Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating
Coefficients in Error Correction Models
Author(s): Peter C. B. Phillips

SOME EXACT DISTRIBUTION THEORY FOR MAXIMUM LIKELIHOOD ESTIMATORS OF COINTEGRATING COEFFICIENTS IN ERROR CORRECTION MODELS

By Peter C. B. Phillips[1]

This paper derives some exact finite sample distributions and characterizes the tail behavior of maximum likelihood estimators of the cointegrating coefficients in error correction models. It is shown that the reduced rank regression estimator has a distribution with Cauchy-like tails and no finite moments of integer order. The maximum likelihood estimator of the coefficients in a particular triangular system representation is studied and shown to have matrix t-distribution tails with finite integer moments to order $T - n + r$ where T is the sample size, n is the total number of variables in the system, and r is the dimension of the cointegration space. These results help to explain some recent simulation studies where extreme outliers are found to occur more frequently for the reduced rank regression estimator than for alternative asymptotically efficient procedures that are based on the triangular representation. In a simple triangular system, the Wald statistic for testing linear hypotheses about the columns of the cointegrating matrix is shown to have an F distribution, analogous to Hotelling's T^2 distribution in multivariate linear regression.

Keywords: Hypergeometric function, invariant measure, matrix Cauchy, maximum likelihood, orthogonal group, reduced rank regression, spherical distribution, Stiefel manifold, triangular system.

1. INTRODUCTION

Several procedures are now available for obtaining estimates of cointegrating coefficients that are known to be asymptotically efficient under Gaussian assumptions. A commonly used method in practical work at present is reduced rank regression—see Johansen (1988) and also Ahn-Reinsel (1988, 1990). This method applies maximum likelihood to a systems error correction model (ECM) formulated with vector autoregression (VAR) dynamics. An alternative parametric maximum likelihood method was developed in Phillips (1991) and is based on a triangular system representation of the cointegration model. The Phillips and Johansen procedures are, in fact, asymptotically equivalent when the same coordinate system for the cointegration space is used (see Park (1990) for a demonstration), but these procedures are not equivalent in finite samples. This is because the triangular system representation explicitly incorporates identifying conditions on the cointegration space that isolate a set of full rank integrated regressors and identify the cointegrating equations as individual structural equations. These conditions are analogous to traditional identifying restrictions in simultaneous equations models. The Johansen reduced rank regression approach places no prior restrictions on the cointegration space but

---

employs normalization rules that uniquely determine empirical estimates of the cointegrating vectors as generalized eigenvectors in a canonical correlation analysis between the levels and the differences of the time series. These distinctions between the procedures do have important effects, some of which will be explored in the present paper. In addition to the above mentioned procedures, several alternative efficient methods of estimation are now available including a fully modified least squares method due to Phillips and Hansen (1990) that also uses the triangular representation.

Some attempts have been made to evaluate these different procedures by simulation. Recent work by Corbae, Ouliaris, and Phillips (1990), Gregory (1991), Hargreaves and Setiabudi (1992), Park and Ogaki (1991), Cappuccio and Lubian (1992), and Toda and Phillips (1993) is relevant in this connection. One common feature to emerge from these simulation studies is that the Johansen reduced rank regression procedure seems occasionally to produce estimates that are very unreliable in the sense of being extreme outliers, where the other methods do not.

One aim of the present paper is to provide an analytical basis for understanding this phenomenon. The methods used are related to some earlier work by the author (1984, 1986a) on the exact distribution of simultaneous equations estimators. Formulae are derived for the exact finite sample distribution of the reduced rank regression estimator in the general case and the leading factor in this distribution is shown to be proportional to a matrix Cauchy distribution. Under a mild dominance requirement on the density of the squared canonical correlation matrix, it is shown that the density of the reduced rank regression estimator has Cauchy-like tail behavior and therefore no finite first moments, thereby providing an explanation for the outlier behavior observed in the aforementioned simulation studies. Maximum likelihood estimators that are based on the triangular representation are studied in a particular case and are shown to have quite different finite sample tail behavior. Unlike the reduced rank regression estimator, these estimators possess finite integer moments up to order $T - m$, where T is the sample size and m is the number of full rank integrated regressors. Some exact results for the same model on the distribution of Wald tests about the cointegrating coefficient matrix are given, including a version of Hotelling's T^2 test in multivariate linear regression.

Throughout this paper we use the symbol " $\equiv$ " to signify equivalence in distribution and $\mathcal{N}(\cdot)$ and $\mathcal{R}(\cdot)$ to represent the null and range spaces of their respective matrix arguments. $O(n) = \{H(n \times n): H'H = I_n\}$ denotes the orthogonal group of $n \times n$ orthogonal matrices, $V_{r,n} = \{A(n \times r): A'A = I_r\}$ is the Stiefel manifold, and $\Gamma_n(a) = \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma(a - (1/2)(i - 1))$, where $\mathrm{Re}(a) > (n - 1)/2$, is the multivariate gamma function. The matrix variate $X(n \times r)$ is said to be spherically distributed if $X \equiv C_1 X C_2$ for all $C_1 \in O(n)$ and $C_2 \in O(r)$. If, in addition, $X'X = I_r$, then X has a uniform distribution on the manifold $V_{r,n}$. This distribution on $V_{r,n}$ is, in fact, uniquely determined by its invariance under the orthogonal transformations C_1 and C_2. The reader is referred to Muirhead (1982, Ch. 1-3), James (1954), and Herz (1955) for more background on these matrix spaces and distributions.

The plan of the paper is as follows. Section 2 develops a distribution theory for the reduced rank regression estimator and examines the tail behavior of this distribution. Section 3 derives the distribution of the maximum likelihood estimator of the cointegrating matrix and some Wald tests in a particular case of the triangular representation. The tail behavior of the distribution of this estimator is obtained and compared with that of the reduced rank regression estimator. Section 4 discusses the implications of these results and draws some conclusions. Technical derivations are given in the Appendix in Section 5.

2. SOME EXACT FINITE SAMPLE THEORY FOR THE REDUCED RANK REGRESSION ESTIMATOR

Suppose the n-vector time series y_t satisfies an ECM of the form

$$(1) \qquad J(L)\,\Delta y_t = \Gamma A' y_{t-1} + \varepsilon_t \quad (t = 1, \ldots, T), \quad J(L) = \sum_0^{k-1} J_i L^i, \quad J_0 = I,$$

$$\varepsilon_t \equiv \text{iid } N(0, \Sigma_\varepsilon).$$

In (1) A' is an $r \times n$ matrix of cointegrating vectors and $\Gamma(n \times r)$ is a matrix of factor loading coefficients.

Suppose that the matrix A in (1) is estimated by reduced rank regression. It may be assumed that the dimension of the cointegrating space $r \geqslant 1$ is pre-scribed or it could be chosen by the likelihood ratio trace statistic of Johansen (1988). (Even when $\Gamma = 0$ this outcome of the test, i.e. $r \geqslant 1$, occurs with positive probability.) The reduced rank regression estimator of A satisfies the optimization problem

$$(2) \qquad \hat{A} = \arg\min_A |S_{00} - S_{01} A (A' S_{11} A)^{-1} A' S_{10}|$$

(see Johansen (1988, p. 234)) subject to the following normalization conditions which ensure the empirical uniqueness of $\hat{A}$:

$$(3) \qquad \hat{A}' S_{11} \hat{A} = I_r,$$

$$(4) \qquad \hat{A}' S_{10} S_{00}^{-1} S_{01} \hat{A} = \hat{\Lambda}_r,$$

where $\hat{\Lambda}_r = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_r)$ and $\hat{\lambda}_1 \geqslant \cdots \geqslant \hat{\lambda}_r \geqslant 0$ are the first r ordered roots of the determinental equation $|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$. The S_{ij} in (2)–(4) are moment matrices of the residuals from the regression of Δy_t $(i = 0)$ and y_{t-1} $(i = 1)$ on the lagged differences Δy_{t-j} $(j = 1, \ldots, k - 1)$ that appear in (1). The notation is the same as Johansen's (1988) except that we use A' in place of his β' and the index 1 in place of his k (the latter since we use y_{t-1} rather than y_{t-k} in the ECM formulation (1)). Let $S = S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$. Then the eigenvalues of S are the squared canonical correlation coefficients, the first r of which comprise the diagonal elements of $\hat{\Lambda}_r$.

Let us now replace the empirical normalization rules (3) and (4) with the requirement that the leading submatrix of the cointegrating matrix A' is the

identity. That is, suppose we set

(5) $A' = [I_r, -B]$.

This condition corresponds to the *a priori* requirement that there be r structural relations of the form

(6) $y_{1t} = By_{2t} + u_{1t}$,

where $y_t' = [y_{1t}', y_{2t}']$, is a partition of y_t that is conformable with (5), y_{2t} is a full rank integrated process, and u_{1t} is a stationary error. This is the type of formulation that occurs frequently in applied econometric work. Note that the explicit form (6) is stronger than the empirical normalization given by (3) and (4). In particular, the form of (6) explicitly recognizes a subvector, y_{2t}, of full rank integrated regressors and (6) involves r restrictions per equation (i.e. $r - 1$ exclusion restrictions and one unit normalization restriction) on the parameters of the cointegrating matrix A', giving r^2 restrictions *in toto*. The normalizations (3) and (4) also involve r^2 restrictions but they apply to the estimate $\hat{A}'$, not the true cointegrating matrix A'. The difference is important because in the limit A' may not be uniquely determined by these empirical restrictions (as happens when there are multiple eigenvalues in the limit matrix Λ_r corresponding to $\hat{\Lambda}_r$ in (4)).

Once the matrix A in (1) is estimated by reduced rank regression, *ex post* estimation of the submatrix B in (5) is always possible and this is often what is done in practice. Thus, partitioning A' and $\hat{A}'$ conformably with (5) as $A' = [A_1, A_2]$, we have $B = -A_1^{-1}A_2$ and the corresponding estimator $\hat{B} = -\hat{A}_1^{-1}\hat{A}_2$. Observe that $\hat{B}$ is invariant to rotations and scalings of the cointegration space defined by $\hat{A}'$. For instance, writing $\bar{\hat{A}} = \hat{A}(\hat{A}'\hat{A})^{-1/2} \in V_{r,n}$ we could replace (3) by $\bar{\hat{A}}'\bar{\hat{A}} = I_r$, and $-\bar{\hat{A}}_1^{-1}\bar{\hat{A}}_2 = -\hat{A}_1^{-1}\hat{A}_2$ is clearly invariant to this rescaling of $\hat{A}$. Thus, $\hat{B}$ is uniquely determined from $\hat{A}$ irrespective of whether the empirical normalizations (3) and (4) are employed. The finite sample distribution of $\hat{B}$ is therefore also invariant to these transformations. The following result gives a general expression for the density of $\hat{B}$.

2.1 THEOREM: *The exact finite sample distribution of the reduced rank regression estimator $\hat{B}$ of the structural coefficient matrix B in (6) is given by the density*

(7) $\text{pdf}(\hat{B}) = \left[\pi^{-n^2/2 + rm/2}\Gamma_r(r/2)\right]^{-1}|I_r + \hat{B}\hat{B}'|^{-(r+m)/2}$

$$\times \int_{0 < \hat{\Lambda} < I} \int_{L \in O(r)} \int_{K \in O(m)} \prod_{i<j}\left(\hat{\lambda}_i - \hat{\lambda}_j\right) f\left(C(\hat{B}, L, K)\hat{\Lambda}C\right)$$

$$\times (\hat{B}, L, K)'\big)(\underline{dL})(\underline{dK})\, d\hat{\Lambda},$$

where $m = n - r$, $f(S)$ is the probability density of $S = S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2}$ and the orthogonal matrix $C = C(\hat{B}, L, K)$ depends on $\hat{B}, L, K$ through the partitioned

decomposition $C = [C_1, C_2]$ *with*

$$(8) \qquad C_1 = \begin{bmatrix} I_r \\ -\hat{B}' \end{bmatrix} \left(I_r + \hat{B}\hat{B}' \right)^{-1/2} L, \qquad C_2 = \begin{bmatrix} \hat{B} \\ I_m \end{bmatrix} \left(I_m + \hat{B}'\hat{B} \right)^{-1/2} K.$$

The matrices L and K in (8) *are in* $O(r)$ *and* $O(m)$, *respectively, and* $(\underline{dL})$ *and* $(\underline{dK})$ *in* (7) *represent the normalized invariant measures on these two spaces.*

2.2 REMARKS: (i) Formula (7) is a useful general expression for the density of $\hat{B}$. The formula applies even when the errors in (1) are not normally distributed, although of course any departure from normality certainly affects the form of the density $f(\cdot)$ in (7).

(ii) The analytic form of the density $f(S)$ of the squared canonical correlation matrix S that appears in (7) is not known even when the ε_t are normally distributed or in the scalar case where $n = 1$ and seems beyond reach, at least with existing methods. The problem of determining $f(\cdot)$ is related to the problem of finding the exact distribution of the estimated coefficients in a vector autoregression, a long standing problem that is unresolved even for the scalar first order autoregressive model. Nevertheless, (7) is useful because under what seem to be very reasonable conditions on the form of $f(\cdot)$ (7) may be used to determine the tail behavior of the estimator $\hat{B}$. In particular, we shall assume that $f(\cdot)$ satisfies the following condition.

CONDITION (D): *There exists a function* $g(\cdot)$ *such that for each* $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_n)$ *with* $0 < \hat{\Lambda} < I$

$$(9) \qquad f(C\hat{\Lambda}C') \leq g(\hat{\Lambda}) \qquad \textit{for all} \qquad C \in O(n) \qquad \textit{and}$$

$$(10) \qquad \int_{\hat{\Lambda}} \prod_{i<j} \left(\hat{\lambda}_i - \hat{\lambda}_j \right) g(\hat{\Lambda}) \, d\hat{\Lambda} < \infty.$$

Condition (D) requires that $f(C\hat{\Lambda}C')$ be dominated above by a function $g(\hat{\Lambda})$ that satisfies the integrability condition (10). This seems like a rather mild requirement but it is also very difficult to confirm that it is satisfied, and we have not been able to do so even under Gaussian errors, let alone more generally. However, there are conditions and model specializations under which (D) holds. For instance, (9) is certainly satisfied whenever $f(C\hat{\Lambda}C')$ is continuous in C because $O(n)$ is compact. Moreover, when $f(\cdot)$ is a symmetric function of its matrix argument (i.e. when $f(\cdot)$ is a function of the n elementary symmetric functions $\sigma_1 = \mathrm{tr}(S), \ldots, \sigma_n = \det(S)$) we have $f(C\hat{\Lambda}C') = f(\Lambda)$ and then (9) and (10) are automatically satisfied. The case where $f(\cdot)$ is a symmetric function of matrix argument does arise in an interesting specialization of the model (1) and that case will be examined further below.

(iii) The leading factor in the density function (7) involves a multiple of the determinantal expression $|I + \hat{B}\hat{B}'|^{-(r+m)/2}$, which is proportional to a matrix Cauchy density (see (15) below and Lemma 5.2) for the $r \times m$ matrix $\hat{B}$. This

factor determines the tail shape of the density (7) as the following corollary shows.

2.3 COROLLARY: *Suppose $f(\cdot)$ satisfies condition (D) and let $\hat{B}$ approach the limits of its domain of definition along the ray $\hat{B} = bB_0$ for some fixed matrix $B_0 \neq 0$ and scalar b which tends to infinity. Then*

(11) $\text{pdf}\left(\hat{B} = bB_0\right) = k_0 |I + b^2 B_0 B_0'|^{-(r+m)/2}(1 + o(1)), \quad as \quad b \to \infty,$

where

(12) $k_0 = \left[\pi^{-n^2/2 + rm/2} \Gamma_r(r/2) \right]^{-1}$

$$\times \int_{0 < \hat{\Lambda} < I} \int_{L \in O(r)} \int_{K \in O(m)} \prod_{i < j} \left(\hat{\lambda}_i - \hat{\lambda}_j \right) f\left(C_0 \hat{\Lambda} C_0' \right) (\underline{dL})(\underline{dK}) \, d\hat{\Lambda}.$$

In (12) $C_0 = [C_{01}, C_{02}] \in O(n)$ *is partitioned conformably with* (8),

(13) $C_{01} = \begin{bmatrix} (F_1 F_1')_0 \\ -B_0' F_2 (F_2' B_0 B_0' F_2)^{-1/2} F_2' \end{bmatrix} L,$

$C_{02} = \begin{bmatrix} B_0 G_2 (G_2' B_0' B_0 G_2)^{-1/2} G_2' \\ (G_1 G_1')_0 \end{bmatrix} K,$

$F = [F_1, F_2] \in O(r)$ *with* $F_1 \in \mathcal{N}(B_0')$, $F_2 \in \mathcal{R}(B_0)$ *and* $G = [G_1, G_2] \in O(m)$ *with* $G_1 \in \mathcal{N}(B_0)$ *and* $G_2 \in \mathcal{R}(B_0')$. *The notation* $(\ )_0$ *in* (13) *signifies that the matrix in parentheses is replaced by a zero matrix when the corresponding null space (i.e. $\mathcal{N}(B_0')$ and $\mathcal{N}(B_0)$) has zero dimension.*

The tail behavior of the density (11) on the ray $\hat{B} = bB_0$ is equivalent (up to a constant multiple) to that of a matrix Cauchy distribution. It follows that the density of the reduced rank regression estimator $\hat{B}$ has no finite sample integer moments. This may go some way to explain the occurrence of the extreme outliers that have been observed in simulation studies of this estimator.

(iv) A special case that is of some independent interest is that of spurious regression. Here the system that is estimated is the vector autoregression (1) with $r > 0$ when the generating mechanism of y_t is a set of n random walks initialized at the origin at $t = 0$, i.e.

(14) $\Delta y_t = \varepsilon_t, \qquad y_0 = 0,$

where ε_t is spherically symmetric, i.e. $\varepsilon_t \equiv H\varepsilon_t$ for all $H \in O(n)$. The case of independent Gaussian random walks is of primary interest and here (14) is a specialization of (1) with $\Gamma = 0$, $J(L) = I$, $\Sigma_\varepsilon = I$, $y_0 = 0$, and $\varepsilon_t \equiv$ iid $N(0, I)$. Under the generating mechanism (14) the density of $\hat{B}$ given in Theorem 2.1 has a very simple form as the next result shows.

2.4 COROLLARY: *Suppose y_t is generated as a vector of random walks as in* (14) *with spherically symmetric errors ε_t and the reduced rank vector autoregression* (1) *is estimated with $r \geqslant 1$. Then the exact distribution of the resulting*

estimator $\hat{B}$ is matrix Cauchy with probability density

(15) $\quad \mathrm{pdf}(\hat{B}) = \left[\pi^{r(n-r)/2}\Gamma_r(r/2)\right]^{-1}\Gamma_r(n/2)|I_r + \hat{B}\hat{B}'|^{-n/2}.$

When $r = 1$ and $m = n - 1$ the distribution (15) has the familiar multivariate Cauchy form, i.e.

(16) $\quad \mathrm{pdf}(\hat{b}) = \left[\pi^{m/2}\Gamma(m/2)\right]^{-1}\Gamma((m+1)/2)(1 + b'b)^{-(m+1)/2},$

where $\hat{B} = \hat{b}'$ is here a $1 \times m$ vector. Both (15) and (16) are preserved under marginalization in the sense that all submatrices of $\hat{B}$ have Cauchy distributions. The fact that elements of $\hat{B}$ have no finite integer moments is the consequence of the fact that $\hat{B}$ is a matrix quotient of submatrices of $\hat{A}$ or $\bar{\hat{A}} = \hat{A}(\hat{A}'\hat{A})^{-1/2}$, viz.

$$\hat{B} = -\hat{A}_1^{-1}\hat{A}_2 = -\bar{\hat{A}}_1^{-1}\bar{\hat{A}}_2 = -\mathrm{adj}\left(\bar{\hat{A}}_1\right)\bar{\hat{A}}_2/\det\left(\bar{\hat{A}}_1\right).$$

Take the (i, j) element of this quotient. We can always find an $\bar{\hat{A}}$ in $V_{r,n}$ for which $[\mathrm{adj}(\bar{\hat{A}}_1)\bar{\hat{A}}_2]_{ij}$ is nonzero and $\det(\bar{\hat{A}}_1) = 0$. Select an $\bar{\hat{A}}$ in $V_{r,n}$ for which the ith row of $\bar{\hat{A}}_1$ is zero while $[\mathrm{adj}(\bar{\hat{A}}_1)\bar{\hat{A}}_2]_{ij}$ is nonzero. This will always be possible because, as shown in the proof of Corollary 2.4, $\bar{\hat{A}}$ is uniformly distributed on $V_{r,n}$. Thus, we have $\mathrm{pdf}(\bar{\hat{A}}) = \mathrm{constant} > 0$ at this value of $\bar{\hat{A}}$, whereas $\det(\bar{\hat{A}}_1) = 0$ and $[\mathrm{adj}(\bar{\hat{A}}_1)\bar{\hat{A}}_2]_{ij} \neq 0$ by construction. This is sufficient to ensure that the random variable $[\hat{B}]_{ij}$ has no finite integer moments—see Sargan (1988; Theorem 1) and Phillips (1983, Theorem 3.9.1).

3. TAIL BEHAVIOR OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE TRIANGULAR SYSTEM REPRESENTATION

In place of (1) let us now suppose that maximum likelihood is applied to the triangular system representation, viz.

(17) $\quad y_{1t} = By_{2t} + u_{1t},$

(18) $\quad \Delta y_{2t} = u_{2t},$

where $u_t = (u'_{1t}, u'_{2t})' \equiv \mathrm{iid}\ N(0, \Sigma_u)$. This system can also be written in the ECM format by setting

(19) $\quad J(L) = I_n, \quad \Gamma = -\begin{bmatrix} I_r \\ 0 \end{bmatrix}, \quad A' = [I, -B],$

$\quad\quad \varepsilon_{1t} = u_{1t} + Bu_{2t}, \quad \text{and} \quad \varepsilon_{2t} = u_{2t}.$

Under this parameterization, observe that (1) is linear in both the variables and the coefficients because the factor loading matrix Γ is now known. (The dependence of the error ε_{1t} on the matrix B can be ignored, since we assume Σ_u is unrestricted.) Because of this difference the maximum likelihood estimator $\tilde{B}$ of B in (17) is different from the reduced rank regression estimator $\hat{B}$. Note that the system (17)–(18) is restrictive because we do not have any time dependence in the error process u_t.

As shown in Phillips (1991), $\tilde{B}$ is equivalent to the OLS estimator of B in the augmented regression equation

(20) $y_{1t} = By_{2t} + D\,\Delta y_{2t} + u_{1\cdot 2t}$

where $D = \Sigma_{12}\Sigma_{22}^{-1}$, $u_{1\cdot 2t} = u_{1t} - \Sigma_{12}\Sigma_{22}^{-1}u_{2t}$, and

$$\Sigma_u = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

is a partition of Σ conformable with that of u_t. Observe that the error $u_{1\cdot 2t}$ in (20) is independent of the regressors y_{2t} and Δy_{2t} by virtue of its construction and $u_{1\cdot 2t} \equiv$ iid $N(0, \Sigma_{11\cdot 2})$, where $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. In conventional partitioned regression notation the estimator $\tilde{B}$ has the form

(21) $\tilde{B} = Y_1'Q_\Delta Y_2(Y_2'Q_\Delta Y_2)^{-1}$.

The error in the estimator $\tilde{B}$ is written as $\tilde{E} = \tilde{B} - B = U_{1\cdot 2}Q_\Delta Y_2(Y_2'Q_\Delta Y_2)^{-1}$, again in conventional regression notation.

Let $\mathscr{F}_2$ be the σ-field generated by $\{u_{2t}\colon t = 1,\ldots,T\}$. Then conditional on $\mathscr{F}_2$, the estimation error $\tilde{E}$ is linear in $U_{1\cdot 2}$ and hence Gaussian. The conditional distribution of $\tilde{E}$ is $N(0, \Sigma_{11\cdot 2} \otimes F)$ with density

(22) $\text{pdf}\left(\tilde{E}|\mathscr{F}_2\right) = (2\pi)^{-rm/2}|F|^{-r/2}\,\text{etr}\left\{-(1/2)\tilde{E}'\Sigma_{11\cdot 2}^{-1}\tilde{E}F^{-1}\right\}$,

where $F = (Y_2'Q_\Delta Y_2)^{-1}$. The marginal density of $\tilde{E}$ is now obtained by integrating out the conditioning variates in (22) with respect to their associated probability measures. The analysis is assisted by transforming to a set of canonical variates which make the integrations easier. The derivations are a little involved and are given in the Appendix. They lead to the following result.

3.1 THEOREM: (a) *The exact finite sample distribution of $\tilde{E} = \tilde{B} - B$ where $\tilde{B}$ is the maximum likelihood estimator of B in the triangular cointegrated system* (17) *and* (18) *is given by*

(23) $\text{pdf}(\tilde{E}) = \pi^{-mr/2}\left[\Gamma_m(T/2)\right]^{-1}\Gamma_m((T+r)/2)|\Sigma_{11\cdot 2}|^{-m/2}|\Sigma_{22}|^{r/2}$

$$\times \int_{V_{m,T}} |K(\Xi_2)|^{r/2}$$

$$\times \int_{O(m)} \left|I_m + H_3 K(\Xi_2)^{-1}H_3'\Sigma_{22}^{1/2}\tilde{E}'\Sigma_{11\cdot 2}^{-1}\tilde{E}\Sigma_{22}^{1/2}\right|^{-(T+r)/2}$$

$$\times (\underline{dH_3})(\underline{d\Xi_2})$$

where

$$K(\Xi_2) = \Xi_2'L'(I - \Xi_2\Xi_2')L\Xi_2,$$

and

$$L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}(T \times T).$$

In the above expressions the matrices $\Xi_2 \in V_{m,T}$ and $H_3 \in O(m)$, and $(\underline{d\Xi_2})$, $(\underline{dH_3})$ signify the normalized invariant measures on $V_{m,T}$ and $O(m)$, respectively.

(b) *The finite sample distribution of the scaled estimation error*

$$E_N = \tilde{E}(Y_2' Q_\Delta Y_2)^{1/2} = (\tilde{B} - B)(Y_2' Q_\Delta Y_2)^{1/2}$$

is matrix normal with variance matrix $\Sigma_{11 \cdot 2} \otimes I_m$.

3.2 REMARKS: (i) Let $\check{E} = \Sigma_{11 \cdot 2}^{-1/2} \tilde{E} \Sigma_{22}^{1/2}$. This is the estimation error of the maximum likelihood estimator in a "standardized model"—see (36) and (37) in the Appendix. Consideration of $\check{E}$ helps to simplify the analysis of the properties of the density (23). The density of $\check{E}$ is

$$(24) \qquad \text{pdf}(\check{E}) = \pi^{-mr/2} \left[\Gamma_m(T/2) \right]^{-1} \Gamma_m((T+r)/2)$$

$$\times \int_{V_{m,T}} |K(\Xi_2)|^{-r/2} \int_{O(m)} |I_m + H_3 K(\Xi_2)^{-1} H_3' \check{E}' \check{E}|^{-(T+2)/2}$$

$$\times (\underline{dH_3})(\underline{d\Xi_2}).$$

It is apparent from the form of (24) that the matrix variate $\check{E}$ is spherically symmetric in the sense that $\check{E} \equiv J_1 \check{E} J_2$ with $J_1 \in O(r)$ and $J_2 \in O(m)$. This is seen by making the replacement $\check{E} \to J_1 \check{E} J_2$ in (24) and noting that $\overline{H}_3 = H_3' J_2' \in O(m)$ so that upon integrating over $\overline{H}_3$ the density (24) is unchanged by the replacement.

(ii) Since $\check{E}$ is spherically symmetric all marginal distributions of $\check{E}$ have the same form. Note that when $r = m = 1$ the density (24) is

$$(25) \qquad \text{pdf}(\check{e}) = \pi^{-1/2} \Gamma(T/2)^{-1} \Gamma((T+1)/2)$$

$$\times \int_{V_{1,T}} k(\Xi_2)^{-1/2} \left[1 + k(\Xi_2)^{-1} \check{e}^2 \right]^{-(T+1)/2} (\underline{d\Xi_2}),$$

which is a mixture of scalar t-variates with T degrees of freedom. More generally when $r = 1$ and $m \geqslant 1$ we have the following density for $\check{E} = \check{e}'$ $(1 \times m)$

$$(26) \qquad \text{pdf}(\check{e}) = \pi^{-m/2} \Gamma_m(T/2)^{-1} \Gamma_m((T+1)/2)$$

$$\times \int_{V_{m,T}} |K(\Xi_2)|^{-1/2}$$

$$\times \int_{O(m)} |I_m + H_3 K(\Xi_2)^{-1} H_3' \check{e} \check{e}'|^{-(T+1)/2} (\underline{dH_3})(\underline{d\Xi_2}),$$

$$= \pi^{-m/2} \left[\Gamma(T - m + 1)/2 \right]^{-1} \Gamma((T+1)/2)$$

$$\times \int_{V_{m,T}} |K(\Xi_2)|^{-1/2}$$

$$\times \int_{O(m)} \left(1 + \check{e}' H_3 K(\Xi_2)^{-1} H_3' \check{e} \right)^{-(T+1)/2} (\underline{dH_3})(\underline{d\Xi_2}),$$

which is a covariance matrix mixture of multivariate t-variates with $T - m + 1$ degrees of freedom.

(iii) The tail behavior of the marginal density of an individual component $\check{e}_1$ of $\check{e}$ can be deduced from (26). First we note that the marginal density of $\check{e}_1$ has the form

$$(27) \qquad \text{pdf}(\check{e}_1) = \pi^{-1/2}\big[\Gamma((T-m+1)/2)\big]^{-1}\Gamma((T-m+2)/2)$$

$$\times \int_{V_{m,T}} \int_{O(m)} k_{11\cdot2}(\Xi_2, H_3)^{1/2}$$

$$\times \big[1 + k_{11\cdot2}(\Xi_2, H_3)\check{e}_1^2\big]^{-(T-m+2)/2}(dH_3)(d\Xi_2),$$

where $k_{11\cdot2}(\Xi_2, H_3)$ is defined by $k_{11\cdot2} = k_{11} - k_{12}K_{22}^{-1}k_{21}$, and the partitioned matrix elements come from

$$H_3 K(\Xi_2)^{-1} H_3' = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & K_{22} \end{bmatrix}.$$

The density (27) is a mixture of scalar t-variates with $T - m + 1$ degrees of freedom. Next we expand the binomial factor in (27) for large $|\check{e}_1|$ and integrate the expansion term by term with respect to Ξ_2 and H_3. This term by term integration of the asymptotic expansion is possible because $V_{m,T}$ and $O(m)$ are compact sets and $k_{11\cdot2} > 0$ uniformly for $\Xi_2 \in V_{m,T}$ and $H_3 \in O(m)$. The asymptotic expansion then holds uniformly in Ξ_2 and H_3, and can therefore be integrated term by term—see Erdélyi (1956, p. 16). Since

$$\big[1 + k_{11\cdot2}(\Xi_2, H_3)^{-1}\check{e}_1^2\big]^{-(T-m+2)/2}$$

$$= k_{11\cdot2}(\Xi_2, H_3)^{(T-m+2)/2}|\check{e}_1|^{-(T-m+2)}\big\{1 + O(\check{e}_1^{-2})\big\}$$

we obtain the expansion

$$(28) \qquad \text{pdf}(\check{e}_1) = c|\check{e}_1|^{-(T-m+2)}\big\{1 + O(\check{e}_1^{-2})\big\}$$

for large $|\check{e}_1|$, where

$$c = \pi^{-1/2}\Gamma((T-m+1)/2)^{-1}\Gamma((T-m+2)/2)$$

$$\times \int_{V_{m,T}} \int_{O(m)} k(\Xi_2, H_3)^{(T-m+2)/2}(d\Xi_2)(dH_3).$$

From (28) it is apparent that the maximal moment exponent of $\check{e}$ is $T - m + 1$ and integer moments of $\check{e}_1$ exist to order $T - m$. The same result holds for an arbitrary element of the matrix variate $\check{E}$. (As remarked in (i) above, all marginal distributions of $\check{E}$ have the same general form. In particular, when $r > 1$ each row of $\check{E}$ has a density of the form (26) leading to tail behavior for an individual element of $\check{E}$ that is of the form given by (28).) The same conclusion on the existence of moments applies to the unstandardized estimator $\tilde{E}$.

(iv) Part (b) of Theorem 3.1 shows that the scaled estimation error $E_N \equiv N(0, \Sigma_{11\cdot2} \otimes I_m)$. This result can be used to mount exact finite sample tests with the maximum likelihood estimator in the triangular system (17) and (18).

Suppose, for example, that we wish to test the null hypothesis

$$(29) \qquad H_0: D_1 B d_2 = d$$

about the coefficient matrix B in (17) for some given matrix $D_1(q \times r)$ of rank q and given q-vectors d_2 and d. The matrix B is estimated by $\tilde{B}$ using OLS on (20), we construct $D_1 \tilde{B} d_2 - d = D_1(\tilde{B} - B)d_2 = D_1 \tilde{E} d_2$, and then form the Wald statistic

$$(30) \qquad W = \left(D_1 \tilde{B} d_2 - d \right)' \left(D_1 \tilde{\Sigma}_{11 \cdot 2} D_1 \right)^{-1} \left(D_1 \tilde{B} d_2 - d \right) / d_2'(Y_2' Q_\Delta Y_2)^{-1} d_2,$$

where $\tilde{\Sigma}_{11 \cdot 2} = T^{-1} Y_1' Q_* Y_1$ is the maximum likelihood estimator of the error variance matrix $\Sigma_{11 \cdot 2}$ in equation (20) and the affix " $*$ " on Q_* signifies that both y_{2t} and Δy_{2t} are included in the regressor set. Our next result shows that the test statistic W is a multiple of a Hotelling's T^2 variate.

3.3. THEOREM: *The exact finite sample distribution of the Wald statistic W for testing the null hypothesis H_0 in (29) is given by*

$$(31) \qquad W \equiv \frac{Tq}{N - q + 1} F_{q, N-q+1}$$

where $N = T - 2m$.

3.4. REMARKS: (i) Theorem 3.3 shows that exact F tests of linear hypotheses of the form (29) can be constructed from the conventional Wald test in specialized triangular systems of the form (17)–(18). The result is entirely analogous to that which applies in linear multivariate regression models with fixed regressors—see Phillips (1986b, 1987).

(ii) More general exact results than (31) can be obtained in a similar way. For instance the Wald statistic for testing $H_0: D_1 B D_2 = D$ is proportional to a Hotelling's T_0^2 statistic and has the same exact finite sample distribution—see Phillips (1987) for the exact density in this case. Moreover, a general exact theory for Wald tests of hypotheses such as $H_0: D \text{ vec}(B) = d$ can be developed using operator algebra along the lines of Phillips (1986b). For the null hypothesis case, the exact theory of that paper holds here conditionally on $\mathscr{F}_2$. The unconditional distributions may then be obtained as in Theorem 3.1 by appropriate subsequent integrations. A complete development of the theory would obviously take us beyond the material concern of the present paper.

(iii) Theorems 3.1 and 3.3 can be extended to apply in a somewhat more general model than (17) and (18). Suppose (18) is replaced by the stationary VAR model

$$(18)' \qquad \Phi(L) \Delta y_{2t} = u_{2t}, \qquad \Phi(L) = \sum_{i=0}^{p} \Phi_i L^i, \qquad \Phi_0 = I_m,$$

where u_{2t} has the same properties as in (18). Then, in place of (20) we have the

augmented regression

$$(20)' \qquad y_{1t} = By_{2t} + \sum_{i=0}^{p} D_i \, \Delta y_{2t-i} + u_{1 \cdot 2t}$$

where $D_i = D\Phi_i$, $D = \Sigma_{12}\Sigma_{22}^{-1}$, and $u_{1 \cdot 2t} = u_{1t} - \Sigma_{12}\Sigma_{22}^{-1}u_{2t}$ as before. In this extended case $u_{1 \cdot 2t}$ is still orthogonal to the regressors in (20)'. Under Gaussian assumptions, the analysis that leads to Theorems 3.1 and 3.3 then goes through in the same way as before when one allows for the expanded regressor set in (20)'. All that changes is the form of the matrix $K(\Xi_2)$ that appears in expression (23) for the density of the estimation error $\tilde{E} = \tilde{B} - B$ where $\tilde{B}$ is now the OLS estimator of B in (20)'. In consequence, the earlier results on the tail behavior of the distribution of $\tilde{B}$ and exact tests of linear hypotheses about B continue to apply in the context of (20)'. As discussed in Phillips and Loretan (1991), the augmented regression (20)' is a case where there is valid conditioning on the regressors in the sense that they are weakly exogenous with respect to the cointegrating matrix B. In more general cases where the errors u_{1t} are temporally dependent and when there is feedback from u_{1t-j} to u_{2t} for $j \geqslant 1$ the appropriate augmented regression equation has coefficient nonlinearities and leads as well as lags of Δy_{2t}. In such cases an exact theory is much more complex and this paper does not deal with these cases.

4. DISCUSSION

This paper demonstrates that some exact finite sample distribution theory is possible for estimators in cointegrating regression models. Our results show that reduced rank regression estimators have Cauchy-like tails and no finite integer moments. Outliers can be expected to occur more frequently for this estimator than maximum likelihood and other efficient estimators that are based on the triangular system representation. The latter estimators are shown, at least in the particular system (17)–(18) considered here, to have finite sample distributions for which integer moments exist to order $T - m = T - n + r$, where T is the sample size, n is the total number of variables in the system, and r is the number of structural cointegrating equations.

It may be worth remarking that the outlier behavior of the reduced rank regression estimator is unrelated to the fact that the information content of the data is random in the limit as $T \to \infty$ for nonstationary data. Both the regression estimator and the reduced rank regression estimator have limiting mixed normal distributions that reflect this random limiting information feature of nonstationary data. Indeed, efficient cointegrating regression estimators and reduced rank regression estimators have equivalent limiting distributions. Yet the finite sample tail behavior of these estimators is very different.

Outlier behavior is only one characteristic of the finite sample distributions of $\hat{B}$ and $\tilde{B}$. We should therefore be careful to avoid concluding that the finite sample performance of $\hat{B}$ is inferior to that of estimators like $\tilde{B}$ solely because of tail behavior. Simultaneous equations theory is instructive in this regard. It is

known, for example, that the LIML estimator in a structural equation model has a distribution that is close to being symmetric in finite samples (producing near median unbiasedness in the estimator) and seems to approach its asymptotic normal distribution quite rapidly while still having no finite first moment. Moreover, the LIML estimator seems generally to have better finite sample characteristics than regression estimators such as 2SLS. Anderson (1982) and Phillips (1983) provide a detailed comparison of the distributions of LIML and 2SLS estimators in this context; and Hillier (1990) and Phillips (1990) explore reasons for the apparent superiority of LIML. The fact that the LIML estimator is invariant to the equation normalization is certainly one factor in determining its near median unbiasedness. In effect, no preference is given to a particular direction (or axis) in the optimization criterion by which the estimator is obtained. A similar factor comes into play in the reduced rank regression estimator $\hat{B}$, which as we have seen is invariant to the empirical normalization criteria (3) and (4) that are employed in obtaining $\hat{A}$. On the other hand, the maximum likelihood regression estimator $\tilde{B}$ relies on the explicit normalization (5) which gives a preference to the coordinates of y_{1t} in determining the value of $\tilde{B}$. Note, however, that in the case considered in Section 3 the distribution of $\tilde{B}$ is actually symmetric about B and hence $\tilde{B}$ is median unbiased. Since $\tilde{B}$ has finite integer moments to order $T - m$, $\tilde{B}$ is also an unbiased estimator of B in this case.

In triangular system cointegrating regression an exact theory of inference is also possible in simple models like (17) and (18). For testing linear hypotheses about the columns (or linear combinations of the columns) of the cointegrating matrix B, the Wald statistic constructed from the augmented regression estimator $\tilde{B}$ is proportional to an F variate, just as in the case of the multivariate linear regression model with fixed regressors and Gaussian errors. This test is analogous to the Hotelling's T^2 test in multivariate regression.

Cowles Foundation for Research in Economics, Yale University, Box 2125 Yale Station, New Haven, CT 06520-2125, U.S.A.
and
University of Auckland, Private Bag, Auckland, New Zealand

APPENDIX

The following two results will be used in the proof of Theorem 2.1.

5.1. LEMMA: *If A is an $n \times n$ positive definite random matrix with density function $f(A)$ then the joint density function of the eigenvalues $l_1, \ldots, l_n$ of A is*

$$\pi^{n^2/2}[\Gamma_n(n/2)]^{-1} \prod_{i<j}^{n} (l_i - l_j) \int_{O(n)} f(HLH')(\underline{dH})$$

where $L = \operatorname{diag}(l_1, \ldots, l_n)$, $l_1 > \cdots > l_n > 0$, and $(\underline{dH})$ is the invariant measure on $O(n)$ normalized so that $\int_{O(n)}(\underline{dH}) = 1$.

PROOF: See Muirhead (1982, Theorem 3.2.17).

5.2 LEMMA: *Let $X \in V_{r,n}$ and let (dX) denote the invariant measure on $V_{r,n}$ and $(\underline{dX})$ the corresponding normalized measure (so that $\int_{V_{r,n}} (\underline{dX}) = 1$). If we parameterize X as*

$$X = \begin{bmatrix} I_r \\ -R' \end{bmatrix} (I + RR')^{-1/2} L,$$

where R is an $r \times (n-r)$ matrix and $L \in O(r)$, then the following equation defines the transformation of measures

$$(\underline{dX}) = \left[\pi^{r(n-r)/2} \Gamma_r(r/2) \right]^{-1} \Gamma_r(n/2) |I_r + RR'|^{-n/2} dR(\underline{dL}),$$

where $(\underline{dL})$ denotes the invariant measure on $O(r)$ and $dR = \prod_{i,j} dr_{ij}$ is Lebesgue measure on $\mathbb{R}^{r(n-r)}$.

PROOF: We need to find the jacobian of the transformation $X \to (R, L)$ and this is most easily accomplished by using exterior algebra. We start by taking matrix differentials of the relation defining X in terms of R and L:

$$dX = \begin{bmatrix} 0 \\ -dR' \end{bmatrix} (I + RR')^{-1/2} L - \begin{bmatrix} I \\ -R' \end{bmatrix} (I + RR')^{-1/2} d\{(I + RR')^{1/2}\}$$

$$\times (I + RR')^{-1/2} L + \begin{bmatrix} I \\ -R' \end{bmatrix} (I + RR')^{-1/2} dL.$$

Next we construct an orthonormal complement to the matrix X, e.g.

$$X_{\perp} = \begin{bmatrix} R \\ I_m \end{bmatrix} (I_m + R'R)^{-1/2}, \qquad m = n - r,$$

and then $[X, X_{\perp}] \in O(n)$.

Now we have

$$X'_{\perp} dX = -(I_m + R'R)^{-1/2} dR' (I + RR')^{-1/2} L,$$

and if we take exterior products of the elements on the left-hand side we obtain (writing $[X, X_{\perp}]$ as $[x_1, x_2, \ldots, x_n]$)

$$(X'_{\perp} dX) := \bigwedge_{i=1}^{r} \bigwedge_{j>r} x'_j dx_i = |I_m + R'R|^{-r/2} |I_r + RR'|^{-m/2} \bigwedge_{i=1}^{r} \bigwedge_{j=1}^{m} dr_{ij}$$

$$= |I_r + RR'|^{-n/2} dR.$$

Next we consider

$$X' dX = -L'(I + RR')^{-1/2} R dR' (I + RR')^{-1/2} L - L'd\{(I + RR')^{-1/2}\}$$

$$\times (I + RR')^{-1/2} L + L' dL$$

$$= L' dL + \text{terms involving the differentials } dr_{ij}.$$

We now take the exterior products of the elements in the subdiagonal positions of the matrix $X' dX$, giving us

$$(X' dX) := \bigwedge_{i=1}^{r} \bigwedge_{j=i+1}^{r} x'_j dx_i$$

$$= \bigwedge_{i=1}^{r} \bigwedge_{j=i+1}^{r} l'_j dl_i + \text{terms involving the differentials } dr_{ij}$$

$$= (dL) + \text{terms involving the differentials } dr_{ij},$$

where $L = [l_1, \ldots, l_r]$. Combining $(X' \, dX)$ and $(X'_\perp \, dX)$ we obtain

$$(dX) := \bigwedge_{i=1}^{r} \bigwedge_{j=i+1}^{n} x'_j \, dx_i = (X' \, dX) \wedge (X'_\perp \, dX) = |I_r + RR'|^{-n/2} \, dR(dL).$$

Note that terms involving the differentials dr_{ij} in $(X' \, dX)$ contribute nothing to this last expression because $dR = \Lambda_{i,j} \, dr_{ij}$ is already a differential form of maximum degree.

The final expression above can be translated into an expression for the corresponding normalized measures by writing

$$(\underline{dX}) = \frac{\text{Vol}[O(r)]}{\text{Vol}[V_{r,n}]} \cdot |I + RR'|^{-n/2} \, dR(\underline{dL}),$$

where $\text{Vol}[O(r)] = \int_{O(r)} (\underline{dL})$ and $\text{Vol}[V_{r,n}] = \int_{V_{r,n}} (\underline{dX})$. Now, these volume integrals have values $\text{Vol}[O(r)] = 2^r \pi^{r^2/2} / \Gamma_r(r/2)$, and $\text{Vol}[V_{r,n}] = 2^r \pi^{rn/2} / \Gamma_r(n/2)$ (e.g., see Muirhead (1982, pp. 70–71)). In consequence, we obtain the lemma as stated. This generalizes a result given originally in the case where $r = 1$ in an earlier paper by the author (1984, p. 254).

5.3. PROOF OF THEOREM 2.1: Define $S = S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$. Let $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_n)$ be the diagonal matrix of the ordered eigenvalues $\hat{\lambda}_1 > \cdots > \hat{\lambda}_n > 0$ of S. Let C be the matrix of corresponding eigenvectors normalized as $C'C = I_n$, so that $S = C\hat{\Lambda}C'$. The transformation $S \to (\hat{\Lambda}, C)$ can be made $1:1$ by requiring that the first element in each column of C be nonnegative (e.g., see Muirhead (1982, p. 104)). Observe that $\hat{\Lambda}_r$ in (8) is the leading $r \times r$ submatrix of $\hat{\Lambda}$ corresponding to the r largest roots. Let C_1 be the corresponding submatrix of the eigenvector matrix C.

By Lemma 5.1 the joint density $f(S)$ of $S = C\hat{\Lambda}C'$ can be decomposed into the joint density of $(\hat{\Lambda}, C)$, whose probability element is

$$(32) \qquad \pi^{n^2/2} [\Gamma_n(n/2)]^{-1} \prod_{i<j} (\hat{\lambda}_i - \hat{\lambda}_j) f(C\hat{\Lambda}C') \, d\hat{\Lambda}(\underline{dC}),$$

where $(\underline{dC})$ represents the normalized invariant measure on $O(n)$ and $d\hat{\Lambda} = \prod_{i=1}^{n} d\hat{\lambda}_i$ is Lebesgue measure on the space of the eigenvalues. Next we partition the matrix C into the eigenvectors C_1 corresponding to $\hat{\Lambda}_r$ and the complementary set C_2, i.e. $C = [C_1, C_2]$. Note that if G $(n \times n - r)$ is any matrix chosen so that $[C_1, G]$ is orthogonal then we can write $C_2 = GK$ for some $K \in O(n-r)$. In this event we can decompose the invariant measure $(\underline{dC})$ on $O(n)$ into two factors as

$$(33) \qquad (\underline{dC}) = (\underline{dC_1})(\underline{dK})$$

where $(\underline{dC_1})$ is the normalized invariant measure on the Stiefel manifold $V_{r,n}$ and $(\underline{dK})$ is the normalized invariant measure on the orthogonal group $O(n-r)$. (The decomposition (33) was shown by Constantine and Muirhead (1976, Lemma 2.2, p. 374).)

Since C_1 is the matrix of eigenvectors corresponding to the largest r roots, $\hat{\Lambda}_r$, of $\hat{\Lambda}$, it is this matrix that, upon suitable normalization to accord with (3) and (4), produces the reduced rank regression estimator $\hat{A}$. Indeed, $\hat{A} = S_{11}^{-1/2} C_1$ satisfies both (3) and (4) and the optimization problem (2). Suppose, instead, we reparameterize C_1 so that it accords with the normalization $A' = [I, -B]$ given in (5). This can be achieved by writing C_1 in the form

$$(34) \qquad C_1 = \begin{bmatrix} I_r \\ -\hat{B}' \end{bmatrix} [I_r + \hat{B}\hat{B}']^{-1/2} L,$$

where $\hat{B}$ is the reduced rank regression estimator of B and $L \in O(r)$. Using Lemma 5.2 we find that the normalized measure $(\underline{dC_1})$ decomposes as follows:

$$(35) \qquad (\underline{dC_1}) = [\pi^{rm/2} \Gamma_r(r/2)]^{-1} \Gamma_r((r+m)/2) |I_r + \hat{B}\hat{B}'|^{-(r+m)/2} \, d\hat{B}(\underline{dL}),$$

where $d\hat{B} = \prod_{i,j} d\hat{b}_{ij}$ is Lebesgue measure on $\mathbb{R}^{rm}$, the support of $\hat{B}$, and $(\underline{dL})$ is the normalized invariant measure on $O(r)$.

We may now write the matrix C in (32) as $C = [C_1, C_2]$ with

$$C_1 = \begin{bmatrix} I_r \\ -\hat{B} \end{bmatrix} (I_r + \hat{B}\hat{B}')^{-1/2} L, \qquad C_2 = \begin{bmatrix} \hat{B} \\ I_{n-r} \end{bmatrix} (I_{n-r} + \hat{B}'\hat{B})^{-1/2} K.$$

Using (32), (33), and (35) we deduce that the probability element of $(\hat{\Lambda}, \hat{B}, L, K)$ is

$$\Gamma_r((r+m)/2) \left[\pi^{-n^2/2+rm/2} \Gamma_r(r/2) \Gamma_n(n/2) \right]^{-1} \prod_{i<j} (\hat{\lambda}_i - \hat{\lambda}_j)$$

$$\times f\left(C(\hat{B}, L, K) \hat{\Lambda} C(\hat{B}, L, K)' \right) |I_r + \hat{B}\hat{B}'|^{-(r+m)/2} \, d\hat{\Lambda} \, d\hat{B} (\underline{dL})(\underline{dK}).$$

Noting that $m = n - r$ and integrating over $\hat{\Lambda}$, L, and K we obtain the density of $\hat{B}$:

$$\mathrm{pdf}(\hat{B}) = \left[\pi^{-n^2/2+rm/2} \Gamma_r(r/2) \right]^{-1} |I_r + \hat{B}\hat{B}'|^{-(r+m)/2}$$

$$\times \int_{\hat{\Lambda}} \int_{L \in O(r)} \int_{K \in O(m)}$$

$$\times \prod_{i<j} (\hat{\lambda}_i - \hat{\lambda}_j) f\left(C(\hat{B}, L, K) \hat{\Lambda} C(\hat{B}, L, K)' \right) (\underline{dL})(\underline{dK}) \, d\hat{\Lambda}$$

as stated.

5.4. PROOF OF COROLLARY 2.3: Under Condition (D), dominated convergence applies and along the ray $\hat{B} = bB_0$ we have

$$\lim_{b \to \infty} \int_{0 < \hat{\Lambda} < I} \int_{L \in O(r)} \int_{K \in O(m)}$$

$$\times \prod_{i<j} (\hat{\lambda}_i - \hat{\lambda}_j) f\left(C(bB_0, L, K) \hat{\Lambda} C(bB_0, L, K)' \right) (\underline{dL})(\underline{dK}) \, d\hat{\Lambda}$$

$$= \int_{0 < \hat{\Lambda} < I} \int_{L \in O(r)} \int_{K \in O(m)}$$

$$\times \prod_{i<j} (\hat{\lambda}_i - \hat{\lambda}_j) f\left(C_0 \hat{\Lambda} C_0' \right) (\underline{dL})(\underline{dK}) \, d\hat{\Lambda}$$

$$= k_0, \qquad \text{say}$$

where $C_0 = \lim_{b \to \infty} C(\hat{B} = bB_0, L, K)$. Note that C lies in the compact space $O(n)$ and is continuous in the elements of $\hat{B}$, so that the limit C_0 exists. To find C_0 we work from the partition $C = [C_1, C_2]$ given in (8) and set $\hat{B} = bB_0$. We define $F = [F_1, F_2] \in O(r)$ where F_1 spans $\mathcal{N}(B_0')$ and F_2 spans $\mathcal{R}(B_0)$. Then $F'[I + b^2 B_0 B_0']F = \mathrm{diag}(I, I + b^2 F_2' B_0 B_0' F_2)$ and we can write C_1 as

$$C_1 = \begin{bmatrix} I_r \\ -bB_0' \end{bmatrix} F \left\{ F'[I + b^2 B_0 B_0'] F \right\}^{-1/2} F'L$$

$$= \begin{bmatrix} I_r \\ -bB_0' \end{bmatrix} F \, \mathrm{diag} \left\{ I, (I + b^2 F_2' B_0 B_0' F_2)^{-1/2} \right\} F'L$$

$$= \begin{bmatrix} I_r \\ -bB_0' \end{bmatrix} \left\{ F_1 F_1' + F_2 (I + b^2 F_2' B_0 B_0' F_2)^{-1/2} F_2' \right\} L$$

$$\to \begin{bmatrix} F_1 F_1' \\ -B_0' F_2 (F_2' B_0 B_0' F_2)^{-1/2} F_2' \end{bmatrix} L = C_{01}, \qquad \text{as} \qquad b \to \infty.$$

If $\dim\{\mathcal{N}(B_0')\} = 0$ then we replace $F_1 F_1'$ by the $r \times r$ zero matrix in the above formula for C_{01}.

Next define $G = [G_1, G_2] \in O(m)$ with $G_1 \in \mathcal{N}(B_0)$ and $G_2 \in \mathcal{R}(B_0')$. Then in a similar way, we get

$$
\begin{aligned}
C_2 &= \begin{bmatrix} bB_0 \\ I_m \end{bmatrix} \left(I_m + b^2 B_0' B_0 \right)^{-1/2} K \\
&= \begin{bmatrix} bB_0 \\ I_m \end{bmatrix} \left\{ G_1 G_1' + G_2 \left(I + b^2 G_2' B_0' B_0 G_2 \right)^{-1/2} G_2' \right\} K \\
&\rightarrow \begin{bmatrix} B_0 G_2 (G_2' B_0' B_0 G_2)^{-1/2} G_2' \\ G_1 G_1' \end{bmatrix} K = C_{02}, \qquad \text{as} \qquad b \rightarrow \infty.
\end{aligned}
$$

Again, we replace $G_1 G_1'$ by the $m \times m$ zero matrix in the above formula for C_{02} if $\dim \{ \mathcal{N}(B_0) \} = 0$. Using these results in the density (7) where $\hat{B} = b B_0$ for some constant matrix $B_0 \neq 0$ we have the following expansion as $b \rightarrow \infty$

$$
\begin{aligned}
\text{pdf} \left(\hat{B} = b B_0 \right) &= \left[\pi^{-n^2/2 + rm/2} \Gamma_r(r/2) \right]^{-1} [k_0 + o(1)] |I + b^2 B_0 B_0'|^{-(r+m)/2} \\
&= k_0 |I + b^2 B_0 B_0'|^{-(r+m)/2} \{ 1 + o(1) \},
\end{aligned}
$$

giving the stated result (11).

5.5. PROOF OF COROLLARY 2.4: Let $H_1 \in O(n)$ be an orthogonal matrix of order n. Under (14) and the spherical symmetry of ε_t we have the distributional equivalence $\Delta y_t = H_1 \Delta y_t$ for all $t \geqslant 1$ and a similar distributional equivalence between the levels of y_t and that of a vector random walk with rotated innovations, i.e. $y_t = \Sigma_1^t \varepsilon_j \equiv \Sigma_1^t H_1 \varepsilon_j = H_1 y_t$. The moment matrices S_{ij} $(i, j = 0, 1)$ that appear in the criterion (2) are, in consequence, also distributionally equivalent under rotations. Thus for any $H_1 \in O(n)$, we have $S_{00} = H_1' S_{00} H_1$, $S_{01} = H_1' S_{01} H_1$, $S_{11} = H_1' S_{11} H_1$, and finally

$$
S = S_{11}^{-1/2} S_{01} S_{00}^{-1} S_{01} S_{11}^{-1/2} \equiv H_1' S H.
$$

Thus, if $f(S)$ is the joint density function of S then $f(S) = f(H_1' S H_1)$ and $f(\cdot)$ is therefore a symmetric function of its matrix argument. It follows that in the general expression (7) for the density of $\hat{B}$ we can replace $f(C \hat{A} C')$ by $f(\hat{A})$. Using Lemma 5.1 and noting that the integrals of the normalized measures $(\underline{dL})$ and $(\underline{dK})$ over $O(r)$ and $O(m)$ in (7) are both unity, we deduce that

$$
\begin{aligned}
\text{pdf} (\hat{B}) &= \left[\pi^{-n^2/2 + rm/2} \Gamma_r(r/2) \right]^{-1} |I_r + \hat{B} \hat{B}'|^{-(r+m)/2} \pi^{-n^2/2} \Gamma_n(n/2) \\
&= \left[\pi^{rm/2} \Gamma_r(r/2) \right]^{-1} \Gamma_n(n/2) |I_r + \hat{B} \hat{B}'|^{-(r+m)/2}.
\end{aligned}
$$

Using the fact that $m = n - r$, we obtain the stated result (15).

It is also useful to derive the distributions of $\hat{A}$ and $\hat{\bar{A}} = \hat{A}(\hat{A}'\hat{A})^{-1/2}$ in this case. As in the proof of Theorem 2.1 we can write the estimator $\hat{A}$ in the form $\hat{A} = S_{11}^{-1/2} C_1$ where C_1 is the matrix of eigenvectors of S corresponding to $\hat{\Lambda}_r$. When $f(\cdot)$ is a symmetric function the joint density (32) splits into the product of the density of the roots $\hat{\Lambda}$ and the density of the eigenvector matrix C which is uniform on the group $O(n)$. In consequence, C_1 is uniform on $V_{r,n}$ and we have $C_1 \equiv H_1' C_1 H_2$ for any $H_1 \in O(n)$ and any $H_2 \in O(r)$. Thus,

$$
H_1' \hat{A} H_2 = H_1' S_{11}^{-1/2} C_1 H_2 = (H_1' S_{11} H_1)^{-1/2} H_1' C_1 H_2 \equiv S_{11}^{-1/2} C_1 = \hat{A}
$$

and $\hat{A}$ is therefore spherically symmetric. Next

$$
\hat{\bar{A}} = \hat{A}(\hat{A}'\hat{A})^{-1/2} \equiv H_1 \hat{A} H_2 (H_2' \hat{A}' \hat{A} H_2)^{-1/2} = H_1' \hat{A} (\hat{A}'\hat{A})^{-1/2} H_2 = H_1' \hat{\bar{A}} H_2
$$

and $\hat{\bar{A}}' \hat{\bar{A}} = I_r$. It follows that $\hat{\bar{A}}$ is uniform on $V_{r,n}$, since the invariance of the distribution under $H_1 \in O(n)$ and $H_2 \in O(r)$ is a defining characteristic of the uniform distribution on $V_{r,n}$.

5.6. PROOF OF THEOREM 3.1: To simplify the derivations we first transform the augmented regression model (20) and (18) as follows:

$$\Sigma_{11\cdot2}^{-1/2}y_{1t} = \left(\Sigma_{11\cdot2}^{-1/2}B\Sigma_{22}^{1/2}\right)\left(\Sigma_{22}^{-1/2}y_{2t}\right) + \left(\Sigma_{11\cdot2}^{-1/2}D\Sigma_{22}^{1/2}\right)\left(\Sigma_{22}^{-1/2}\Delta y_{2t}\right) + \Sigma_{11\cdot2}^{-1/2}u_{1\cdot2t},$$

$$\Sigma_{22}^{-1/2}\Delta y_{2t} = \Sigma_{22}^{-1/2}u_{2t}.$$

With this standardized system as

(36) $\bar{y}_{1t} = \bar{B}\bar{y}_{2t} + \bar{D}\Delta\bar{y}_{2t} + \bar{u}_{1\cdot2t},$

(37) $\Delta\bar{y}_{2t} = \bar{u}_{2t},$

and note that $(\bar{u}'_{1\cdot2t}, \bar{u}'_{2t}) \equiv$ iid $N(0, I_n)$. If $\check{B}$ is the OLS estimator of $\bar{B}$ in (36) then

(38) $\check{E} = \check{B} - \bar{B} = \Sigma_{11\cdot2}^{-1/2}(\tilde{B} - B)\Sigma_{22}^{1/2} = \Sigma_{11\cdot2}^{-1/2}\tilde{E}\Sigma_{22}^{1/2}$

so that the density of $\tilde{E}$ is easily deduced from that of $\check{E}$ by reversing the standardizing transformation (38). In what follows we will omit the upper bars in our notation and simply assume that these transformations have been performed.

The conditional density of $\check{E} = U'_{1\cdot2}Q_{\Delta}Y_2(Y'_2Q_{\Delta}Y_2)^{-1}$ given $\mathscr{F}_2 = \sigma(u_{2t}: t = 1, \ldots, T)$ is the Gaussian distribution

(39) $\text{pdf}\left(\check{E}|\mathscr{F}_2\right) = (2\pi)^{-rm/2}|F|^{-r/2}\,\text{etr}\left\{-(1/2)\check{E}'\check{E}F^{-1}\right\}$

where $F = (Y'_2Q_{\Delta}Y_2)^{-1}$. To simplify (39) we write $Y_2 = LU_2$ where $U'_2 = [u_{21}, \ldots, u_{2T}]$ and L is the $T \times T$ matrix

$$L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Next, let $\Xi_2 = U_2(U'_2U_2)^{-1/2}$ and

$$Y'_2Q_{\Delta}Y_2 = U'_2L'\left(I - U_2(U'_2U_2)^{-1}U'_2\right)LU_2 = (U'_2U_2)^{1/2}K(U'_2U_2)^{1/2},$$

where $K = \Xi'_2L'(I - \Xi_2\Xi'_2)L\Xi_2 = K(\Xi_2)$. Let $R = U'_2U_2$ and note that R is central Wishart with T degrees of freedom and covariance matrix I_m. We write $R \equiv W_m(T, I_m)$ and note that the density of R is

(40) $\text{pdf}(R) = \left[2^{mT/2}\Gamma_m(T/2)\right]^{-1}\text{etr}\left\{-(1/2)R\right\}|R|^{(T-m-1)/2}.$

The distribution of Ξ_2 is independent of R and is uniform on the Stiefel manifold $V_{m,T}$. We write the normalized invariant measure of this manifold as $(d\Xi_2)$.

From (39) and (40) the conditional density of $\check{E}$ given Ξ_2 is

(41) $\text{pdf}\left(\check{E}|\Xi_2\right) = 2^{-m(T+r)/2}\pi^{-rm/2}\left[\Gamma_m(T/2)\right]^{-1}|K(\Xi_2)|^{-r/2}$

$$\times \int_{R>0}|R|^{r/2+(T-m-1)/2}\,\text{etr}\left\{-(1/2)R\right\}$$

$$\cdot\,\text{etr}\left\{-(1/2)\check{E}'\check{E}R^{1/2}K(\Xi_2)^{-1}R^{1/2}\right\}dR.$$

Next observe that the distribution of $\check{E}$ in the standardized model (36) is spherical and, in particular, $\check{E} \equiv \check{E}H_3$ for any $H_3 \in O(m)$. It follows that we may replace $\check{E}$ in (41) by $\check{E}H_3$ and integrate over the normalized group $O(m)$. Now using (dH_3) to signify the normalized measure on the group we have

(42) $\int_{O(r)}\text{etr}\left\{-(1/2)H'_3\check{E}'\check{E}H_3R^{1/2}K(\Xi_2)^{-1}R^{1/2}\right\}(dH_3)$

$$= {}_0F_0^{(m)}\left(R^{1/2}K(\Xi_2)^{-1}R^{1/2}, -\check{E}'\check{E}\right)$$

(e.g., Muirhead (1982, Theorem 7.3.3, p. 260)), where $_0F_0^{(m)}$ is a hypergeometric function of two matrix arguments. Since this function depends only on the latent roots of the argument matrices we may replace the right side of (42) by the simpler expression

$$_0F_0^{(m)}\big(RK(\Xi_2)^{-1}, -\check{E}'\check{E}\big).$$

Hence (41) has the alternate form

$$\text{pdf}\big(\check{E}|\Xi_2\big) = 2^{-m(T+r)/2}\pi^{-rm/2}\big[\Gamma_m(T/2)\big]^{-1}|K(\Xi_2)|^{-r/2}$$

$$\times \int_{R>0} |R|^{r/2+(T-m-1)/2}\,\text{etr}\{-(1/2)R\}$$

$$_0F_0^{(m)}\big(RK(\Xi_2)^{-1}, -\check{E}'\check{E}\big)\,dR,$$

and using the matrix Laplace transform of the $_0F_0^{(m)}$ function (e.g., Muirhead (1982, Theorem 7.3.4, p. 260)) we obtain

$$\text{pdf}\big(\check{E}|\Xi_2\big) = \pi^{-mr/2}\big[\Gamma_m(T/2)\big]^{-1}|K(\Xi_2)|^{-r/2}\Gamma_m((T+r)/2)\,_1F_0^{(m)}$$

$$\times((T+r)/2); K(\Xi_2)^{-1}, -\check{E}'\check{E})$$

$$= \pi^{-mr/2}\big[\Gamma_m(T/2)\big]^{-1}\Gamma_m((T+r)/2)|K(\Xi_2)|^{-r/2}$$

$$\times \int_{O(m)}\,_1F_0\big((T+r)/2; -H_3K(\Xi_2)^{-1}H_3'\check{E}'\check{E}\big)(dH_3)$$

$$= \pi^{-mr/2}\big[\Gamma_m(T/2)\big]^{-1}\Gamma_m((T+r)/2)|K(\Xi_2)|^{-r/2}$$

$$\times \int_{O(m)}\big|I_m + H_3K(\Xi_2)^{-1}H_3'\check{E}'\check{E}\big|^{-(T+r)/2}(dH_3).$$

The density of $\check{E}$ is obtained by integrating over the manifold $V_{m,T}$, leading to

$$\text{pdf}(\check{E}) = \pi^{-mr/2}\big[\Gamma_m(T/2)\big]^{-1}\Gamma_m((T+r)/2)\int_{V_{m,T}}|K(\Xi_2)|^{-r/2}$$

$$\times \int_{O(m)}\big|I_m + H_3K(\Xi_2)^{-1}H_3'\check{E}'\check{E}\big|^{-(T+r)/2}(dH_3)(d\Xi_2).$$

The probability density of the estimation error in the unstandardized model is obtained from the above expression by reversing the standardizing transformations. Using $\tilde{E}$ to denote the estimation error in the unstandardized model we have $\tilde{E} = \Sigma_{11\cdot2}^{1/2}\check{E}\Sigma_{22}^{-1/2}$ and then

$$\text{pdf}(\tilde{E}) = \pi^{-mr/2}\big[\Gamma_m(T/2)\big]^{-1}\Gamma_m((T+r)/2)|\Sigma_{11\cdot2}|^{-m/2}|\Sigma_{22}|^{r/2}$$

$$\times \int_{V_{m,T}}|K(\Xi_2)|^{-r/2}$$

$$\times \int_{O(m)}\big|I_m + H_3K(\Xi_2)^{-1}H_3'\Sigma_{22}^{1/2}\tilde{E}'\Sigma_{11\cdot2}^{-1}\tilde{E}\Sigma_{22}^{1/2}\big|^{-(T+r)/2}(dH_3)(d\Xi_2),$$

giving the stated result (23).

To prove part (b) of Theorem 3.1 we simply note that $E_N = U_{1\cdot2}'Q_\Delta Y_2(Y_2'Q_\Delta Y_2)^{-1/2}$, whose conditional distribution given $\mathscr{F}_2$ is $N(0, \Sigma_{11\cdot2} \otimes I_m)$. Since this is independent of $\mathscr{F}_2$ it is also the unconditional distribution, giving the required result.

5.7. Proof of Theorem 3.3: Define the variate

$$X = D_1U_{1\cdot2}Q_\Delta Y_2(Y_2'Q_\Delta Y_2)^{-1}d_2/\big\{d_2'(Y_2'Q_\Delta Y_2)^{-1}d_2\big\}^{1/2}$$

$$= D_1(\tilde{B} - B)d_2/\big\{d_2'(Y_2'Q_\Delta Y_2)^{-1}d_2\big\}^{1/2}.$$

Conditional on $\mathscr{F}_2$ we have the distribution

$$(43) \qquad X|_{\mathscr{F}_2} \equiv N(0, D_1 \Sigma_{11 \cdot 2} D_1').$$

Now $T\bar{\Sigma}_{11 \cdot 2} = Y_1' Q_* Y_1 = U_{1 \cdot 2}' Q_* U_{1 \cdot 2}$, so that $T\bar{\Sigma}_{11 \cdot 2}|_{\mathscr{F}_2} \equiv W_r(T - 2m, \Sigma_{11 \cdot 2})$, i.e., central Wishart with covariance matrix $\Sigma_{11 \cdot 2}$ and degrees of freedom $T - 2m$. It follows that

$$(44) \qquad TD_1 \bar{\Sigma}_{11 \cdot 2} D_1'|_{\mathscr{F}_2} = W_q(T - 2m, D_1 \Sigma_{11 \cdot 2} D_1').$$

Combining (43) and (44) and setting $S_X = D_1 \bar{\Sigma}_{11 \cdot 2} D_1'$ we have the Hotelling's T^2 variate

$$X' S_X^{-1} X|_{\mathscr{F}_2} \equiv \frac{Tq}{N - q + 1} F_{q, n - q + 1},$$

where $N = T - 2m$. Since the distribution is independent of $\mathscr{F}_2$, it is also the unconditional distribution.

Now the Wald statistic is

$$W = \left(D_1 \bar{B} d_2 - d \right)' \left(D_1 \bar{\Sigma}_{11 \cdot 2} D_1 \right)^{-1} \left(D_1 \bar{B} - d \right) / \left\{ d_2' (Y_2' Q_\Delta Y_2)^{-1} d_2 \right\}$$

and under H_0 we can write this as $W = X' S_X^{-1} X$. Hence, under H_0 we have

$$W \equiv X' S_X^{-1} X \equiv \frac{Tq}{N - q + 1} F_{q, N - q + 1}.$$

Thus, W is proportional to an $F_{q, N - q + 1}$ variate, as required.

REFERENCES

AHN, S. K., AND G. C. REINSEL (1988): "Nested Reduced Rank Autoregressive Models for Multiple Time Series," *Journal of the American Statistical Association*, 83, 849–856.
——— (1990): "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.
ANDERSON, T. W. (1982): "Some Recent Developments on the Distributions of Single-Equation Estimators," in *Advances in Econometrics*, ed. by W. Hildenbrand. Cambridge: Cambridge University Press.
CAPPUCCIO, N., AND D. LUBIAN (1992): "The Relationships Among Some Estimators of the Cointegrating Coefficient," Mimeographed, Universita di Padova.
CONSTANTINE, A. G., AND R. J. MUIRHEAD (1976): "Asymptotic Expansions for Distributions of Latent Roots in Multivariate Analysis," *Journal of Multivariate Analysis*, 6, 369–391.
CORBAE, D., S. OULIARIS, AND P. C. B. PHILLIPS (1990): "A Reexamination of the Permanent Income Hypothesis Using Frequency Domain Regressions," Mimeographed, University of Iowa.
ERDÉLYI, A. (1956): *Asymptotic Expansions*. New York: Dover.
GREGORY, A. W. (1991): "Testing for Cointegration in Linear Quadratic Models," Mimeographed, Queens University.
HARGREAVES, C., AND A. SETIABUDI (1992): "Properties of Five Cointegration Estimators," Mimeographed, Australian National University.
HERZ, C. S. (1955): "Bessel Functions of Matrix Argument," *Annals of Mathematics*, 61, 474–523.
HILLIER, G. H. (1990): "On the Normalization of Structural Equations: Properties of Direction Estimators," *Econometrica*, 58, 1181–1194.
JAMES, A. T. (1954): "Normal Multivariate Analysis and the Orthogonal Group," *Annals of Mathematical Statistics*, 25, 40–75.
JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.
MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. New York: Wiley.
PARK, J. Y. (1990): "Maximum Likelihood Estimation of Simultaneous Cointegrated Models," Mimeographed, Cornell University.

PARK, J. Y., AND M. OGAKI (1991): "Inference in Cointegrated Models Using VAR Prewhitening to Estimate Short Run Dynamics," Rochester Center for Economic Research Working Paper No. 281.

PHILLIPS, P. C. B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, ed. by M. D. Intriligator and Z. Griliches. Amsterdam: North-Holland, Ch. 8 and pp. 449–516.

——— (1984): "The Exact Distribution of LIML: I," *International Economic Review*, 25, 249–261.

——— (1986a): "The Distribution of FIML in the Leading Case," *International Economic Review*, 27, 239–243.

——— (1986b): "The Exact Distribution of the Wald Statistic," *Econometrica*, 54, 881–895.

——— (1987): "An Everywhere Convergent Series Representation of Hotelling's Generalized T_0^2," *Journal of Multivariate Analysis*, 21, 238–248.

——— (1990): "T. W. Anderson's Contributions to the Study of Structural Equation Estimation," in *The Collected Works of T. W. Anderson*, ed. by G. Styan. New York: Wiley, pp. 1643–1649.

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

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with $I(1)$ Processes," *Review of Economic Studies*, 57, 99–125.

PHILLIPS, P. C. B., AND M. LORETAN (1991): "Estimating Long-run Economic Equilibria," *Review of Economic Studies*, 59, 407–436.

SARGAN, J. D. (1988): "The Existence of Moments of Estimated Reduced Form Coefficients," in *Contributions to Econometrics*, ed. by E. Maasoumi. Cambridge: Cambridge University Press, Ch. 6, pp. 133–157.

TODA, H., AND P. C. B. PHILLIPS (1993): "Vector Autoregression and Causality: A Theoretical Overview and Simulation Study," *Econometric Reviews* (forthcoming).