1} + \hat{\mathbf{u}}_t, \qquad [20.2.4]$$

where $\hat{\boldsymbol{\Pi}}_i$ denotes an $(n \times n)$ matrix of *OLS* coefficient estimates and $\hat{\mathbf{u}}_t$ denotes the $(n \times 1)$ vector of *OLS* residuals. We also estimate a second battery of regressions, regressing the scalar $y_{i,t-1}$ on a constant and $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$ for $i = 1, 2, \ldots, n$. Write this second set of *OLS* regressions as[1]

$$\mathbf{y}_{t-1} = \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\aleph}}_1\Delta\mathbf{y}_{t-1} + \hat{\boldsymbol{\aleph}}_2\Delta\mathbf{y}_{t-2} + \cdots + \hat{\boldsymbol{\aleph}}_{p-1}\Delta\mathbf{y}_{t-p+1} + \hat{\mathbf{v}}_t, \qquad [20.2.5]$$

with $\hat{\mathbf{v}}_t$ the $(n \times 1)$ vector of residuals from this second battery of regressions.

[1]Johansen (1991) described his procedure as calculating $\check{\mathbf{v}}_t$ in place of $\hat{\mathbf{v}}_t$, where $\check{\mathbf{v}}_t$ is the *OLS* residual from a regression of $\mathbf{y}_{t-p}$ on a constant and $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$. Since $\mathbf{y}_{t-p} = \mathbf{y}_{t-1} - \Delta\mathbf{y}_{t-1} - \Delta\mathbf{y}_{t-2} - \cdots - \Delta\mathbf{y}_{t-p+1}$, the residual $\check{\mathbf{v}}_t$ is numerically identical to $\hat{\mathbf{v}}_t$ described in the text.

Step 2: Calculate Canonical Correlations

Next calculate the sample variance-covariance matrices of the *OLS* residuals $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$:

$$\hat{\Sigma}_{VV} \equiv (1/T) \sum_{t=1}^{T} \hat{\mathbf{v}}_t \hat{\mathbf{v}}_t' \qquad [20.2.6]$$

$$\hat{\Sigma}_{UU} \equiv (1/T) \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \qquad [20.2.7]$$

$$\hat{\Sigma}_{UV} \equiv (1/T) \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{v}}_t' \qquad [20.2.8]$$

$$\hat{\Sigma}_{VU} \equiv \hat{\Sigma}_{UV}'.$$

From these, find the eigenvalues of the matrix

$$\hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{VU} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{UV} \qquad [20.2.9]$$

with the eigenvalues ordered $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$. The maximum value attained by the log likelihood function subject to the constraint that there are h cointegrating relations is given by

$$\begin{aligned} \mathcal{L}^* = & -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\Sigma}_{UU}| \\ & - (T/2) \sum_{i=1}^{h} \log(1 - \hat{\lambda}_i). \end{aligned} \qquad [20.2.10]$$

Step 3: Calculate Maximum Likelihood Estimates of Parameters

If we are interested only in a likelihood ratio test of the number of cointegrating relations, step 2 provides all the information needed. If maximum likelihood estimates of parameters are also desired, these can be calculated as follows. Let $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_h$ denote the $(n \times 1)$ eigenvectors of [20.2.9] associated with the h largest eigenvalues. These provide a basis for the space of cointegrating relations; that is, the maximum likelihood estimate is that any cointegrating vector can be written in the form

$$\mathbf{a} = b_1\hat{\mathbf{a}}_1 + b_2\hat{\mathbf{a}}_2 + \cdots + b_h\hat{\mathbf{a}}_h$$

for some choice of scalars $(b_1, b_2, \ldots; b_h)$. Johansen suggested normalizing these vectors $\hat{\mathbf{a}}_i$ so that $\hat{\mathbf{a}}_i'\hat{\Sigma}_{VV}\hat{\mathbf{a}}_i = 1$. For example, if the eigenvectors $\tilde{\mathbf{a}}_i$ of [20.2.9] are calculated from a standard computer program that normalizes $\tilde{\mathbf{a}}_i'\tilde{\mathbf{a}}_i = 1$, Johansen's estimate is $\hat{\mathbf{a}}_i = \tilde{\mathbf{a}}_i \div \sqrt{\tilde{\mathbf{a}}_i'\hat{\Sigma}_{VV}\tilde{\mathbf{a}}_i}$. Collect the first h normalized vectors in an $(n \times h)$ matrix $\hat{\mathbf{A}}$:

$$\hat{\mathbf{A}} \equiv [\hat{\mathbf{a}}_1 \quad \hat{\mathbf{a}}_2 \quad \cdots \quad \hat{\mathbf{a}}_h]. \qquad [20.2.11]$$

Then the *MLE* of $\boldsymbol{\zeta}_0$ is given by

$$\hat{\boldsymbol{\zeta}}_0 = \hat{\Sigma}_{UV}\hat{\mathbf{A}}\hat{\mathbf{A}}'. \qquad [20.2.12]$$

The *MLE* of $\boldsymbol{\zeta}_i$ for $i = 1, 2, \ldots, p - 1$ is

$$\hat{\boldsymbol{\zeta}}_i = \hat{\Pi}_i - \hat{\boldsymbol{\zeta}}_0\hat{\aleph}_i, \qquad [20.2.13]$$

and the *MLE* of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\pi}}_0 - \hat{\boldsymbol{\zeta}}_0\hat{\boldsymbol{\theta}}. \qquad [20.2.14]$$

The *MLE* of $\boldsymbol{\Omega}$ is

$$\hat{\boldsymbol{\Omega}} = (1/T) \sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \hat{\boldsymbol{\zeta}}_0 \hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \hat{\boldsymbol{\zeta}}_0 \hat{\mathbf{v}}_t)']. \qquad [20.2.15]$$

We now review the logic behind each of these steps in turn.

Motivation for Auxiliary Regressions

The first step involves *concentrating* the likelihood function.[2] This means taking $\boldsymbol{\Omega}$ and $\boldsymbol{\zeta}_0$ as given and maximizing [20.2.3] with respect to $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$. This restricted maximization problem takes the form of seemingly unrelated regressions of the elements of the $(n \times 1)$ vector $\Delta\mathbf{y}_t - \boldsymbol{\zeta}_0\mathbf{y}_{t-1}$ on a constant and the explanatory variables $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$. Since each of the n regressions in this system has the identical explanatory variables, the estimates of $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$ would come from *OLS* regressions of each of the elements of $\Delta\mathbf{y}_t - \boldsymbol{\zeta}_0\mathbf{y}_{t-1}$ on a constant and $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$. Denote the values of $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$ that maximize [20.2.3] for a given value of $\boldsymbol{\zeta}_0$ by

$$[\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0), \ldots, \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0)].$$

These values are characterized by the condition that the following residual vector must have sample mean zero and be orthogonal to $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$:

$$\begin{aligned}[][\Delta\mathbf{y}_t - \boldsymbol{\zeta}_0\mathbf{y}_{t-1}] - \{\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0) + \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0)\Delta\mathbf{y}_{t-1} + \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0)\Delta\mathbf{y}_{t-2} \\ + \cdots + \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0)\Delta\mathbf{y}_{t-p+1}\}.\end{aligned} \qquad [20.2.16]$$

But notice that the *OLS* residuals $\hat{\mathbf{u}}_t$ in [20.2.4] and $\hat{\mathbf{v}}_t$ in [20.2.5] each satisfy this orthogonality requirement, and therefore the vector $\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t$ also has sample mean zero and is orthogonal to $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$. Moreover, $\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t$ is of the form of expression [20.2.16],

$$\begin{aligned}\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t = (\Delta\mathbf{y}_t - \hat{\boldsymbol{\pi}}_0 - \hat{\boldsymbol{\Pi}}_1\Delta\mathbf{y}_{t-1} - \hat{\boldsymbol{\Pi}}_2\Delta\mathbf{y}_{t-2} - \cdots - \hat{\boldsymbol{\Pi}}_{p-1}\Delta\mathbf{y}_{t-p+1}) \\ - \boldsymbol{\zeta}_0(\mathbf{y}_{t-1} - \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\aleph}}_1\Delta\mathbf{y}_{t-1} - \hat{\boldsymbol{\aleph}}_2\Delta\mathbf{y}_{t-2} - \cdots - \hat{\boldsymbol{\aleph}}_{p-1}\Delta\mathbf{y}_{t-p+1}),\end{aligned}$$

with

$$\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0) = \hat{\boldsymbol{\pi}}_0 - \boldsymbol{\zeta}_0\hat{\boldsymbol{\theta}} \qquad [20.2.17]$$

$$\hat{\boldsymbol{\zeta}}_i^*(\boldsymbol{\zeta}_0) = \hat{\boldsymbol{\Pi}}_i - \boldsymbol{\zeta}_0\hat{\boldsymbol{\aleph}}_i \qquad \text{for } i = 1, 2, \ldots, p-1. \qquad [20.2.18]$$

Thus, the vector in [20.2.16] is given by $\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t$.

The concentrated log likelihood function (to be denoted $\mathcal{M}$) is found by replacing $(\boldsymbol{\alpha}, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_{p-1})$ in [20.2.3] with $[\hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0), \ldots, \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0)]$:

$$\begin{aligned}\mathcal{M}(\boldsymbol{\Omega}, \boldsymbol{\zeta}_0) &\equiv \mathcal{L}\{\boldsymbol{\Omega}, \hat{\boldsymbol{\zeta}}_1^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\zeta}}_2^*(\boldsymbol{\zeta}_0), \ldots, \hat{\boldsymbol{\zeta}}_{p-1}^*(\boldsymbol{\zeta}_0), \hat{\boldsymbol{\alpha}}^*(\boldsymbol{\zeta}_0), \boldsymbol{\zeta}_0\} \\ &= -(Tn/2)\log(2\pi) - (T/2)\log|\boldsymbol{\Omega}| \\ &\quad - (1/2)\sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)'\boldsymbol{\Omega}^{-1}(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)].\end{aligned} \qquad [20.2.19]$$

The idea behind concentrating the likelihood function in this way is that if we can find the values of $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\zeta}}_0$ for which $\mathcal{M}$ is maximized, then these same values (along with $\hat{\boldsymbol{\alpha}}^*(\hat{\boldsymbol{\zeta}}_0)$ and $\hat{\boldsymbol{\zeta}}_i^*(\hat{\boldsymbol{\zeta}}_0)$) will maximize [20.2.3].

[2]See Koopmans and Hood (1953, pp. 156–58) for more background on concentration of likelihood functions.

Continuing the concentration one step further, recall from the analysis of [11.1.25] that the value of $\boldsymbol{\Omega}$ that maximizes [20.2.19] (still regarding $\boldsymbol{\zeta}_0$ as fixed) is given by

$$\hat{\boldsymbol{\Omega}}^*(\boldsymbol{\zeta}_0) = (1/T)\sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)']. \qquad [20.2.20]$$

As in expression [11.1.32], the value obtained for [20.2.19] when evaluated at [20.2.20] is then

$$\begin{aligned} \mathcal{N}(\boldsymbol{\zeta}_0) &\equiv \mathcal{M}\{\hat{\boldsymbol{\Omega}}^*(\boldsymbol{\zeta}_0), \boldsymbol{\zeta}_0\} \\ &= -(Tn/2)\log(2\pi) - (T/2)\log|\hat{\boldsymbol{\Omega}}^*(\boldsymbol{\zeta}_0)| - (Tn/2) \\ &= -(Tn/2)\log(2\pi) - (Tn/2) \qquad [20.2.21] \\ &\quad - (T/2)\log\left|(1/T)\sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)']\right|. \end{aligned}$$

Expression [20.2.21] represents the biggest value one can achieve for the log likelihood for any given value of $\boldsymbol{\zeta}_0$. Maximizing the likelihood function thus comes down to choosing $\boldsymbol{\zeta}_0$ so as to minimize

$$\left|(1/T)\sum_{t=1}^{T} [(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)']\right| \qquad [20.2.22]$$

subject to the constraint of [20.2.2].

Motivation for Canonical Correlation Analysis

To see the motivation for calculating canonical correlations, consider first a simpler problem. Suppose that by an astounding coincidence, $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$ were already in canonical form,

$$\begin{aligned} \hat{\mathbf{u}}_t &= \hat{\boldsymbol{\eta}}_t \\ \hat{\mathbf{v}}_t &= \hat{\boldsymbol{\xi}}_t, \end{aligned}$$

with

$$(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t' = \mathbf{I}_n \qquad [20.2.23]$$

$$(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t' = \mathbf{I}_n \qquad [20.2.24]$$

$$(1/T)\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\eta}}_t' = \hat{\mathbf{R}} \qquad [20.2.25]$$

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}_1 & 0 & \cdots & 0 \\ 0 & \hat{r}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \hat{r}_n \end{bmatrix}. \qquad [20.2.26]$$

Suppose that for these canonical data we were asked to choose $\boldsymbol{\zeta}_0$ so as to minimize

$$\left|(1/T)\sum_{t=1}^{T} \left[(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0\hat{\boldsymbol{\xi}}_t)(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0\hat{\boldsymbol{\xi}}_t)'\right]\right| \qquad [20.2.27]$$

subject to the constraint that $\zeta_0\hat{\xi}_t$ could make use of only h linear combinations of $\hat{\xi}_t$. If there were no restrictions on ζ_0 (so that $h = n$), then expression [20.2.27] would be minimized by OLS regressions of $\hat{\eta}_{it}$ on $\hat{\xi}_t$ for $i = 1, 2, \ldots, n$. Conditions [20.2.24] and [20.2.25] establish that the ith regression would have an estimated coefficient vector of

$$\left\{(1/T)\sum_{t=1}^{T}\hat{\xi}_t\hat{\xi}_t'\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}\hat{\xi}_t\hat{\eta}_{it}\right\} = \hat{r}_i \cdot \mathbf{e}_i,$$

where $\mathbf{e}_i$ denotes the ith column of $\mathbf{I}_n$. Thus, even if all n elements of $\hat{\xi}_t$ appeared in the regression, only the ith element $\hat{\xi}_{it}$ would have a nonzero coefficient in the regression used to explain $\hat{\eta}_{it}$. The average squared residual for this regression would be

$$\begin{aligned}\left\{(1/T)\sum_{t=1}^{T}(\hat{\eta}_{it})^2\right\} - \left\{(1/T)\sum_{t=1}^{T}(\hat{\eta}_{it}\hat{\xi}_t')\right\}&\left\{(1/T)\sum_{t=1}^{T}(\hat{\xi}_t\hat{\xi}_t')\right\}^{-1}\left\{(1/T)\sum_{t=1}^{T}(\hat{\xi}_t\hat{\eta}_{it})\right\} \\ &= 1 - \hat{r}_i \cdot \mathbf{e}_i' \cdot \mathbf{I}_n \cdot \mathbf{e}_i \cdot \hat{r}_i \\ &= 1 - \hat{r}_i^2.\end{aligned}$$

Moreover, conditions [20.2.23] through [20.2.25] imply that the residual for the ith regression, $\hat{\eta}_{it} - \hat{r}_i\hat{\xi}_{it}$, would be orthogonal to the residual from the jth regression, $\hat{\eta}_{jt} - \hat{r}_j\hat{\xi}_{jt}$, for $i \neq j$. Thus, if ζ_0 were unrestricted, the optimal value for the matrix in [20.2.27] would be a diagonal matrix with $(1 - \hat{r}_i^2)$ in the row i, column i position and zero elsewhere.

Now suppose that we are restricted to use only h linear combinations of $\hat{\xi}_t$ as regressors. From the preceding analysis, we might guess that the best we can do is use the h elements of $\hat{\xi}_t$ that have the highest correlations with elements of $\hat{\boldsymbol{\eta}}_t$, that is, choose $(\hat{\xi}_{1t}, \hat{\xi}_{2t}, \ldots, \hat{\xi}_{ht})$ as regressors.[3] When this set of regressors is used to explain $\hat{\eta}_{it}$ for $i \leq h$, the average squared residual will be $(1 - \hat{r}_i^2)$, as before. When this set of regressors is used to explain $\hat{\eta}_{it}$ for $i > h$, all of the regressors are orthogonal to $\hat{\eta}_{it}$ and would receive regression coefficients of zero. The average squared residual for the latter regression is simply $(1/T)\Sigma_{t=1}^{T}\hat{\eta}_{it}^2 = 1$ for $i = h + 1, h + 2, \ldots, n$. Thus, if we are restricted to using only h linear combinations of $\hat{\xi}_t$, the optimized value of [20.2.27] will be

$$\begin{aligned}&\left|(1/T)\sum_{t=1}^{T}[(\hat{\boldsymbol{\eta}}_t - \zeta_0^*\hat{\xi}_t)(\hat{\boldsymbol{\eta}}_t - \zeta_0^*\hat{\xi}_t)']\right| \\ &= \begin{vmatrix} 1-\hat{r}_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1-\hat{r}_2^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1-\hat{r}_h^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{vmatrix} \qquad [20.2.28] \\ &= \prod_{i=1}^{h}(1 - \hat{r}_i^2).\end{aligned}$$

[3]See Johansen (1988) for a more formal demonstration of this claim.

Of course, the actual data $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$ will not be in exact canonical form. However, the previous section described how to find $(n \times n)$ matrices $\hat{\mathcal{K}}$ and $\hat{\mathcal{A}}$ such that

$$\hat{\boldsymbol{\eta}}_t = \hat{\mathcal{K}}'\hat{\mathbf{u}}_t \qquad [20.2.29]$$

$$\hat{\boldsymbol{\xi}}_t = \hat{\mathcal{A}}'\hat{\mathbf{v}}_t. \qquad [20.2.30]$$

The columns of $\hat{\mathcal{A}}$ are given by the eigenvectors of the matrix in [20.2.9], normalized by the condition $\hat{\mathcal{A}}'\hat{\boldsymbol{\Sigma}}_{VV}\hat{\mathcal{A}} = \mathbf{I}_n$. The eigenvalues of [20.2.9] give the squares of the canonical correlations:

$$\hat{\lambda}_i = \hat{r}_i^2. \qquad [20.2.31]$$

The columns of $\hat{\mathcal{K}}$ correspond to the normalized eigenvectors of the matrix $\hat{\boldsymbol{\Sigma}}_{UU}^{-1}\hat{\boldsymbol{\Sigma}}_{UV}\hat{\boldsymbol{\Sigma}}_{VV}^{-1}\hat{\boldsymbol{\Sigma}}_{VU}$, though it turns out that $\hat{\mathcal{K}}$ does not actually have to be calculated in order to use the following results. Assuming that $\hat{\mathcal{K}}$ and $\hat{\mathcal{A}}$ are nonsingular, [20.2.29] and [20.2.30] allow [20.2.22] to be written

$$\begin{aligned}
&\left|(1/T)\sum_{t=1}^{T}\left[(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)(\hat{\mathbf{u}}_t - \boldsymbol{\zeta}_0\hat{\mathbf{v}}_t)'\right]\right| \\
&= \left|(1/T)\sum_{t=1}^{T}\left[[(\hat{\mathcal{K}}')^{-1}\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t][(\hat{\mathcal{K}}')^{-1}\hat{\boldsymbol{\eta}}_t - \boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t]'\right]\right| \\
&= \left|(\hat{\mathcal{K}}')^{-1}(1/T)\sum_{t=1}^{T}\left[[\hat{\boldsymbol{\eta}}_t - \hat{\mathcal{K}}'\boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t][\hat{\boldsymbol{\eta}}_t - \hat{\mathcal{K}}'\boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}\hat{\boldsymbol{\xi}}_t]'\right](\hat{\mathcal{K}})^{-1}\right| \\
&= |(\hat{\mathcal{K}}')^{-1}|\left|(1/T)\sum_{t=1}^{T}\left[[\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t][\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t]'\right]\right||(\hat{\mathcal{K}})^{-1}| \\
&= \left|(1/T)\sum_{t=1}^{T}\left[[\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t][\hat{\boldsymbol{\eta}}_t - \hat{\boldsymbol{\Pi}}\hat{\boldsymbol{\xi}}_t]'\right]\right| \div |\hat{\mathcal{K}}|^2,
\end{aligned} \qquad [20.2.32]$$

where

$$\hat{\boldsymbol{\Pi}} \equiv \hat{\mathcal{K}}'\boldsymbol{\zeta}_0(\hat{\mathcal{A}}')^{-1}. \qquad [20.2.33]$$

Recall that maximizing the concentrated log likelihood function for the actual data [20.2.21] is equivalent to choosing $\boldsymbol{\zeta}_0$ so as to minimize the expression in [20.2.32] subject to the requirement that $\boldsymbol{\zeta}_0$ can be written as $\mathbf{BA}'$ for some $(n \times h)$ matrices $\mathbf{B}$ and $\mathbf{A}$. But $\boldsymbol{\zeta}_0$ can be written in this form if and only if $\hat{\boldsymbol{\Pi}}$ in [20.2.33] can be written in the form $\boldsymbol{\beta}\boldsymbol{\gamma}'$ for some $(n \times h)$ matrices $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Thus, the task can be described as choosing $\hat{\boldsymbol{\Pi}}$ so as to minimize [20.2.32] subject to this condition. But this is precisely the problem solved in [20.2.28]—the solution is to use as regressors the first h elements of $\hat{\boldsymbol{\xi}}_t$. The value of [20.2.32] at the optimum is given by

$$\prod_{i=1}^{h}(1 - \hat{r}_i^2) \div |\hat{\mathcal{K}}|^2. \qquad [20.2.34]$$

Moreover, the matrix $\hat{\mathcal{K}}$ satisfies

$$\mathbf{I}_n = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t' = (1/T)\sum_{t=1}^{T}\hat{\mathcal{K}}'\hat{\mathbf{u}}_t\hat{\mathbf{u}}_t'\hat{\mathcal{K}} = \hat{\mathcal{K}}'\hat{\boldsymbol{\Sigma}}_{UU}\hat{\mathcal{K}}. \qquad [20.2.35]$$

Taking determinants of both sides of [20.2.35] establishes

$$1 = |\hat{\mathcal{K}}'|\,|\hat{\Sigma}_{UU}|\,|\hat{\mathcal{K}}|$$

or

$$1/|\hat{\mathcal{K}}|^2 = |\hat{\Sigma}_{UU}|.$$

Substituting this back into [20.2.34], it appears that the optimized value of [20.2.32] is equal to

$$|\hat{\Sigma}_{UU}| \times \prod_{i=1}^{h} (1 - \hat{r}_i^2).$$

Comparing [20.2.32] with [20.2.21], it follows that the maximum value achieved for the log likelihood function is given by

$$\mathcal{L}^* = \mathcal{N}(\hat{\zeta}_0) = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log\left\{|\hat{\Sigma}_{UU}| \times \prod_{i=1}^{h} (1 - \hat{r}_i^2)\right\},$$

as claimed in [20.2.10].

Motivation for Maximum Likelihood Estimates of Parameters

We have seen that the concentrated log likelihood function [20.2.21] is maximized by selecting as regressors the first h elements of $\hat{\boldsymbol{\xi}}_t$. Since $\hat{\boldsymbol{\xi}}_t = \hat{\mathcal{A}}'\hat{\mathbf{v}}_t$, this means using $\hat{\mathbf{A}}'\hat{\mathbf{v}}_t$ as regressors, where the $(n \times h)$ matrix $\hat{\mathbf{A}}$ denotes the first h columns of the $(n \times n)$ matrix $\hat{\mathcal{A}}$. Thus,

$$\zeta_0\hat{\mathbf{v}}_t = -\mathbf{B}\hat{\mathbf{A}}'\hat{\mathbf{v}}_t \qquad [20.2.36]$$

for some $(n \times h)$ matrix $\mathbf{B}$. This verifies the claim that $\hat{\mathbf{A}}$ is the maximum likelihood estimate of a basis for the space of cointegrating vectors.

Given that we want to choose $\hat{\mathbf{w}}_t \equiv \hat{\mathbf{A}}'\hat{\mathbf{v}}_t$ as regressors, the value of $\mathbf{B}$ for which the concentrated likelihood function will be maximized is obtained from *OLS* regressions of $\hat{\mathbf{u}}_t$ on $\hat{\mathbf{w}}_t$:

$$\hat{\mathbf{B}} = -\left[(1/T)\sum_{t=1}^{T}\hat{\mathbf{u}}_t\hat{\mathbf{w}}_t'\right]\left[(1/T)\sum_{t=1}^{T}\hat{\mathbf{w}}_t\hat{\mathbf{w}}_t'\right]^{-1}. \qquad [20.2.37]$$

But $\hat{\mathbf{w}}_t$ is composed of h canonical variates, meaning that

$$\left[(1/T)\sum_{t=1}^{T}\hat{\mathbf{w}}_t\hat{\mathbf{w}}_t'\right] = \mathbf{I}_h. \qquad [20.2.38]$$

Moreover,

$$\begin{aligned}\left[(1/T)\sum_{t=1}^{T}\hat{\mathbf{u}}_t\hat{\mathbf{w}}_t'\right] &= \left[(1/T)\sum_{t=1}^{T}\hat{\mathbf{u}}_t\hat{\mathbf{v}}_t'\hat{\mathbf{A}}\right] \\ &= \hat{\Sigma}_{UV}\hat{\mathbf{A}}.\end{aligned} \qquad [20.2.39]$$

Substituting [20.2.39] and [20.2.38] into [20.2.37],

$$\hat{\mathbf{B}} = -\hat{\Sigma}_{UV}\hat{\mathbf{A}},$$

and so, from [20.2.2], the maximum likelihood estimate of $\boldsymbol{\zeta}_0$ is given by

$$\hat{\boldsymbol{\zeta}}_0 = \hat{\Sigma}_{UV}\hat{\mathbf{A}}\hat{\mathbf{A}}'$$

as claimed in [20.2.12].

Expressions [20.2.17] and [20.2.18] gave the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}_i$ that maximized the likelihood function for any given value of $\boldsymbol{\zeta}_0$. Since the likelihood function is maximized with respect to $\boldsymbol{\zeta}_0$ by choosing $\hat{\boldsymbol{\zeta}}_0$ according to [20.2.12], it is maximized with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}_i$ by substituting $\hat{\boldsymbol{\zeta}}_0$ into [20.2.17] and [20.2.18], as claimed in [20.2.14] and [20.2.13]. Finally, substituting $\hat{\boldsymbol{\zeta}}_0$ into [20.2.20] verifies [20.2.15].

Maximum Likelihood Estimation in the Absence of Deterministic Time Trends

The preceding analysis assumed that $\boldsymbol{\alpha}$, the $(n \times 1)$ vector of constant terms in the *VAR*, was unrestricted. The value of $\boldsymbol{\alpha}$ contributes h constant terms for the h cointegrating relations, along with $g \equiv n - h$ deterministic time trends that are common to each of the n elements of $\mathbf{y}_t$. In some applications it might be of interest to allow constant terms in the cointegrating relations but to rule out deterministic time trends for any of the variables. We saw in equation [19.1.45] that this would require

$$\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\mu}_1^*, \qquad [20.2.40]$$

where $\mathbf{B}$ is the $(n \times h)$ matrix appearing in [20.2.2] while $\boldsymbol{\mu}_1^*$ is an $(h \times 1)$ vector corresponding to the unconditional mean of $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$. Thus, for this restricted case, we want to estimate only the h elements of $\boldsymbol{\mu}_1^*$ rather than all n elements of $\boldsymbol{\alpha}$.

To maximize the likelihood function subject to the restrictions that there are h cointegrating relations and no deterministic time trends in any of the series, Johansen's (1991) first step was to concentrate out $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots,$ and $\boldsymbol{\zeta}_{p-1}$ (but not $\boldsymbol{\alpha}$). For given $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}_0$, this is achieved by *OLS* regression of $(\Delta\mathbf{y}_t - \boldsymbol{\alpha} - \boldsymbol{\zeta}_0\mathbf{y}_{t-1})$ on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$. The residuals from this regression are related to the residuals from three separate regressions:

(1) A regression of $\Delta\mathbf{y}_t$ on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$ with no constant term,

$$\Delta\mathbf{y}_t = \tilde{\mathbf{\Pi}}_1\Delta\mathbf{y}_{t-1} + \tilde{\mathbf{\Pi}}_2\Delta\mathbf{y}_{t-2} + \cdots + \tilde{\mathbf{\Pi}}_{p-1}\Delta\mathbf{y}_{t-p+1} + \tilde{\mathbf{u}}_t; \qquad [20.2.41]$$

(2) A regression of a constant term on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$,

$$1 = \tilde{\boldsymbol{\omega}}_1'\Delta\mathbf{y}_{t-1} + \tilde{\boldsymbol{\omega}}_2'\Delta\mathbf{y}_{t-2} + \cdots + \tilde{\boldsymbol{\omega}}_{p-1}'\Delta\mathbf{y}_{t-p+1} + \tilde{w}_t; \qquad [20.2.42]$$

(3) A regression of $\mathbf{y}_{t-1}$ on $(\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1})$ with no constant term,

$$\mathbf{y}_{t-1} = \tilde{\aleph}_1\Delta\mathbf{y}_{t-1} + \tilde{\aleph}_2\Delta\mathbf{y}_{t-2} + \cdots + \tilde{\aleph}_{p-1}\Delta\mathbf{y}_{t-p+1} + \tilde{\mathbf{v}}_t. \qquad [20.2.43]$$

The concentrated log likelihood function is then

$$\begin{aligned}\tilde{\mathcal{M}}(\mathbf{\Omega}, \boldsymbol{\alpha}, \boldsymbol{\zeta}_0) = {} & -(Tn/2)\log(2\pi) - (T/2)\log|\mathbf{\Omega}| \\ & - (1/2)\sum_{t=1}^{T}[(\tilde{\mathbf{u}}_t - \boldsymbol{\alpha}\tilde{w}_t - \boldsymbol{\zeta}_0\tilde{\mathbf{v}}_t)'\mathbf{\Omega}^{-1}(\tilde{\mathbf{u}}_t - \boldsymbol{\alpha}\tilde{w}_t - \boldsymbol{\zeta}_0\tilde{\mathbf{v}}_t)].\end{aligned}$$

Further concentrating out $\mathbf{\Omega}$ results in

$$\begin{aligned}&\tilde{\mathcal{N}}(\boldsymbol{\alpha}, \boldsymbol{\zeta}_0) \\ &= -(Tn/2)\log(2\pi) - (Tn/2) \qquad [20.2.44] \\ &\quad - (T/2)\log\left|\sum_{t=1}^{T}(1/T)\left\{(\tilde{\mathbf{u}}_t - \boldsymbol{\alpha}\tilde{w}_t - \boldsymbol{\zeta}_0\tilde{\mathbf{v}}_t)(\tilde{\mathbf{u}}_t - \boldsymbol{\alpha}\tilde{w}_t - \boldsymbol{\zeta}_0\tilde{\mathbf{v}}_t)'\right\}\right|.\end{aligned}$$

Imposing the constraints $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\mu}_1^*$ and $\boldsymbol{\zeta}_0 = -\mathbf{B}\mathbf{A}'$, the magnitude in [20.2.44] can be written

$$\bar{\mathcal{N}}(\boldsymbol{\alpha}, \boldsymbol{\zeta}_0) = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log\left|\sum_{t=1}^{T}(1/T)\{(\tilde{\mathbf{u}}_t + \mathbf{B}\tilde{\mathbf{A}}'\tilde{\mathbf{w}}_t)(\tilde{\mathbf{u}}_t + \mathbf{B}\tilde{\mathbf{A}}'\tilde{\mathbf{w}}_t)'\}\right|, \qquad [20.2.45]$$

where

$$\underset{(n+1)\times 1}{\tilde{\mathbf{w}}_t} \equiv \begin{bmatrix} \hat{w}_t \\ \tilde{\mathbf{v}}_t \end{bmatrix}$$

$$\underset{h\times(n+1)}{\tilde{\mathbf{A}}'} \equiv [-\boldsymbol{\mu}_1^* \quad \mathbf{A}']. \qquad [20.2.46]$$

But setting $\boldsymbol{\zeta}_0 = -\mathbf{BA}'$ in [20.2.21] produces an expression of exactly the same form as [20.2.45], with $\mathbf{A}$ in [20.2.21] replaced by $\tilde{\mathbf{A}}$ and $\hat{\mathbf{v}}_t$ replaced by $\tilde{\mathbf{w}}_t$. Thus, the restricted log likelihood is maximized simply by replacing $\hat{\mathbf{v}}_t$ in the analysis of [20.2.21] with $\tilde{\mathbf{w}}_t$.

To summarize, construct

$$\tilde{\Sigma}_{\mathrm{WW}} = (1/T)\sum_{t=1}^{T}\tilde{\mathbf{w}}_t\tilde{\mathbf{w}}_t'$$

$$\tilde{\Sigma}_{\mathrm{UU}} = (1/T)\sum_{t=1}^{T}\tilde{\mathbf{u}}_t\tilde{\mathbf{u}}_t'$$

$$\tilde{\Sigma}_{\mathrm{UW}} = (1/T)\sum_{t=1}^{T}\tilde{\mathbf{u}}_t\tilde{\mathbf{w}}_t'$$

and find the eigenvalues of the $(n + 1) \times (n + 1)$ matrix

$$\tilde{\Sigma}_{\mathrm{WW}}^{-1}\tilde{\Sigma}_{\mathrm{WU}}\hat{\Sigma}_{\mathrm{UU}}^{-1}\hat{\Sigma}_{\mathrm{UW}}, \qquad [20.2.47]$$

ordered $\tilde{\lambda}_1 > \tilde{\lambda}_2 > \cdots > \tilde{\lambda}_{n+1}$. The maximum value achieved for the log likelihood function subject to the constraint that there are h cointegrating relations and no deterministic time trends is

$$\tilde{\mathcal{L}}_h = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\tilde{\Sigma}_{\mathrm{UU}}| - (T/2)\sum_{i=1}^{h}\log(1 - \tilde{\lambda}_i). \qquad [20.2.48]$$

Let $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \ldots, \tilde{\mathbf{a}}_{n+1}$ denote the eigenvectors of [20.2.47] normalized by $\tilde{\mathbf{a}}_i'\tilde{\Sigma}_{\mathrm{WW}}\tilde{\mathbf{a}}_i = 1$. Then the maximum likelihood estimate of $\tilde{\mathbf{A}}$ is given by the matrix $[\tilde{\mathbf{a}}_1 \; \tilde{\mathbf{a}}_2 \; \cdots \; \tilde{\mathbf{a}}_h]$. The maximum likelihood estimate of $\mathbf{B}\tilde{\mathbf{A}}'$ is

$$\tilde{\mathbf{B}}\tilde{\mathbf{A}}' = -\tilde{\Sigma}_{\mathrm{UW}}\tilde{\mathbf{A}}\tilde{\mathbf{A}}'. \qquad [20.2.49]$$

Recall from [20.2.46] that

$$\begin{aligned}\mathbf{B}\tilde{\mathbf{A}}' &= [-\mathbf{B}\boldsymbol{\mu}_1^* \quad \mathbf{BA}'] \\ &= [-\boldsymbol{\alpha} \quad -\boldsymbol{\zeta}_0].\end{aligned} \qquad [20.2.50]$$

Thus, [20.2.49] implies that the maximum likelihood estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}_0$ are given by

$$[\tilde{\boldsymbol{\alpha}} \quad \tilde{\boldsymbol{\zeta}}_0] = \tilde{\Sigma}_{\mathrm{UW}}\tilde{\mathbf{A}}\tilde{\mathbf{A}}'.$$

The *MLE* of $\boldsymbol{\zeta}_i$ is

$$\tilde{\boldsymbol{\zeta}}_i = \tilde{\boldsymbol{\Pi}}_i - \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\omega}}_i' - \tilde{\boldsymbol{\zeta}}_0\tilde{\boldsymbol{\aleph}}_i \qquad \text{for } i = 1, 2, \ldots, p - 1,$$

while the MLE of $\mathbf{\Omega}$ is

$$\bar{\mathbf{\Omega}} = (1/T) \sum_{t=1}^{T} [(\tilde{\mathbf{u}}_t - \tilde{\boldsymbol{\alpha}}\tilde{w}_t - \tilde{\boldsymbol{\zeta}}_0\tilde{\mathbf{v}}_t)(\tilde{\mathbf{u}}_t - \tilde{\boldsymbol{\alpha}}\tilde{w}_t - \tilde{\boldsymbol{\zeta}}_0\tilde{\mathbf{v}}_t)'].$$

20.3. *Hypothesis Testing*

We saw in the previous chapter that tests of the null hypothesis of no cointegration typically involve nonstandard asymptotic distributions, while tests about the value of the cointegrating vector under the maintained hypothesis that cointegration is present will have asymptotic χ^2 distributions, provided that suitable allowance is made for the serial correlation in the data. These results generalize to *FIML* analysis. The asymptotic distribution of a test of the number of cointegrating relations is nonstandard, but tests about the cointegrating vector are often χ^2.

Testing the Null Hypothesis of h *Cointegrating Relations*

Suppose that an $(n \times 1)$ vector $\mathbf{y}_t$ can be characterized by a $VAR(p)$ in levels, which we write in the form of [20.2.1]:

$$\Delta\mathbf{y}_t = \boldsymbol{\zeta}_1\Delta\mathbf{y}_{t-1} + \boldsymbol{\zeta}_2\Delta\mathbf{y}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \boldsymbol{\zeta}_0\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \qquad [20.3.1]$$

Under the null hypothesis H_0 that there are exactly h cointegrating relations among the elements of $\mathbf{y}_t$, this VAR is restricted by the requirement that $\boldsymbol{\zeta}_0$ can be written in the form $\boldsymbol{\zeta}_0 = -\mathbf{BA}'$, for $\mathbf{B}$ an $(n \times h)$ matrix and $\mathbf{A}'$ an $(h \times n)$ matrix. Another way of describing this restriction is that only h linear combinations of the levels of $\mathbf{y}_{t-1}$ can be used in the regressions in [20.3.1]. The largest value that can be achieved for the log likelihood function under this constraint was given by [20.2.10]:

$$\begin{aligned}\mathcal{L}_0^* = &-(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\boldsymbol{\Sigma}}_{UU}| \\ &- (T/2)\sum_{i=1}^{h}\log(1 - \hat{\lambda}_i).\end{aligned} \qquad [20.3.2]$$

Consider the alternative hypothesis H_A that there are n cointegrating relations, where n is the number of elements of $\mathbf{y}_t$. This amounts to the claim that *every* linear combination of $\mathbf{y}_t$ is stationary, in which case $\mathbf{y}_{t-1}$ would appear in [20.3.1] without constraints and no restrictions are imposed on $\boldsymbol{\zeta}_0$. The value for the log likelihood function in the absence of constraints is given by

$$\begin{aligned}\mathcal{L}_A^* = &-(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\boldsymbol{\Sigma}}_{UU}| \\ &- (T/2)\sum_{i=1}^{n}\log(1 - \hat{\lambda}_i).\end{aligned} \qquad [20.3.3]$$

A likelihood ratio test of H_0 against H_A can be based on

$$\mathcal{L}_A^* - \mathcal{L}_0^* = -(T/2)\sum_{i=h+1}^{n}\log(1 - \hat{\lambda}_i).$$

If the hypothesis involved just $I(0)$ variables, we would expect twice the log likelihood ratio,

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T\sum_{i=h+1}^{n}\log(1 - \hat{\lambda}_i), \qquad [20.3.4]$$

to be asymptotically distributed as χ^2. In the case of H_0, however, the hypothesis involves the coefficient on $\mathbf{y}_{t-1}$, which, from the Stock-Watson common trends representation, depends on the value of $g \equiv (n - h)$ separate random walks. Let $\mathbf{W}(r)$ be g-dimensional standard Brownian motion. Suppose that the true value of the constant term $\boldsymbol{\alpha}$ in [20.3.1] is zero, meaning that there is no intercept in any of the cointegrating relations and no deterministic time trend in any of the elements of $\mathbf{y}_t$. Suppose further that no constant term is included in the auxiliary regressions [20.2.4] and [20.2.5] that were used to construct $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$. Johansen (1988) showed that under these conditions the asymptotic distribution of the statistic in [20.3.4] is the same as that of the trace of the following matrix:

$$\mathbf{Q} \equiv \left[\int_0^1 \mathbf{W}(r)\, d\mathbf{W}(r)'\right]' \left[\int_0^1 \mathbf{W}(r)\mathbf{W}(r)'\, dr\right]^{-1} \left[\int_0^1 \mathbf{W}(r)\, d\mathbf{W}(r)'\right]. \qquad [20.3.5]$$

Percentiles for the trace of the matrix in [20.3.5] are reported in the case 1 portion of Table B.10. These are based on Monte Carlo simulations.

If the number of cointegrating relations (h) is 1 less than the number of variables (n), then $g = 1$ and [20.3.5] describes the following scalar:

$$Q = \frac{\left\{\int_0^1 W(r)\, dW(r)\right\}^2}{\left\{\int_0^1 [W(r)]^2\, dr\right\}} = \frac{(1/2)^2\left\{[W(1)]^2 - 1\right\}^2}{\left\{\int_0^1 [W(r)]^2\, dr\right\}}, \qquad [20.3.6]$$

where the second equality follows from [18.1.15]. Expression [20.3.6] will be recognized as the square of the statistic [17.4.12] that described the asymptotic distribution of the Dickey-Fuller test based on the *OLS t* statistic. For example, if we are considering an autoregression involving a single variable ($n = 1$), the null hypothesis of no cointegrating relations ($h = 0$) amounts to the claim that $\zeta_0 = 0$ in [20.3.1] or that Δy_t follows an $AR(p - 1)$ process. Thus, Johansen's procedure provides an alternative approach to testing for unit roots in univariate series, an idea explored further in Exercise 20.4.

Another approach would be to test the null hypothesis of h cointegrating relations against the alternative of $h + 1$ cointegrating relations. Twice the log likelihood ratio for this case is given by

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T \log(1 - \hat{\lambda}_{h+1}). \qquad [20.3.7]$$

Again, under the assumption that the true value of $\boldsymbol{\alpha} = \mathbf{0}$ and that no constant term is included in [20.2.4] or [20.2.5], the asymptotic distribution of the statistic in [20.3.7] is the same as that of the largest eigenvalue of the matrix $\mathbf{Q}$ defined in [20.3.5]. Monte Carlo estimates of this distribution are reported in the case 1 section of Table B.11.

Note that if $g = 1$, then $n = h + 1$. In this case the statistics [20.3.4] and [20.3.7] are identical. For this reason, the first row in Table B.10 is the same as the first row of Table B.11.

Typically, the cointegrating relations could include nonzero intercepts, in which case we would want to include constants in the auxiliary regressions [20.2.4] and [20.2.5]. As one might guess from the analysis in Chapter 18, the asymptotic distribution in this case depends on whether or not any of the series exhibit deterministic time trends. Suppose that the true value of $\boldsymbol{\alpha}$ is such that there are no deterministic trends in any of the series, so that the true $\boldsymbol{\alpha}$ satisfies $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\mu}_1^*$ as in [20.2.40]. Assuming that no restrictions are imposed on the constant term in the estimation of the auxiliary regressions [20.2.4] and [20.2.5], then the asymptotic distribution of [20.3.4] is given in the case 2 section of Table B.10, while the asymptotic distribution of [20.3.7] is given in the case 2 panel of Table B.11. By contrast, if any of the variables exhibit deterministic time trends (one or more elements of $\boldsymbol{\alpha} - \mathbf{B}\boldsymbol{\mu}_1^*$ are nonzero), then the asymptotic distribution of [20.3.4] is that of the variable in the case 3 section of Table B.10, while the asymptotic distribution of [20.3.7] is given in the case 3 section of Table B.11.

When $g = 1$ and $\boldsymbol{\alpha} \neq \mathbf{B}\boldsymbol{\mu}_1^*$, the single random walk that is common to $\mathbf{y}_t$ is dominated by a deterministic time trend. In this situation, Johansen and Juselius (1990, p. 180) noted that the case 3 analog of [20.3.6] has a $\chi^2(1)$ distribution, for reasons similar to those noted by West (1988) and discussed in Chapter 18. The modest differences between the first row of the case 3 part of Table B.10 or B.11 and the first row of Table B.2 are presumably due to sampling error implicit in the Monte Carlo procedure used to generate the values in Tables B.10 and B.11.

Application to Exchange Rate Data

Consider for illustration the monthly data for Italy and the United States plotted in Figure 19.2. The systems of equations in [20.2.4] and [20.2.5] were estimated by *OLS* for $\mathbf{y}_t = (p_t, s_t, p_t^*)'$, where p_t is 100 times the log of the U.S. price level, s_t is 100 times the log of the dollar-lira exchange rate, and p_t^* is 100 times the log of the Italian price level. The regressions were estimated over t = 1974:2 through 1989:10 (so that the number of observations used for estimation was $T = 189$); $p = 12$ lags were assumed for the *VAR* in levels.

The sample variance-covariance matrices for the residuals $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$ were calculated from [20.2.6] through [20.2.8] to be

$$\hat{\Sigma}_{UU} = \begin{bmatrix} 0.0435114 & -0.0316283 & 0.0154297 \\ -0.0316283 & 4.68650 & 0.0319877 \\ 0.0154297 & 0.0319877 & 0.179927 \end{bmatrix}$$

$$\hat{\Sigma}_{VV} = \begin{bmatrix} 427.366 & -370.699 & 805.812 \\ -370.699 & 424.083 & -709.036 \\ 805.812 & -709.036 & 1525.45 \end{bmatrix}$$

$$\hat{\Sigma}_{UV} = \begin{bmatrix} -0.484857 & 0.498758 & -0.837701 \\ -1.81401 & -2.95927 & -2.46896 \\ -1.80836 & 1.46897 & -3.58991 \end{bmatrix}.$$

The eigenvalues of the matrix in [20.2.9] are then[4]

$$\hat{\lambda}_1 = 0.1105$$
$$\hat{\lambda}_2 = 0.05603$$
$$\hat{\lambda}_3 = 0.03039$$

with

$$T\log(1 - \hat{\lambda}_1) = -22.12$$
$$T\log(1 - \hat{\lambda}_2) = -10.90$$
$$T\log(1 - \hat{\lambda}_3) = -5.83.$$

[4]Calculations were based on more significant digits than reported, and so the reader may find slight discrepancies in trying to reproduce these results from the figures reported.

The likelihood ratio test of the null hypothesis of $h = 0$ cointegrating relations against the alternative of $h = 3$ cointegrating relations is then calculated from [20.3.4] to be

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = 22.12 + 10.90 + 5.83 = 38.85. \qquad [20.3.8]$$

Here the number of unit roots under the null hypothesis is $g = n - h = 3$. Given the evidence of deterministic time trends, the magnitude in [20.3.8] is to be compared with the case 3 section of Table B.10. Since $38.85 > 29.5$, the null hypothesis of no cointegration is rejected at the 5% level. Similarly, the likelihood ratio test [20.3.7] of the null hypothesis of no cointegrating relations ($h = 0$) against the alternative of a single cointegrating relation ($h = 1$) is given by 22.12. Comparing this with the case 3 section of Table B.11, we see that $22.12 > 20.8$, so that the null hypothesis of no cointegration is also rejected by this test.

This differs from the conclusion of the Phillips-Ouliaris test for no cointegration between these series, on the basis of which the null hypothesis of no cointegration for these variables was found to be accepted in Chapter 19.

Searching for evidence of a possible second cointegrating relation, consider the likelihood ratio test of the null hypothesis of $h = 1$ cointegrating relation against the alternative of $h = 3$ cointegrating relations:

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = 10.90 + 5.83 = 16.73.$$

For this test, $g = 2$. Since $16.73 > 15.2$, the null hypothesis of a single cointegrating relation is rejected at the 5% level. The likelihood ratio test of the null hypothesis of $h = 1$ cointegrating relation against the alternative of $h = 2$ relations is $10.90 < 14.0$; hence, the two tests offer conflicting evidence as to the presence of a second cointegrating relation.

The eigenvector $\hat{\mathbf{a}}_1$ of the matrix in [20.2.9] associated with $\hat{\lambda}_1$, normalized so that $\hat{\mathbf{a}}_1'\hat{\Sigma}_{VV}\hat{\mathbf{a}}_1 = 1$, is given by

$$\hat{\mathbf{a}}_1' = [-0.7579 \quad 0.02801 \quad 0.4220]. \qquad [20.3.9]$$

It is natural to renormalize this by taking the first element to be unity:

$$\tilde{\mathbf{a}}_1' = [1.00 \quad -0.04 \quad -0.56].$$

This is virtually identical to the estimate of the cointegrating vector based on *OLS* from [19.2.49].

Likelihood Ratio Tests About the Cointegrating Vector

Consider a system of n variables that is assumed (under both the null and the alternative) to be characterized by h cointegrating relations. We might then want to test a restriction on these cointegrating vectors, such as that only q of the variables are involved in the cointegrating relations. For example, we might be interested in whether the middle coefficient in [20.3.9] is zero, that is, in whether the cointegrating relation involves solely the U.S. and Italian price levels. For this example $h = 1$, $q = 2$, and $n = 3$. In general it must be the case that $h \leq q \leq n$. Since h linear combinations of the q variables included in the cointegrating relations are stationary, if $q = h$, then all q of the included variables would have to be stationary in levels. If $q = n$, then the null hypothesis places no restrictions on the cointegrating relations.

Consider the general restriction that there is a known $(q \times n)$ matrix $\mathbf{D}'$ such that the cointegrating relations involve only $\mathbf{D}'\mathbf{y}_t$. For the preceding example,

$$\mathbf{D}' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad [20.3.10]$$

Hence, the error-correction term in [20.3.1] will take the form

$$\zeta_0 \mathbf{y}_{t-1} = -\mathbf{B}\mathbf{A}'\mathbf{D}'\mathbf{y}_{t-1},$$

where $\mathbf{B}$ is now an $(n \times h)$ matrix and $\mathbf{A}'$ is an $(h \times q)$ matrix. Maximum likelihood estimation proceeds exactly as in the previous section, where $\hat{\mathbf{v}}_t$ in [20.2.5] is replaced by the *OLS* residuals from regressions of $\mathbf{D}'\mathbf{y}_{t-1}$ on a constant and $\Delta\mathbf{y}_{t-1}$, $\Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$. This is equivalent to replacing $\hat{\boldsymbol{\Sigma}}_{\mathbf{VV}}$ in [20.2.6] and $\hat{\boldsymbol{\Sigma}}_{\mathbf{UV}}$ in [20.2.8] with

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{VV}} \equiv \mathbf{D}'\hat{\boldsymbol{\Sigma}}_{\mathbf{VV}}\mathbf{D} \qquad [20.3.11]$$

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{UV}} \equiv \hat{\boldsymbol{\Sigma}}_{\mathbf{UV}}\mathbf{D}. \qquad [20.3.12]$$

Let $\tilde{\lambda}_i$ denote the ith largest eigenvalue of

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{VV}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\mathbf{VU}}\hat{\boldsymbol{\Sigma}}_{\mathbf{UU}}^{-1}\tilde{\boldsymbol{\Sigma}}_{\mathbf{UV}}. \qquad [20.3.13]$$

The maximized value for the restricted log likelihood is then

$$\mathcal{L}_0^* = -(Tn/2)\log(2\pi) - (Tn/2) - (T/2)\log|\hat{\boldsymbol{\Sigma}}_{\mathbf{UU}}| - (T/2)\sum_{i=1}^{h}\log(1 - \tilde{\lambda}_i).$$

A likelihood ratio test of the null hypothesis that the h cointegrating relations only involve $\mathbf{D}'\mathbf{y}_t$ against the alternative hypothesis that the h cointegrating relations could involve any elements of $\mathbf{y}_t$ would then be

$$2(\mathcal{L}_A^* - \mathcal{L}_0^*) = -T\sum_{i=1}^{h}\log(1 - \hat{\lambda}_i) + T\sum_{i=1}^{h}\log(1 - \tilde{\lambda}_i). \qquad [20.3.14]$$

In this case, the null hypothesis involves only coefficients on $I(0)$ variables (the error-correction terms $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$), and standard asymptotic distribution theory turns out to apply. Johansen (1988, 1991) showed that the likelihood ratio statistic [20.3.14] has an asymptotic χ^2 distribution with $h \cdot (n - q)$ degrees of freedom.

For illustration, consider the restriction represented by [20.3.10] that the exchange rate has a coefficient of zero in the cointegrating vector [20.3.9]. From [20.3.11] and [20.3.12], we calculate

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{VV}} = \begin{bmatrix} 427.366 & 805.812 \\ 805.812 & 1525.45 \end{bmatrix}$$

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{UV}} = \begin{bmatrix} -0.484857 & -0.837701 \\ -1.81401 & -2.46896 \\ -1.80836 & -3.58991 \end{bmatrix}.$$

The eigenvalues for the matrix in [20.3.13] are then

$$\tilde{\lambda}_1 = 0.1059 \qquad \tilde{\lambda}_2 = 0.04681,$$

with

$$T\log(1 - \tilde{\lambda}_1) = -21.15 \qquad T\log(1 - \tilde{\lambda}_2) = -9.06.$$

The likelihood ratio statistic [20.3.14] is

$$2(\mathscr{L}_A^* - \mathscr{L}_0^*) = 22.12 - 21.15$$
$$= 0.97.$$

The degrees of freedom for this statistic are

$$h \cdot (n - q) = 1 \cdot (3 - 2) = 1;$$

the null hypothesis imposes a single restriction on the cointegrating vector. The 5% critical value for a $\chi^2(1)$ variable is seen from Table B.2 to be 3.84. Since $0.97 < 3.84$, the null hypothesis that the exchange rate does not appear in the cointegrating relation is accepted. The restricted cointegrating vector (normalized with the coefficient on the U.S. price level to be unity) is

$$\tilde{\mathbf{a}}_1' = [1.00 \quad 0.00 \quad -0.54].$$

As a second example, consider the hypothesis that originally suggested interest in a possible cointegrating relation between these three variables. This was the hypothesis that the real exchange rate is stationary, or that the cointegrating vector is proportional to $(1, -1, -1)'$. For this hypothesis, $\mathbf{D}' = (1, -1, -1)$ and

$$\hat{\Sigma}_{VV} = 88.5977$$

$$\hat{\Sigma}_{UV} = \begin{bmatrix} -0.145914 \\ 3.61422 \\ 0.312582 \end{bmatrix}.$$

In this case, the matrix [20.3.13] is the scalar 0.0424498, and so $\tilde{\lambda}_1 = 0.0424498$ and $T \log(1 - \tilde{\lambda}_1) = -8.20$. Thus, the likelihood ratio test of the null hypothesis that the cointegrating vector is proportional to $(1, -1, -1)'$ is

$$2(\mathscr{L}_A^* - \mathscr{L}_0^*) = 22.12 - 8.20$$
$$= 13.92.$$

In this case, the degrees of freedom are

$$h \cdot (n - q) = 1 \cdot (3 - 1) = 2.$$

The 5% critical value for a $\chi^2(2)$ variable is 5.99. Since $13.92 > 5.99$, the null hypothesis that the cointegrating vector is proportional to $(1, -1, -1)'$ is rejected.

Other Hypothesis Tests

A number of other hypotheses can be tested in this framework. For example, Johansen (1991) showed that the null hypothesis that there are no deterministic time trends in any of the series can be tested by taking twice the difference between [20.2.10] and [20.2.48]. Under the null hypothesis, this likelihood ratio statistic is asymptotically χ^2 with $g = n - h$ degrees of freedom. Johansen also discussed construction of Wald-type tests of hypotheses involving the cointegrating vectors.

Not all hypothesis tests about the coefficients in Johansen's framework are asymptotically χ^2. Consider an error-correction *VAR* of the form of [20.2.1] where $\zeta_0 = -\mathbf{BA}'$. Suppose we are interested in the null hypothesis that the last n_3 elements of $\mathbf{y}_t$ fail to Granger-cause the first n_1 elements of $\mathbf{y}_t$. Toda and Phillips (forthcoming) showed that a Wald test of this null hypothesis can have a nonstandard distribution. See Mosconi and Giannini (1992) for further discussion.

Comparison Between FIML *and Other Approaches*

Johansen's *FIML* estimation represents the short-run dynamics of a system in terms of a vector autoregression in differences with the error-correction vector $\mathbf{z}_{t-1}$ added. Short-run dynamics can also be modeled with what are sometimes called nonparametric methods, such as the Bartlett window used to construct the fully modified Phillips-Hansen (1990) estimator in equation [19.3.53]. Related nonparametric estimators have been proposed by Phillips (1990, 1991a), Park (1992), and Park and Ogaki (1991). Park (1990) established the asymptotic equivalence of the parametric and nonparametric approaches, and Phillips (1991a) discussed the sense in which any *FIML* estimator is asymptotically efficient. Johansen (1992) provided a further discussion of the relation between limited-information and full-information estimation strategies.

In practice, the parametric and nonparametric approaches differ not just in their treatment of short-run dynamics but also in the normalizations employed. The fact that Johansen's method seeks to estimate the *space* of cointegrating relations rather than a particular set of coefficients can be both an asset and a liability. It is an asset if the researcher has no prior information about which variables appear in the cointegrating relations and is concerned about inadvertently normalizing $a_{11} = 1$ when the true value of $a_{11} = 0$. On the other hand, Phillips (1991b) has stressed that if the researcher wants to make structural interpretations of the separate cointegrating relations, this logically requires imposing further restrictions on the matrix $\mathbf{A}'$.

For example, let r_t denote the nominal interest rate on 3-month corporate debt, i_t the nominal interest rate on 3-month government debt, and π_t the 3-month inflation rate. Suppose that these three variables appear to be $I(1)$ and exhibit two cointegrating relations. A natural view is that these cointegrating relations represent two stabilizing relations. The first reflects forces that keep the risk premium stationary, so that

$$r_t = \mu_{11}^* + \gamma_1 i_t + z_{1t}^*, \qquad [20.3.15]$$

with $z_{1t}^* \sim I(0)$. A second force is the Fisher effect, which tends to keep the real interest rate stationary:

$$\pi_t = \mu_{21}^* + \gamma_2 i_t + z_{2t}^*, \qquad [20.3.16]$$

with $z_{2t}^* \sim I(0)$. The system of [20.3.15] and [20.3.16] will be recognized as an example of Phillips's (1991a) triangular representation [19.1.20] for the vector $\mathbf{y}_t = (r_t, \pi_t, i_t)'$. Thus, in this example theoretical considerations suggest a natural ordering of variables for which the normalization used by Phillips would be of particular interest for structural inference—the coefficients μ_{11}^* and γ_1 tell us about the risk premium, and the coefficients μ_{21}^* and γ_2 tell us about the Fisher effect.

20.4. *Overview of Unit Roots—To Difference or Not to Difference?*

The preceding chapters have explored a number of issues in the statistical analysis of unit roots. This section attempts to summarize what all this means in practice.

Consider a vector of variables $\mathbf{y}_t$ whose dynamics we would like to describe and some of whose elements may be nonstationary. For concreteness, let us assume that the goal is to characterize these dynamics in terms of a vector autoregression.

One option is to ignore the nonstationarity altogether and simply estimate the *VAR* in levels, relying on standard t and F distributions for testing any hypotheses. This strategy has the following features to recommend it. (1) The parameters that describe the system's dynamics are estimated consistently. (2) Even if the true model is a VAR in differences, certain functions of the parameters and hypothesis tests based on a VAR in levels have the same asymptotic distribution as would estimates based on differenced data. (3) A Bayesian motivation can be given for the usual t or F distributions for test statistics even when the classical asymptotic theory for these statistics is nonstandard.

A second option is routinely to difference any apparently nonstationary variables before estimating the VAR. If the true process is a VAR in differences, then differencing should improve the small-sample performance of all of the estimates and eliminate altogether the nonstandad asymptotic distributions associated with certain hypothesis tests. The drawback to this approach is that the true process may not be a VAR in differences. Some of the series may in fact have been stationary, or perhaps some linear combinations of the series are stationary, as in a cointegrated VAR. In such circumstances a VAR in differenced form is misspecified.

Yet a third approach is to investigate carefully the nature of the nonstationarity, testing each series individually for unit roots and then testing for possible cointegration among the series. Once the nature of the nonstationarity is understood, a stationary representation for the system can be estimated. For example, suppose that in a four-variable system we determine that the first variable y_{1t} is stationary while the other variables (y_{2t}, y_{3t}, and y_{4t}) are each individually $I(1)$. Suppose we further conclude that y_{2t}, y_{3t}, and y_{4t} are characterized by a single cointegrating relation. For $\mathbf{y}_{2t} \equiv (y_{2t}, y_{3t}, y_{4t})'$, this implies a vector error-correction representation of the form

$$\begin{bmatrix} y_{1t} \\ \Delta \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} + \begin{bmatrix} \zeta_{11}^{(1)} & \zeta_{12}^{(1)} \\ \zeta_{21}^{(1)} & \zeta_{22}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ \Delta \mathbf{y}_{2,t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{11}^{(2)} & \zeta_{12}^{(2)} \\ \zeta_{21}^{(2)} & \zeta_{22}^{(2)} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ \Delta \mathbf{y}_{2,t-2} \end{bmatrix} + \cdots$$
$$+ \begin{bmatrix} \zeta_{11}^{(p-1)} & \zeta_{12}^{(p-1)} \\ \zeta_{21}^{(p-1)} & \zeta_{22}^{(p-1)} \end{bmatrix} \begin{bmatrix} y_{1,t-p+1} \\ \Delta \mathbf{y}_{2,t-p+1} \end{bmatrix} + \begin{bmatrix} \zeta_1^{(0)} \\ \zeta_2^{(0)} \end{bmatrix} \mathbf{y}_{2,t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix},$$

where the (4×3) matrix $\begin{bmatrix} \zeta_1^{(0)} \\ \zeta_2^{(0)} \end{bmatrix}$ is restricted to be of the form $\mathbf{ba}'$ where $\mathbf{b}$ is (4×1) and $\mathbf{a}'$ is (1×3). Such a system can then be estimated by adapting the methods described in Section 20.2, and most hypothesis tests on this system should be asymptotically χ^2.

The disadvantage of the third approach is that, despite the care one exercises, the restrictions imposed may still be invalid—the investigator may have accepted a null hypothesis even though it is false, or rejected a null hypothesis that is actually true. Moreover, alternative tests for unit roots and cointegration can produce conflicting results, and the investigator may be unsure as to which should be followed.

Experts differ in the advice offered for applied work. One practical solution is to employ parts of all three approaches. This eclectic strategy would begin by estimating the VAR in levels without restrictions. The next step is to make a quick assessment as to which series are likely nonstationary. This assessment could be based on graphs of the data, prior information about the series and their likely cointegrating relations, or any of the more formal tests discussed in Chapter 17. Any nonstationary series can then be differenced or expressed in error-correction form and a stationary VAR could then be estimated. For example, to estimate a VAR that includes the log of income (y_t) and the log of consumption (c_t), these two variables might be included in a stationary VAR as Δy_t and $(c_t - y_t)$. If the VAR for the data in levels form yields similar inferences to those for the VAR in stationary form, then the researcher might be satisfied that the results were not governed by the assumptions made about unit roots. If the answers differ, then some attempt to reconcile the results should be made. Careful efforts along the lines of the third strategy described in this section might convince the investigator that the stationary formulation was misspecified, or alternatively that the levels results can be explained by the appropriate asymptotic theory. A nice example of how asymptotic theory could be used to reconcile conflicting findings was provided by Stock and Watson (1989). Alternatively, Christiano and Ljungqvist (1988) proposed simulating data from the estimated levels model, and seeing whether incorrectly fitting such simulated data with the stationary specification would spuriously produce the results found when the stationary specification was fitted to the actual data. Similarly, data could be simulated from the stationary model to see if it could account for the finding of the levels specification. If we find that a single specification can account for both the levels and the stationary results, then our confidence in that specification increases.

APPENDIX 20.A. *Proof of Chapter 20 Proposition*

■ **Proof of Proposition 20.1.**

(a) First we show that $\lambda_i < 1$ for $i = 1, 2, \ldots, n_1$. Any eigenvalue λ of [20.1.8] satisfies

$$|\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} - \lambda \mathbf{I}_{n_1}| = 0.$$

Since Σ_{YY} is positive definite, this will be true if and only if

$$|\lambda\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}| = 0. \qquad [20.A.1]$$

But from the triangular factorization of Σ in equation [4.5.26], the matrix

$$\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \qquad [20.A.2]$$

is positive definite. Hence, the determinant in [20.A.1] could not be zero at $\lambda = 1$. Note further that

$$\lambda\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} = (\lambda - 1)\Sigma_{YY} + [\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}]. \qquad [20.A.3]$$

If $\lambda > 1$, then the right side of expression [20.A.3] would be the sum of two positive definite matrices and so would be positive definite. The left side of [20.A.3] would then be positive definite, implying that the determinant in [20.A.1] could not be zero for $\lambda > 1$. Hence, $\lambda \geq 1$ is not consistent with [20.A.1].

To see that $\lambda_i \geq 0$, notice that if λ were less than zero, then $\lambda\Sigma_{YY}$ would be a negative number times a positive definite matrix so that $\lambda\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$ would also be a negative number times a positive definite matrix. Hence, the determinant in [20.A.1] could not be zero for any value of $\lambda < 0$.

Parallel arguments establish that $0 \leq \mu_j < 1$ for $j = 1, 2, \ldots, n_2$.

(b) Let $\mathbf{k}_i$ be an eigenvector associated with a nonzero eigenvalue λ_i of [20.1.8]:

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_i = \lambda_i\mathbf{k}_i. \qquad [20.A.4]$$

Premultiplying both sides of [20.A.4] by Σ_{XY} results in

$$[\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}][\Sigma_{XY}\mathbf{k}_i] = \lambda_i[\Sigma_{XY}\mathbf{k}_i]. \qquad [20.A.5]$$

But $[\Sigma_{XY}\mathbf{k}_i]$ cannot be zero, for if $[\Sigma_{XY}\mathbf{k}_i]$ did equal zero, then the left side of [20.A.4] would be zero, implying that $\lambda_i = 0$. Thus, [20.A.5] implies that λ_i is also an eigenvalue of the matrix $[\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}]$ associated with the eigenvector $[\Sigma_{XY}\mathbf{k}_i]$. Recall further that eigenvalues are unchanged by transposition of a matrix:

$$[\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}]' = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX},$$

which is the matrix [20.1.12]. This proves that if λ_i is a nonzero eigenvalue of [20.1.8], then it is also an eigenvalue of [20.1.12]. Exactly parallel calculations show that if μ_i is a nonzero eigenvalue of [20.1.12], then it is also an eigenvalue of [20.1.8].

(c) Premultiply [20.1.10] by $\mathbf{k}_j'\Sigma_{YY}$:

$$\mathbf{k}_j'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_i = \lambda_i\mathbf{k}_j'\Sigma_{YY}\mathbf{k}_i. \qquad [20.A.6]$$

Similarly, replace i with j in [20.1.10]:

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_j\mathbf{k}_j, \qquad [20.A.7]$$

and premultiply by $\mathbf{k}_i'\Sigma_{YY}$;

$$\mathbf{k}_i'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_j\mathbf{k}_i'\Sigma_{YY}\mathbf{k}_j. \qquad [20.A.8]$$

Subtracting [20.A.8] from [20.A.6], we see that

$$0 = (\lambda_i - \lambda_j)\mathbf{k}_j'\Sigma_{YY}\mathbf{k}_i. \qquad [20.A.9]$$

If $i \neq j$, then $\lambda_i \neq \lambda_j$ and [20.A.9] establishes that $\mathbf{k}_j'\Sigma_{YY}\mathbf{k}_i = 0$ for $i \neq j$. For $i = j$, we normalized $\mathbf{k}_i'\Sigma_{YY}\mathbf{k}_i = 1$ in [20.1.11]. Thus we have established condition [20.1.3] for the case of distinct eigenvalues.

Virtually identical calculations show that [20.1.13] and [20.1.14] imply [20.1.4].

(d) Transpose [20.1.13] and postmultiply by $\Sigma_{XY}\mathbf{k}_j$:

$$\mathbf{a}_i'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_i\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j. \qquad [20.A.10]$$

Similarly, premultiply [20.A.7] by $\mathbf{a}_i'\Sigma_{XY}$:

$$\mathbf{a}_i'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_j = \lambda_j\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j. \qquad [20.A.11]$$

Subtracting [20.A.11] from [20.A.10] results in

$$0 = (\lambda_i - \lambda_j)\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j.$$

This shows that $\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j = 0$ for $\lambda_i \neq \lambda_j$, as required by [20.1.5].

To find the value of $\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_j$ for $i = j$, premultiply [20.1.13] by $\mathbf{a}_i'\Sigma_{XX}$, making use of [20.1.14]:

$$\mathbf{a}_i'\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\mathbf{a}_i = \lambda_i. \qquad [20.A.12]$$

Let us suppose for illustration that n_1 is the smaller of n_1 and n_2; that is, $n = n_1$.[5] Then the matrix of eigenvectors $\mathcal{K}$ is $(n \times n)$ and nonsingular. In this case, [20.1.3] implies that

$$\Sigma_{YY} = [\mathcal{K}']^{-1}\mathcal{K}^{-1},$$

or, taking inverses,

$$\Sigma_{YY}^{-1} = \mathcal{K}\mathcal{K}'. \qquad [20.A.13]$$

Substituting [20.A.13] into [20.A.12], we find that

$$\mathbf{a}_i'\Sigma_{XY}\mathcal{K}\mathcal{K}'\Sigma_{YX}\mathbf{a}_i = \lambda_i. \qquad [20.A.14]$$

Now,

$$\begin{aligned} \mathbf{a}_i'\Sigma_{XY}\mathcal{K} &= \mathbf{a}_i'\Sigma_{XY}[\mathbf{k}_1 \quad \mathbf{k}_2 \quad \cdots \quad \mathbf{k}_n] \\ &= [\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_1 \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_2 \quad \cdots \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i \quad \cdots \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_n] \qquad [20.A.15] \\ &= [0 \quad 0 \quad \cdots \quad \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i \quad \cdots \quad 0]. \end{aligned}$$

Substituting [20.A.15] into [20.A.14], it follows that

$$(\mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i)^2 = \lambda_i.$$

Thus, the ith canonical correlation,

$$r_i \equiv \mathbf{a}_i'\Sigma_{XY}\mathbf{k}_i,$$

is given by the square root of the eigenvalue λ_i, as claimed:

$$r_i^2 = \lambda_i. \quad \blacksquare$$

[5]In the converse case when $n = n_2$, a parallel argument can be constructed using the fact that

$$\mathbf{k}_i'\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\mathbf{k}_i = \lambda_i.$$

Chapter 20 Exercises

20.1. In this problem you are asked to verify the claim in the text that the first canonical variates η_{1t} and ξ_{1t} represent the linear combinations of $\mathbf{y}_t$ and $\mathbf{x}_t$ with maximum possible correlation. Consider the following maximization problem:

$$\max_{\{\mathbf{k}_1, \mathbf{a}_1\}} E(\mathbf{k}_1'\mathbf{y}_t\mathbf{x}_t'\mathbf{a}_1)$$

subject to

$$E(\mathbf{k}_1'\mathbf{y}_t\mathbf{y}_t'\mathbf{k}_1) = 1$$

$$E(\mathbf{a}_1'\mathbf{x}_t\mathbf{x}_t'\mathbf{a}_1) = 1.$$

Show that the maximum value achieved for this problem is given by the square root of the largest eigenvalue of the matrix $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$, and that $\mathbf{a}_1$ is the associated eigenvector normalized as stated. Show that $\mathbf{k}_1$ is the normalized eigenvector of $\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$, associated with this same eigenvalue.

20.2. It was claimed in the text that the maximized log likelihood function under the null hypothesis of h cointegrating relations was given by [20.3.2]. What is the nature of the restriction on the VAR in [20.3.1] when $h = 0$? Show that the value of [20.3.2] for this case is the same as the log likelihood for a $VAR(p - 1)$ process fitted to the differenced data $\Delta\mathbf{y}_t$.

20.3. It was claimed in the text that the maximized log likelihood function under the alternative hypothesis of n cointegrating relations was given by [20.3.3]. This case involves regressing $\Delta\mathbf{y}_t$ on a constant, $\mathbf{y}_{t-1}$, and $\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}$ without restrictions. Let $\hat{\mathbf{g}}_t$ denote the residuals from this unrestricted regression, with $\hat{\Sigma}_{GG} = (1/T)\Sigma_{t=1}^{T} \hat{\mathbf{g}}_t\hat{\mathbf{g}}_t'$. Equation [11.1.32] would then assert that the maximized log likelihood function should be given by

$$\mathcal{L}_A^* = -(Tn/2)\log(2\pi) - (T/2)\log|\hat{\Sigma}_{GG}| - (Tn/2).$$

Show that this number is the same as that given by formula [20.3.3].

20.4. Consider applying Johansen's likelihood ratio test to univariate data ($n = 1$). Show that the test of the null hypothesis that y_t is nonstationary ($h = 0$) against the alternative that y_t is stationary ($h = 1$) can be written

$$T[\log(\hat{\sigma}_0^2) - \log(\hat{\sigma}_1^2)],$$

where $\hat{\sigma}_0^2$ is the average squared residual from a regression of Δy_t on a constant and $\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1}$ while $\hat{\sigma}_1^2$ is the average squared residual when y_{t-1} is added as an explanatory variable to this regression.

Chapter 20 References

Ahn, S. K., and G. C. Reinsel. 1990. "Estimation for Partially Nonstationary Multivariate Autoregressive Models." *Journal of the American Statistical Association* 85:813–23.

Christiano, Lawrence J., and Lars Ljungqvist. 1988. "Money Does Granger-Cause Output in the Bivariate Money-Output Relation." *Journal of Monetary Economics* 22:217–35.

Johansen, Søren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12:231–54.

———. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59:1551–80.

———. 1992. "Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis." *Journal of Econometrics* 52:389–402.

——— and Katarina Juselius. 1990. "Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money." *Oxford Bulletin of Economics and Statistics* 52:169–210.

Koopmans, Tjalling C., and William C. Hood. 1953. "The Estimation of Simultaneous Linear Economic Relationships," in William C. Hood and Tjalling C. Koopmans, eds., *Studies in Econometric Method*. New York: Wiley.

Mosconi, Rocco, and Carlo Giannini. 1992. "Non-Causality in Cointegrated Systems: Representation, Estimation and Testing," *Oxford Bulletin of Economics and Statistics*. 54:399–417.

Park, Joon Y. 1990. "Maximum Likelihood Estimation of Simultaneous Cointegrated Models." University of Aarhus. Mimeo.

———. 1992. "Canonical Cointegrating Regressions." *Econometrica* 60:119–43.

——— and Masao Ogaki. 1991. "Inference in Cointegrated Models Using VAR Prewhitening to Estimate Shortrun Dynamics." University of Rochester. Mimeo.

Phillips, Peter C. B. 1990. "Spectral Regression for Cointegrated Time Series," in William Barnett, James Powell, and George Tauchen, eds., *Nonparametric and Semiparametric Methods in Economics and Statistics*. New York: Cambridge University Press.

———. 1991a. "Optimal Inference in Cointegrated Systems." *Econometrica* 59:283–306.

———. 1991b. "Unidentified Components in Reduced Rank Regression Estimation of ECM's." Yale University. Mimeo.

——— and Bruce E. Hansen. 1990. "Statistical Inference in Instrumental Variables Regression with I(1) Processes." *Review of Economic Studies* 57:99–125.

——— and S. Ouliaris. 1990. "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica* 58:165–93.

Stock, James H., and Mark W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* 83:1097–1107.

——— and ———. 1989. "Interpreting the Evidence on Money-Income Causality." *Journal of Econometrics* 40:161–81.

Toda, H. Y., and Peter C. B. Phillips. Forthcoming. "Vector Autoregression and Causality." *Econometrica*.

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