

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 999

NOTE: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

THE TAIL BEHAVIOR OF MAXIMUM LIKELIHOOD
ESTIMATORS OF COINTEGRATING COEFFICIENTS
IN ERROR CORRECTION MODELS

by

Peter C. B. Phillips

October 1991

**THE TAIL BEHAVIOR OF MAXIMUM LIKELIHOOD ESTIMATORS
OF COINTEGRATING COEFFICIENTS
IN ERROR CORRECTION MODELS**

by

Peter C. B. Phillips*

*Cowles Foundation for Research in Economics
Yale University*

September, 1991

*My thanks go to Glenna Ames for wordprocessing and to the NSF for research support under Grant no. SES 882-1180.

O. ABSTRACT

This paper derives exact finite sample distributions of maximum likelihood estimators of the cointegrating coefficients in error correction models. The distributions are derived for the leading case where the variables in the system are independent random walks. But important aspects of the theory, in particular the tail behavior of the distributions, continue to apply when the system is cointegrated. The reduced rank regression estimator is shown to have a distribution with Cauchy-like tails and no finite moments of integer order. The maximum likelihood estimator of the coefficients in the triangular system representation has 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 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.

JEL Classification: 211

Keywords: Hypergeometric function; Invariant measure; Leading case; 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. The most commonly used method in practical work at present is reduced rank regression -- see Johansen (1988, 1991) and also Ahn-Reinsel (1988, 1990). This method applies maximum likelihood to a systems error correction model (ECM) formulated with vector autoregression (VAR) transient 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 (see Park, 1990) for a demonstration), but they are not equivalent in finite samples. This is because the triangular system representation explicitly incorporates identifying conditions that isolates a set of full rank integrated regressors and identifies the cointegrating equations as individual structural equations. The reduced rank regression approach, on the other hand, employs a normalization rule that uniquely determines only the cointegration space not the individual structural cointegrating relations. This difference does have important effects, some of which will be explored in the present paper. In addition to these procedures the main alternative efficient methods that are now available are: fully modified least squares (Phillips-Hansen, 1990) canonical cointegrating regression (Park, 1991), spectral regression (Phillips, 1991), lead + lag augmented regression (Phillips-Loretan (1991), Saikkonen (1991), Stock-Watson (1991)) and a three step estimator of Engle-Yoo (1989). With the exception of the latter all of these use the triangular representation.

Some attempts have been made to evaluate these different procedures by simulation. Recent work by Corbae-Ouliaris-Phillips (1990), Gregory (1991), Park-Ogaki (1991), Stock-Watson (1991) and Toda-Phillips (1991) is relevant in this connection. One common feature to emerge from these simulation studies is that the 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.

The present paper seeks to provide an analytical basis for understanding this phenomena. The approach adopted is related to earlier work by the author (1984, 1986) on the exact distribution of simultaneous equations estimators. In that work, some carefully chosen simplifications in the structural simultaneous equations model facilitate analytic derivations without sacrificing important elements of generality. The same is true in the present time series context. Here we deal with a "leading case" model where the data is generated by n independent Gaussian random walks. We derive the exact finite sample distribution of the reduced rank

regression estimator for this leading case. The distribution is shown to be matrix Cauchy and has no finite integer order moments. Cauchy-like tail behavior persists even when the leading case hypothesis is relaxed, thereby providing an analytical explanation for the outlier behavior observed in the aforementioned simulation studies. Maximum likelihood and other efficient estimators that are based on the triangular representation are shown to have quite different finite sample tail behavior. Unlike the reduced rank regression estimator, these estimators typically 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.

Throughout this paper we use the symbol "=" to signify equivalence in distribution. $O(n) = \{H(n \times n) : H'H = I_n\}$ 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 $\Gamma_n(a) = \pi^{n(n-1)/4} \prod_{i=1}^n \Gamma(a - (1/2)(i-1))$, $\text{Re}(a) > (n-1)/2$, is the multivariate gamma function.

2. AN EXACT FINITE SAMPLE THEORY FOR THE REDUCED RANK REGRESSION ESTIMATOR IN THE LEADING CASE

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

$$(1) \quad \begin{aligned} J(L)\Delta y_t &= \Gamma A' y_{t-1} + \varepsilon_t \quad (t = 1, \dots, T); \quad J(L) = \sum_0^{k-1} J_i L^i, \quad J_0 = I \\ \varepsilon_t &\equiv \text{iid } N(0, \Sigma_\varepsilon). \end{aligned}$$

In (1) A' is an $r \times n$ matrix of cointegrating vectors. Our discussion will concentrate on what we will describe as the leading case of (1) where

$$(2) \quad \Gamma = 0, \quad J(L) = I, \quad \Sigma_\varepsilon = I, \quad y_0 = 0.$$

Under (2) the generating mechanism of y_t is a set of n independent Gaussian random walks initialized at the origin at $t = 0$, i.e.

$$(3) \quad \Delta y_t = \varepsilon_t; \quad y_0 = 0, \quad \varepsilon_t \equiv \text{iid } N(0, I_n).$$

We shall investigate the finite sample properties of the reduced rank regression estimator of the "cointegrating" matrix A' in (1) under the leading case hypothesis (2) i.e. when the true generating mechanism is (3). As in the case of the LIML estimator explored in earlier work (1984, 1985), the exact distribution that is obtained in this leading case provides the first term in the series representation of the exact distribution in the general case viz. the case where the auxiliary parameters are not all zero as in (2).

Let $H_1 \in O(n)$ be an orthogonal matrix of order n . We have the distributional equivalence

$$(4) \quad \Delta y_t = H_1 \Delta y_t; \forall t \geq 1,$$

and a similar distributional equivalence between the output of (3) and that of a vector random walk with rotated innovations i.e.

$$(5) \quad y_t = \Sigma_1' \epsilon_j = \Sigma_1' H_1 \epsilon_j = H y_t.$$

Now suppose that the matrix A in (1) is estimated by reduced rank regression. It may be assumed that $r \geq 1$ is prescribed or it could be chosen by the likelihood ratio trace statistic of Johansen (1988). (Even under (3) this outcome of the test occurs with positive probability in finite samples.) The estimator of A satisfies the optimization problem

$$(6) \quad \hat{A} = \operatorname{argmin}_A |S_{00} - S_{01} A (A' S_{11} A)^{-1} A' S_{10}|$$

(see Johansen (1988), p. 234) subject to a suitable normalization condition on \hat{A} . We employ the condition

$$(7) \quad \hat{A}' \hat{A} = I_r$$

since it is most convenient for our theory (Johansen (1988) uses $\hat{A}' S_{11} \hat{A} = I_r$). The S_{ij} in (6) 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} ($j = 1, \dots, k-1$) 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)).

In view of (4) and (5) the moment matrices are distributionally equivalent under rotations. That is for any $H_1 \in O(n)$ we have

$$(8) \quad S_{00} = H_1' S_{00} H_1, S_{01} = H_1' S_{01} H_1, S_{11} = H_1' S_{11} H_1.$$

The distribution of the criterion function in (6) is therefore invariant under these transformations. It follows that the distribution of the optimizing value of A is also invariant. Moreover, for any $H_2 \in O(r)$ we find that the normalization condition (7) is also preserved if the rows of A' are rotated by H_2' i.e.

$$H_2' \hat{A}' \hat{A} H_2 = H_2' \hat{A}' H_1 H_1' \hat{A} H_2 = I_r.$$

Under these two transformations the optimizing value \hat{A} in (6) translates as

$$(9) \quad \hat{A} \rightarrow H_1' \hat{A} H_2; H_1 \in O(n), H_2 \in O(r).$$

But since the distributions of the moment matrices S_{ij} and the criterion in (6) are invariant under H_1 , and the normalization is invariant under H_2 it follows that the distribution of \hat{A} is invariant under the simultaneous action of H_1 and H_2 in (9). That is

$$(10) \quad \hat{A} = H_1' \hat{A} H_2$$

for all $H_1 \in O(n)$ and all $H_2 \in O(r)$.

The normalization (7) ensures that \hat{A} is an element of the Stiefel manifold $V_{r,n} = \{A (n \times r) : A'A = I_r\}$. The distributional equivalence (10) establishes that \hat{A} has a distribution on $V_{r,n}$ that is invariant under the simultaneous orthogonal transformations H_1 and H_2 in (9). But the only distribution on $V_{r,n}$ with this property is normalized Haar measure on $V_{r,n}$ i.e. the uniform distribution on $V_{r,n}$ (e.g. see James (1954)). Thus, the exact distribution of \hat{A} on $V_{r,n}$ is uniform.

Let us now change the normalization in (7) to the requirement that the leading submatrix of A' is the identity. That is, suppose we set

$$(11) \quad A' = [I_r \ -B].$$

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

$$(12) \quad y_{1t} = B y_{2t} + u_{1t}$$

where $y_t' = [y_{1t}', y_{2t}']$ is a partition conformable with (11), y_{2t} is a full rank integrated process and u_{1t} is a stationary error. Note that the explicit form (12) is stronger than the normalization (7). In particular, the form of (12) explicitly recognizes a subvector, y_{2t} , of full rank integrated regressors. Moreover, (12) involves r restrictions per equation (i.e. $r-1$ exclusion restrictions and one unit normalization restriction) giving r^2 restrictions *in toto*. By contrast, (7) involves only $r(r+1)/2$ restrictions *in toto*, an insufficient number to identify individual structural equations when $r > 1$. Indeed, as remarked above \hat{A}' is only determined up to an orthogonal rotation $H_2 \in O(r)$. This is, of course, sufficient to identify the cointegration space, which is the principal goal of reduced rank regression. But it is insufficient to identify individual structural relations such as those given explicitly in (12).

Once A is estimated by reduced rank regression, *ex post* estimation of the submatrix B in (11) is always possible. Thus, partitioning A' and \hat{A}' conformably with (11) as $A' = [A_1, A_2]$ we have $B = -A_1^{-1}A_2$ with the corresponding estimator

$$(13) \quad \hat{B} = -\hat{A}_1^{-1}\hat{A}_2.$$

Observe that \hat{B} is invariant to rotations of the cointegration space defined by A' i.e. \hat{B} is invariant under the translation $A' \rightarrow H_2 A'$ for $H_2 \in O(r)$. Hence, \hat{B} is uniquely determined from \hat{A} . The finite sample distribution of \hat{B} is easily deduced from that of \hat{A} . We use the following result.

LEMMA (Phillips, 1989). *If the matrix variate X ($n \times r$) is spherical (i.e. $X = C_1XC_2$ where $C_1 \in O(n)$, $C_2 \in O(r$)) and partitioned as*

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \begin{matrix} r \\ m = n-r \end{matrix}$$

then the matrix quotient variate $R = X_2X_1^{-1}$ is distributed as matrix Cauchy with density

$$\text{pdf}(R) = \left[\pi^{m/2} \Gamma_r(r/2) \right]^{-1} \Gamma_r((r+m)/2) |I_r + R'R|^{-(r+m)/2},$$

where $\Gamma_r(\cdot)$ is the multivariate gamma function. \square

Since \hat{A} is uniform on $V_{r,n}$ and thereby spherically distributed we deduce from this lemma and the form of (13) that \hat{B} is distributed as matrix Cauchy. Collecting these results together we have:

THEOREM 1: *For the leading case (i.e. data generated by (3)) the finite sample distribution of the reduced rank regression estimator \hat{A} of A in the ECM (1) is the uniform distribution on the manifold $V_{r,n}$. The exact distribution of the implied estimator \hat{B} of the structural coefficient submatrix B in (12) is matrix Cauchy with probability density*

$$(14) \quad \text{pdf}(B) = \left[\pi^{r(n-r)/2} \Gamma_r(r/2) \right]^{-1} \Gamma_r(n/2) |I_r + BB'|^{-n/2}. \quad \square$$

REMARKS

(i) Observe that if $r = 1$, $m = n-1$ the distribution (14) has the familiar multivariate Cauchy form, viz.

$$(15) \quad \text{pdf}(b) = \left[\pi^{m/2} \Gamma(m/2) \right]^{-1} \Gamma((m+1)/2) (1 + b'b)^{-(m+1)/2}$$

where $B = b'$ ($1 \times m$) in this case.

(ii) Spherical distributions like (14) and (15) are preserved under marginalization. In particular, individual elements of \hat{B} have univariate Cauchy distributions and, consequently, no finite integer order moments. This latter property continues to apply when the leading case hypothesis (2) is relaxed. The Cauchy distribution then becomes the leading term in the full series representation of the density in the general case with nonzero auxiliary parameters. The presence of this term in the series causes the tail behavior of the resulting density to be of the Cauchy type, thereby resulting in the nonexistence of integer moments of any order.

This goes some way to explain the presence of the extreme outliers that have been observed in simulation studies of the reduced rank regression estimator.

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.

$$(16) \quad y_{1t} = By_{2t} + u_{1t},$$

$$(17) \quad \Delta y_{2t} = u_{2t}$$

where $u_t = (u_{1t}', u_{2t}')' = \text{iid } N(0, \Sigma_u)$. This system is, in fact, a specialization of (1) with

$$J(L) = \begin{bmatrix} r & m \\ 0 & 0 \\ 0 & I_m \end{bmatrix}, \Gamma = \begin{bmatrix} I_r \\ 0 \end{bmatrix}, A = [I, -B]$$

and with

$$u_{1t-1} = -e_{1t}, u_{2t} = e_{2t}.$$

As in Phillips (1991), generalizations may be considered in which u_t is generated by a parametric linear process. But the above prototype triangular system is all that is required for the purpose of our illustration of the tail behavior of estimates of B .

Our interest is in the maximum likelihood estimator \hat{B} of B in (16). As shown in the (1991) paper, \hat{B} is equivalent to the OLS estimator of B in the augmented regression equation

$$(18) \quad y_{1t} = By_{2t} + D\Delta y_{2t} + u_{1,2t} \quad (t = 1, \dots, T)$$

where

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

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 . In conventional regression notation \hat{B} has the form

$$(19) \quad \bar{B} = Y_1' Q_{\Delta} Y_2 (Y_2' Q_{\Delta} Y_2)^{-1}.$$

Let us now suppose that the true generating mechanism of the data is the system of independent random walks given in (3) above. Let \mathcal{F}_2 be the σ field generated by $\{\epsilon_{2t} : t = 1, \dots, T\}$. Then conditional on \mathcal{F}_2 , \bar{B} in (19) is linear in Y_1 and hence Gaussian. The conditional density of \bar{B} is

$$(20) \quad \text{pdf}(B|\mathcal{F}_2) = (2\pi)^{-m/2} |F|^{-r/2} \text{etr}\left\{-\frac{1}{2} B' B F^{-1}\right\}$$

where

$$F = (Y_2' Q_{\Delta} Y_2)^{-1} Y_2' Q_{\Delta} V Q_{\Delta} Y_2 (Y_2' Q_{\Delta} Y_2)^{-1}$$

and

$$V = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & \cdots & 3 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & 2 & 3 & \cdots & T \end{bmatrix}$$

The form of (20) follows directly from the density of $Y' = N(0, I, \otimes V)$.

To simplify (20) we write $Y_2 = L E_2$ where $E_2' = [\epsilon_{21}, \dots, \epsilon_{2T}]$ and

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

Next, let $\Xi_2 = E_2(E_2'E_2)^{1/2}$ and

$$\begin{aligned} Y_2' Q_{\Delta} Y_2 &= E_2' L' (I - E_2(E_2'E_2)^{-1} E_2') L E_2 \\ &= (E_2'E_2)^{1/2} K (E_2'E_2)^{1/2}, \end{aligned}$$

where

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

Then

$$F = (E_2'E_2)^{-1/2} G(\Xi_2) (E_2'E_2)^{-1/2},$$

where

$$G(\Xi_2) = K^{-1}\Xi_2'L'(I - \Xi_2\Xi_2')LL'(I - \Xi_2\Xi_2')L\Xi_2K^{-1} = J'LL'J, \text{ say.}$$

Let $R = E_2'E_2$ and note that R is central Wishart with T degrees of freedom and covariance matrix I_m . We write $R = W_m(T, I_m)$ and note that the density of R is

$$(21) \quad \text{pdf}(R) = \left[2^{mT/2}\Gamma_m(T/2)\right]^{-1} \text{etr}\{-(1/2)R\} |R|^{(T-m-1)/2}.$$

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 (20) and (21) the conditional density of \tilde{B} given Ξ_2 is

$$(22) \quad \text{pdf}(B|\Xi_2) = 2^{-m(T+r)/2} \pi^{-m/2} \left[\Gamma_m(T/2)\right]^{-1} |G(\Xi_2)|^{-r/2} \int_{R>0} |R|^{r/2+(T-m-1)/2} \text{etr}\{-(1/2)R\} \\ \cdot \text{etr}\{-(1/2)B'BR^{1/2}G(\Xi_2)^{-1}R^{1/2}\} dR.$$

Next observe that the distribution of \tilde{B} is spherical and, in particular, $\tilde{B} = \tilde{B}H_3$ for any $H_3 \in O(m)$. It follows that we may replace B in (22) by BH_3 and integrate over the normalized group $O(r)$. Now using (dH_3) to signify the normalized measure on the group we have:

$$(23) \quad \int_{O(r)} \text{etr}\{-(1/2)H_3'B'BH_3R^{1/2}G(\Xi_2)^{-1}R^{1/2}\} (dH_3) = {}_0F_0^{(m)}(R^{1/2}G(\Xi_2)^{-1}R^{1/2}, -B'B),$$

(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 (23) by the simpler expression

$${}_0F_0^{(m)}(RG(\Xi_2)^{-1}, -B'B).$$

Hence (22) has the alternate form

$$\text{pdf}(B|\Xi_2) = 2^{-m(T+r)/2} \pi^{-m/2} \left[\Gamma_m(T/2)\right]^{-1} |G(\Xi_2)|^{-r/2} \int_{R>0} |R|^{r/2+(T-m-1)/2} \text{etr}\{-(1/2)R\} \\ \cdot {}_0F_0^{(m)}(RG(\Xi_2)^{-1}, -B'B) dR$$

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

$$\begin{aligned}
\text{pdf}(B | \Xi_2) &= \pi^{-mr/2} [\Gamma_m(T/2)]^{-1} |G(\Xi_2)|^{-r/2} \Gamma_m((T+r)/2) {}_1F_0^{(m)}((T+r)/2; G(\Xi_2)^{-1}, -B'B) \\
&= \pi^{-m(r+T)/2} [\Gamma_m(T/2)]^{-1} \Gamma_m((T+r)/2) |G(\Xi_2)|^{-r/2} \\
&\quad \int_{O(m)} {}_1F_0^{(m)}((T+r)/2; -H_3 G(\Xi_2)^{-1} H_3' B' B) (dH_3) \\
&= \pi^{-m(r+T)/2} [\Gamma_m(T/2)]^{-1} \Gamma_m((T+r)/2) |G(\Xi_2)|^{-r/2} \\
&\quad \int_{O(m)} |I_m + H_3 G(\Xi_2)^{-1} H_3' B' B|^{-(T+r)/2} (dH_3).
\end{aligned}$$

The density of \bar{B} is obtained by integrating over the manifold $V_{m,T}$ leading to the following result.

THEOREM 2. *For the leading case (i.e. data generated by (3)) the finite sample distribution of the maximum likelihood estimator \bar{B} of B in the triangular cointegrated system (16) and (17) is given by*

$$\begin{aligned}
(24) \quad \text{pdf}(B) &= \pi^{-mr/2} [\Gamma_m(T/2)]^{-1} \Gamma_m((T+r)/2) \\
&\quad \int_{V_{m,T}} |G(\Xi_2)|^{-r/2} \int_{O(m)} |I_m + H_3 G(\Xi_2)^{-1} H_3' B' B|^{-(T+r)/2} (dH_3) (d\underline{\Xi}_2) \quad \square
\end{aligned}$$

REMARKS

(i) The matrix variate \bar{B} with density (24) is spherically symmetric in the sense that $\bar{B} = C_1 \bar{B} C_2$ with $C_1 \in O(r)$, $C_2 \in O(m)$. This is readily seen from the form of \bar{B} in (19), since $C_1 Y_1' = Y_1'$ and $C_2 Y_2' = Y_2'$; and it is also apparent from the form of the density (24). Since \bar{B} is spherically symmetric all marginal distributions of \bar{B} have the same form. Note that when $r = m = 1$ the density (24) is

$$(25) \quad \text{pdf}(b) = \pi^{-1/2} \Gamma(T/2)^{-1} \Gamma((T+1)/2) \int_{V_{1,T}} g(\Xi_2)^{-1/2} [1 + g(\Xi_2)^{-1} b^2]^{-(T+1)/2} (d\underline{\Xi}_2)$$

which is a mixture of scalar t -variates with T degrees of freedom.

More generally when $r = 1$ and $m \geq 1$ we have the following density for $\bar{B} = \bar{b}'$ ($1 \times m$)

$$\begin{aligned}
(26) \quad \text{pdf}(b) &= \pi^{-m/2} \Gamma_m(T/2)^{-1} \Gamma_m((T+1)/2) \\
&\quad \int_{V_{m,T}} |G(\Xi_2)|^{-1/2} \int_{O(m)} |I_m + H_3 G(\Xi_2)^{-1} H_3' b b'|^{-(T+1)/2} (dH_3) (d\underline{\Xi}_2) \\
&= \pi^{-m/2} [\Gamma(T-m+1)/2]^{-1} \Gamma((T+1)/2) \\
&\quad \int_{V_{m,T}} |G(\Xi_2)|^{-1/2} \int_{O(m)} |I_m + H_3 G(\Xi_2)^{-1} H_3' b b'|^{-(T+1)/2} (dH_3) (d\underline{\Xi}_2)
\end{aligned}$$

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

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

$$(27) \quad \text{pdf}(b_1) = \pi^{-1/2} [\Gamma(T-m+1)/2]^{-1} \Gamma((T-m+2)/2) \\ \int_{V_{m,T}} \int_{O(m)} g_{11,2}(\Xi_2, H_3)^{1/2} \left[1 + g_{11,2}(\Xi_2, H_3) b_1^2 \right]^{-(T-m+2)/2} (dH_3)(d\Xi_2)$$

where $g_{11,2}(\Xi_2, H_3)$ is defined by

$$g_{11,2} = g_{11} - g_{12} G_{22}^{-1} g_{21}$$

and the partitioned matrix elements come from

$$H_3 G(\Xi_2)^{-1} H_3' = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & G_{22} \end{bmatrix}.$$

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 $|b_1|$ 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. Since

$$\left[1 + g_{11,2}(\Xi_2, H_3)^{-1} b_1^2 \right]^{-(T-m+2)/2} = g_{11,2}(\Xi_2, H_3)^{(T+1)/2} |b_1|^{-(T-m+2)} \left\{ 1 + O(b_1^{-2}) \right\}$$

we obtain the expansion

$$(28) \quad \text{pdf}(b) = c |b_1|^{-(T-m+2)} \left\{ 1 + O(b_1^{-2}) \right\}$$

for large $|b_1|$, where

$$c = \pi^{-1/2} \Gamma((T-m+1)/2)^{-1} \Gamma((T-m+2)/2) \int_{V_{m,T}} \int_{O(m)} g(\Xi_2, H_3)^{T/2} (d\Xi_2)(dH_3).$$

From (28) it is apparent that the maximal moment exponent of \bar{b} is $T-m+1$ and integer moments of \bar{b}_1 exist to order $T-m$. The same result holds for an arbitrary element of the matrix variate \bar{B} .

(iii) Using arguments similar to those that lead to Theorem 2 we can establish that the OLS estimator of B in (16) has an exact finite sample distribution that is of the same form as (24). The only difference arises in the expression for $G(\Xi_2)$, which for OLS is given by

$$G(\Xi_2) = (\Xi_2' L' L \Xi_2)^{-1} \Xi_2' L' L L' L \Xi_2 (\Xi_2' L' L \Xi_2)^{-1}.$$

Marginal densities of individual components of the OLS estimator have expansions of the form (28). Thus these distributions also possess finite integer moments to order $T-m$.

Estimators that involve corrections to the OLS estimator, such as the fully modified OLS estimator of Phillips-Hansen (1991) and the canonical cointegrating regression estimator of Park (1991), similarly can be expected to have finite integer moments to order $T-m$.

4. CONCLUSION

This paper shows that an exact distribution theory is possible for a variety of estimators in cointegrating regressions for the leading case of data from independent random walks. In such cases cointegrating regressions are, of course, spurious in the sense of Granger-Newbold (1974). The present study therefore complements the author's (1986b) earlier work on the asymptotics of spurious regression. However, the results of this paper reach further than the spurious regression model. The leading case distributions given here in (14) and (24) are also relevant in models where there is cointegration. For in that case these distributions furnish leading terms in series representations of the distributions that apply in the presence of cointegration. The leading terms then determine the tail behavior and moment existence criteria for the distributions under cointegration.

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 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.

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.
- Ahn, S. K. and G. C. Reinsel (1990). "Estimation for partially nonstationary multivariate autoregressive models," *Journal of the American Statistical Association*, 85, 813-823.
- 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.
- Engle, R. F. and S. S. Yoo (1989). "Cointegrated time series: A survey with new results," mimeographed, UC San Diego.
- Granger, C. W. J. and P. Newbold (1974). "Spurious Regressions in Econometrics," *Journal of Econometrics* 74, 111-120.
- Gregory, A. W. (1991). "Testing for cointegration in linear quadratic models," mimeographed, Queens University.
- 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.
- Johansen, S. (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," University of Copenhagen, to appear in *Econometrica*, 1991.
- 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. (1991). "Canonical cointegrating regressions," *Econometrica* (forthcoming).
- 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. (1984). "The exact distribution of LIML: I," *International Economic Review*, 25, 249-261.
- Phillips, P. C. B. (1986a). "The Distribution of FIML in the leading case," *International Economic Review*, 27, 239-243.
- Phillips, P. C. B. (1986b). "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311-340.
- Phillips, P. C. B. (1989). "Spherical matrix distributions and Cauchy quotients," *Statistics and Probability Letters*, 8, 51-53.
- Phillips, P. C. B. (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.
- Saikkonen, P. (1991). "Asymptotically efficient estimation of cointegration regressions," *Econometric Theory*, 7, 1-21.

Stock, J. H. and M. W. Watson (1991). "A simple estimator of cointegrating vectors in higher order integrated systems," mimeographed, University of California, Berkeley.

Toda, H. and P. C. B. Phillips (1991). "Vector autoregression and causality: A theoretical overview and simulation study," mimeographed, Yale University.