On the Meaning of Certain
Cointegration Tests

by

Phoebus J. Dhrymes, Columbia University

Discussion Paper No. 704

ON THE MEANING OF CERTAIN COINTEGRATION TESTS*

Phoebus J. Dhrymes†

Columbia University

08/27/94

Abstract

In this paper we examine (some of) the cointegration tests suggested in Engle and Granger (1987), henceforth EG, those suggested by Johansen (1988), (1991), henceforth J, and those suggested in Dhrymes (1994b), henceforth D. We also explore the relations among the models underlying the various procedures. We find that the tests suggested by Johansen cannot possibly test either for the presence of cointegration or for its rank, but they can determine the cointegration vectors, if cointegration is **known** to hold and its **rank is known** as well.

Key words: Integrated Processes, Stationary Processes, Collinearity, Cointegration, Cointegration tests.

1 Introduction and Preliminaries

This paper narrowly focuses on some of the suggestions in EG, J and the tests in D. In Dickey, Jansen and Thornton (1991) (DJT) there is a good review of the procedures in EG and J, so that we shall not focus on them in great detail unless, as in the case of J, the argument warrants it.

In all three approaches the basic problem is that we are confronted with a sequence $\{X_{t\cdot} : t \geq 1\}$ which is known to be $I(1)$, and where $X_{t\cdot}$ is a q-element **row** vector. The question posed is whether or not it is

† This is a preliminary verision and is not to be quoted, except by permission of the author. Comments, however, are welcome.

cointegrated. For a definition of cointegration, rank thereof etc. see, for example, Dhrymes (1994). The underlying models are

$$\begin{aligned}
(I-L)X_{t\cdot}' &= A(L)\epsilon_{t\cdot}', \quad A(L)=\sum_{j=0}^{k} A_j L^j, \quad \text{Dhrymes} \\
(I-L)X_{t\cdot}' &= C(L)\epsilon_{t\cdot}', \quad C(L)=\sum_{j=0}^{\infty} C_j L^j, \quad \text{Engle-Granger} \\
\Pi(L)X_{t\cdot}' &= \epsilon_{t\cdot}', \qquad\quad \Pi(L)=\sum_{j=1}^{p} \Pi_j L^j, \quad \text{Johansen.}
\end{aligned} \tag{1}$$

In the first two, the ϵ-sequence is taken to be a $MWN(\Sigma)$, i.e. a multivariate white noise process with mean zero and covariance matrix $\Sigma > 0$. In either case if the argument required it, the authors would not be averse to asserting that higher moments existed as well. In the third case, J, the ϵ-sequence is asserted to be a **Gaussian** $MWN(\Sigma)$.

The first two formulations are essentially the same, in practical terms, since in D the order of the moving average, k, need not be known a priori. Moreover, it is shown therein that the EG formulation is not very apt, given the problem to be investigated. If it is taken seriously, i.e. if it is applied *verbatim* in conjunction with the usual definition of cointegration, **it cannot admit of cointegration**. See the examination of the example given in EG, which makes this point in Dhrymes (1994a). Essentially, the problem is the presence of the requisite initial conditions, in the absence of which no such discussions may proceed. The nature of the test in D, however, may be viewed as in the spirit of the original EG intent, if not necessarily the original exposition. The third (J) formulation is somewhat different, in that it neither contains nor is contained, as a special case, in the other two formulations.

2 Models and Implied Tests

In the first formulation (D), a test for cointegration is immediately **available** once we establish that

$$\begin{aligned}
\mathrm{Cov}(X_{t\cdot}') &= \sum_{r=0}^{t-1} S_{0,r}\Sigma S_{0,r}', \quad \text{if } t \leq k \\
&= (t-k)S_{0,k}\Sigma S_{0,k}' + \sum_{r=0}^{k-1} S_{0,r}\Sigma S_{0,r}', \quad \text{if } \ t \geq k+1, \text{ or}
\end{aligned}$$

$$= t\Phi + D, \quad \Phi \ = \ S_{0,k}\Sigma S_{0,k}',$$

$$D \ = \ \sum_{r=0}^{k-1}\left(S_{0,k}\Sigma S_{0,k}' - S_{0,r}\Sigma S_{0,r}'\right), \tag{2}$$

where

$$S_{i,j} = \sum_{s=i}^{j} A_s, \quad i,j = 0,1,2,\ldots,k, \quad S_{0,0} = A_0 = I_q. \tag{3}$$

Since, **by the definition of cointegration**, a necessary and sufficient (nas) condition for cointegration of rank r, is the existence of a $q \times r$ matrix B of rank r, such that the *cointegral vector*, $Z_{t\cdot} = X_{t\cdot}B$, has a stationary covariance matrix (or more precisely is covariance stationary), it is clear that B must be in the **null space** of $S_{0,k}\Sigma S_{0,k}'$ **but not** in that of D, and the latter must of rank at least r. Hence, **in this formulation** the (a) set of cointegrating vectors is simply the basis of the null space of $S_{0,k}\Sigma S_{0,k}'$. Since $\Sigma > 0$, we may also state that **the (a) set of cointegrating vectors is simply the basis of the null space of** $S_{0,k}$. Thus, in this framework we have cointegration **if and only if** there is **at least one zero** among the roots of $S_{0,k}\Sigma S_{0,k}'$ and the cointegrating rank is simply **the number of such zero roots.**

The instrumentality of carrying out this test is the estimator

$$M_{T,(1)} = \frac{1}{T}\sum_{t=1}^{T}\frac{X_{t\cdot}'X_{t\cdot}}{t}, \quad \text{or, alternatively,} \quad M_{T,(2)} = \frac{1}{T^2}\sum_{j=1}^{T}X_{t\cdot}'X_{t\cdot}. \tag{4}$$

The first is an **asymptotically unbiased** estimator of the desired matrix, while the second is an **asymptotically biased** estimator of Φ, the limit of its expectation being $(1/2)\Phi$. The problem with both estimators is that under the hypothesis that the X-sequence is $I(1)$, their limiting distribution is nonstandard and needs to be established and tabulated.

In the D context to test whether the X-process is $I(0)$ is equivalent to testing whether **all roots of** Φ are null. An alternative recommended in Dhrymes (1994b) is to actually test for zero roots in

$$M_{T,(0)} = \frac{1}{T}\sum_{t=1}^{T}X_{t\cdot}'X_{t\cdot}, \tag{5}$$

under the hypothesis that the X-process is $I(0)$. The argument here is that **if** zero roots are detected in this context, it would mean that we have collinearity among the underlying variables, and whatever else is obtained by a proper cointegration test may simply be a reflection of this collinearity, and the well known phenomenon noted first by Stone (1947), and discussed in the author's early work, Dhrymes (1970). The desirability of this procedure, as a way of ruling out contaminated results is strengthened by

Proposition 1. Consider the matrices $M_{T,(i)}$, $i = 0, 1$, of the previous discussion; the matrix $M_{T,(0)} - M_{T,(1)}$, conditionally on the sample, is positive semidefinite.

Proof: Neglecting the factor $(1/T)$, we find

$$X'X - X'N^2X = X'DX, \quad N = \mathrm{diag}(\frac{1}{t^{1/2}}), \quad D = \mathrm{diag}(\frac{t-1}{t}),$$

for $t = 1, 2, \ldots, T$. Since $D \geq 0$, it follows that $X'DX \geq 0$.

q.e.d.

Corollary 1. If, conditionally on the sample, $M_{T,(0)}$ has r_0 roots that obey $\mu_j^{(0)} \leq \delta$ then there exist $r_1 \geq r_0$ characteristic roots of $M_{T,(1)}$, say $\mu_s^{(1)}$ such that $\mu_s^{(1)} \leq \delta$, $s = 1, 2 \ldots r_1 \geq r_0$.

Proof: We note, Bellman (1960) p. 115, that since

$$X'X = X'N^2X + X'DX, \tag{6}$$

and both matrices on the right are at least positive semidefinite, it follows that

$$\mu_j^0 \geq \mu_j^{(1)}, \quad j = 1, 2, 3, \ldots, q. \tag{7}$$

It follows, therefore, that

$$\mu_j^{(1)} \leq \mu_j^{(0)} \leq \delta, \quad j = 1, 2, 3, \ldots, r_0. \tag{8}$$

q.e.d.

Remark 1. Evidently, the implications to be derived from Proposition 1 and Corollary 1 above are suggestive, not conclusive.

We now consider the model put forth by J. In this formulation we begin with a $MAR(p)$ model, generally known in macroeconomics as a VAR, i.e.

$$\Pi(L)X'_{t\cdot} = \epsilon'_{t\cdot}, \quad \Pi(L) = \sum_{j=0}^{p} \Pi_j L^j, \ \Pi_0 = I_q, \tag{9}$$

where the ϵ-process is a $MWN(\Sigma)$ **Gaussian** process. After considerable manipulation, **but without any transformation of** the error process or of the underlying model, we may write

$$(I-L)X'_{t\cdot} = -\Pi(I)X'_{t-1\cdot} - \sum_{j=1}^{p} \Pi_j^*(I-L)X'_{t-j\cdot} + \epsilon'_{t\cdot}, \tag{10}$$

where, of course, $\Pi(I)$ is the **operator** $(\sum_{i=0}^{p} \Pi_i)I$. In current paractice, the literature does not distinguish between the operator $\Pi(I)$, and the **matrix** $\Pi(I) = \sum_{i=0}^{p} \Pi_i$. Following this custom, we shall not distinguish between the two in subsequenct discussion.[1] It is important to understand that the conceptual basis of this approach is different from those in the D or EG formulation, and we shall return to this issue below. For the moment, let us note that if the series **is cointegrated of rank** r **then** $\Pi(I)$ must be a $q \times q$ matrix of rank r. This is so since the left and right members of Eq. (10) are all $I(0)$, except possibly for the first term of the right member. Since we **assert** that the series X is **cointegrated of rank** r then the term $\Pi(I)X_{t\cdot}'$ must be, at most, a **multiple of the cointegral** vector $Z_{t-1\cdot} = X_{t-1\cdot}B$, where B is a $q \times r$ matrix of rank r, whose **columns** are the **cointegrating** vectors. Thus, by the singular value decomposition theorem, we may write

$$\Pi(I) = \Gamma B', \text{ where } \Gamma,\ B \text{ are (both) } q \times r \text{ matrices of rank } r. \tag{11}$$

It bears stressing that **in order to write** $\Pi(I)$ as in Eq. (11) we **must assert** cointegration of rank r. Otherwise $\Pi(I)$ is just another parameter matrix, whose rank can be determined by a test performed utilizing the **estimated** parameter (or a function thereof) as a test statistic. Under the **maintained hypothesis** that we **have cointegration of rank** r, the J procedure maximizes the LF with respect to the ancillary parameters Σ, Π_j^*, $j = 1, 2, \ldots p$, and Γ, thus obtaining the concentrated LF

$$L^*(B) = -\frac{qT}{2}[\ln(2\pi) + 1] - \frac{T}{2}\ln|S|, \quad S = \frac{1}{T}W'N^*W. \tag{12}$$

To justify the representation in Eq. (12), define

$$\begin{aligned}
y_{t\cdot} &= (I - L)X_{t\cdot} - X_{t-1\cdot}\Pi(I)' = \Delta X_{t\cdot} - X_{t-1\cdot}B\Gamma'; \\
x_{t\cdot} &= (\Delta X_{t-1\cdot}, \Delta X_{t-2\cdot}, \Delta X_{t-3\cdot}, \ldots, \Delta X_{t-n+1\cdot}), \quad \epsilon_{t\cdot} = u_{t\cdot}, \\
\Pi^* &= (-\Pi_1^{*'}, -\Pi_2^{*'}, \ldots, \Pi_{n-1}^{*'})', \quad W \;=\; N\Delta P, \quad V \;=\; NP_{-1}, \\
Y &= (y_{t\cdot}), \quad X = (x_{t\cdot}), \quad U = (u_{t\cdot}), \; t = 1, 2, 3, \ldots T, \\
P &= (X_{t\cdot}), \quad P_{-i} = (X_{t-i\cdot}), \quad t = 1, 2, 3, \ldots, T, \quad i = 1, 2, 3, \ldots n - 1 \\
p &= \mathrm{vec}(P), \quad p_{-i} \;=\; \mathrm{vec}(P_{-i}), \quad N \;=\; I_T - X(X'X)^{-1}X'. \\
N^* &= I_T - NP_{-1}B(B'P_{-1}'NP_{-1}B)^{-1}B'P_{-1}'N, \qquad\qquad (13) \\
y &= \mathrm{vec}(Y), \quad \pi^* = \mathrm{vec}(\Pi^*), \quad u = \mathrm{vec}(U), \quad \gamma = \mathrm{vec}(\Gamma'),
\end{aligned}$$

[1] Note that in our notation we distringuish between the identity operator I and the $q \times q$ **identity matrix** I_q.

so that we can write

$$y_{t\cdot} = x_{t\cdot}\Pi^* + u_{t\cdot}, \quad Y = X\Pi^* + U, \quad \text{or} \quad y = (I_q \otimes X)\pi^* + u, \tag{14}$$

for a single observation and the entire sample, respectively. The concentrated LF is obtained by partially maximizing the loglikelihood (LF) of the T observations

$$L_0 = -\frac{qT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{tr}(Y - X\Pi^*)\Sigma^{-1}(Y - X\Pi^*)'. \tag{15}$$

From Dhrymes (1984) p. 106, we have

$$\mathrm{tr}(Y - X\Pi^*)\Sigma^{-1}(Y - X\Pi^*)' = [y - (I_q \otimes X)\pi^*]'(\Sigma^{-1} \otimes I_T)[y - (I_q \otimes X)\pi^*],$$

and solving the first order condition $(\partial L_0/\partial\pi^*) = 0$, we find

$$\hat{\pi}^* = [I_q \otimes (X'X)^{-1}X']y. \tag{16}$$

Inserting this in Eq. (12), we obtain the concentrated LF

$$L_1 = -\frac{qT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}y'(\Sigma^{-1} \otimes N)y. \tag{17}$$

Next, we note that

$$y = \Delta p - (I_q \otimes P_{-1}B)\gamma, \tag{18}$$

and the concentrated LF may be rendered as

$$\begin{aligned} L_1 &= -\frac{qT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| \\ &\quad -\frac{1}{2}[\Delta p - (I_q \otimes P_{-1}B)\gamma]'(\Sigma^{-1} \otimes N)[\Delta p - (I_q \otimes P_{-1}B)\gamma]; \end{aligned} \tag{19}$$

solving the first order conditions $(\partial L_1/\partial\gamma) = 0$, we find

$$\hat{\gamma} = [I_q \otimes (B'P_{-1}'NP_{-1}B)^{-1}B'P_{-1}'N]\Delta p. \tag{20}$$

Noting that

$$(I_q \otimes N)[\Delta p - (I_q \otimes P_{-1}B)\hat{\gamma}] = (I_q \otimes N^*)(I_q \otimes N)\Delta p, \tag{21}$$

we may write the (once again) concentrated LF as

$$\begin{aligned} L_2 &= -\frac{qT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| \\ &\quad -\frac{1}{2}[(I_q \otimes N^*N)\Delta p]'(\Sigma^{-1} \otimes I_T)[(I_q \otimes N^*N)\Delta p]. \end{aligned} \tag{22}$$

Using the results in Dhrymes (1984) p. 106 and "rematricizing" the last expression in the LF, we may rewrite the latter as

$$L_2 = -\frac{qT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{T}{2}\mathrm{tr}\Sigma^{-1}S, \quad S = \frac{1}{T}(N\Delta P)'N^*(N\Delta P). \quad (23)$$

Maximizing L_2 with respect to the elements of Σ^{-1}, we obtain

$$\frac{\partial L_2}{\partial \mathrm{vec}(\Sigma^{-1})} = \frac{T}{2}\mathrm{vec}(\Sigma)' - \frac{T}{2}\mathrm{vec}(S) = 0, \quad \text{or} \quad \hat{\Sigma} = S. \quad (24)$$

Inserting this in Eq. (23), we find the ultimately concentrated LF, L^*, as it appears in Eq. (12). This is now to be **maximized with respect to** B, i.e. we are to find a B that **maximizes the LF**. It bears stressing again that, **in the absence of prior restrictions**, to complete the estimation procedure implicit in the J formulation **nothing more and nothing less** is required by that formulation than simply to maximize the function in Eq. (12) with respect to the elements of the matrix B. For the moment, let us proceed with this unrestricted framework and see where it leads us. Without loss of generality, and for notational simplicity, we neglect the factor $(1/T)$ in the definition of S and, equivalently, pose the problem: minimize S with respect to B. Now, put

$$D(B) = T^q|S| = |W'W - W'V(W'W)^{-1}V'W|,$$

and note that, after some manipulation, we obtain

$$\begin{aligned} D(B) &= |W'W|\,|B'V'VB|^{-1}\,|B'(F-H)B|, \\ F &= V'V, \quad H = V'W(W'W)^{-1}W'V. \end{aligned} \quad (25)$$

Since $D(B)$ is **homogeneous of degree zero in** B, the problem is not well defined, and a **normalization is required**. A convenient normalization in this context is $B'V'VB = I_q$. It is a rather straightforward procedure to show that the minimizing value of B is given by the matrix of the characteristic vectors of H in the **metric of** F.[2] Thus, the

[2] This may be illustrated quite simply when B is taken to be a single vector, say b. The problem then is to *maximize* $b'Hb$ subject to $b'Fb = 1$. Set up the Lagrangian $\mathcal{L} = b'Hb + \lambda(1 - b'Fb)$. The first order conditions yield $(\lambda F - H)b = 0$, and the solution is evidently given as one of the characteristic vectors of H in the metric of F; since we wish to maximize, we choose as the solution the vector, say b^*, that corresponds to the **largest** root, λ^*. Thus we obtain, for the orginal problem,

$$b^{*\prime}(F-H)b^* = 1 - \lambda^*,$$

which yields, **for** $b'(F-H)b$, the minimum we wish. Since the equation that defines this solution is the characteristic equation of H in the metric of F, which has q roots and corresponding characteristic vectors, the reader would easily grasp the conclusion that in a problem that involves all such roots and vectors, such as the minimization of a determinant, the solution would involve all characteristic roots and vectors as well.

minimum value of $D(B)$, obtained **without any prior restrictions,** is given by

$$\min_{B'V'VB=I_q} D(B) = |W'W| \prod_{j=1}^{q}(1-\lambda_j), \tag{26}$$

where the λ_j are the **characteristic roots** of H, **in the metric of** F. The question that arises now, in view of the fact that we are dealing with the **loglikelihood** function, relates to the magnitude of these roots. For if some roots are greater than one the **maximum** of the LF is not well defined, and thus the problem has no solution. To investigate this issue consider the matrix

$$G = (P_{-1}, \Delta P)'N(P_{-1}, \Delta P) = \begin{bmatrix} V'V & V'W \\ W'V & W'W \end{bmatrix},$$

and note that since neither P_{-1}, nor ΔP is in the null space of N, $G \geq 0$. Defining

$$K = \begin{bmatrix} I & -V'W(W'W)^{-1} \\ 0 & I \end{bmatrix},$$

we conclude from

$$KGK' = \begin{bmatrix} F-H & 0 \\ 0 & W'W \end{bmatrix} \tag{27}$$

that $F-H = V'V - V'W(W'W)^{-1}W'V$ is also **positive semidefinite.** Hence, the characteristic roots of H **in the metric** of F, obey $\lambda_j \in [0,1)$. **If** the matrix $F-H$ is **singular** then at least one of the λ_j is equal to **one**, and hence the maximum of the LF is not well defined.[3] Zero roots, if any, are due to the singularity of H. Thus, provided no **root** is **unity**, the LF maximized without prior restrictions yields

$$\max_{B,\Sigma,\Pi^*,\Gamma} L_3 = -\frac{qT}{2}\ln[(2\pi)+1] - \frac{1}{T}\ln|W'W| - \frac{T}{2}\sum_{j=1}^{q}\ln(1-\lambda_j). \tag{28}$$

Since $1-\lambda_j > 0$, if there are **zero** roots their inclusion or exclusion is a matter of complete indifference from the point of view of estimation since the LF is flat relative to this eventuality. Thus, whether to choose characteristic vectors corresponding to zero roots is akin to the situation encountered in many nonlinear problems where the likelihood function

[3] The argument in this connection is to be understood as an a.c. (almost certain) argument. If one wished, one could convert all such entities to their parametric equivalents, if they are ascertained to exist. There is no difficulty with $W'W$ which involves **stationary** entities only. There may be a problem with $V'V$, since $V = NP_{-1}$, and P_{-1} involves elements that may not be stationary; on the other hand $V'V/T^2$, as well as H/T^2 converge, at least in distribution, and the a.c. argument may be made without difficulty.

may have more than one maxima. Strictly speaking this results **in lack of identifiction**! Thus, the problem as posed by J **does not admit of a solution**, unless **restrictions are imposed** a priori. In fact the only unambiguous result that may be obtained from this procedure, **absent any prior restrictions**, is in the case where it is conlcuded that B is **nonsingular, without the occasion of zero roots**!

But even if we arrive at the nonsingularity of B without the intervention of zero roots, we still face a problem of degeneracy. To see this, return to the definition of S and note that if B is nonsingular, the determinant $|S|$ **does not involve** B, hence B is **arbitrary** subject to its being nonsingular. Moreover, returning to Eq. (20) and noting the nonsingularity of B we obtain, after rematricizing the expression for $\hat{\Gamma}$,

$$\hat{\Pi}(I) \;=\; \hat{\Gamma}'B \;=\; (P_{-1}'NP_{-1})^{-1}P_{-1}'N\Delta P, \tag{29}$$

whose right member has the appearance of a "2SLS-like" estimator of $\Pi(I)$ in the pseudo-relation $\Delta P = NP_{-1}\Pi(I) + error$.

If now we **impose the prior restrictions** (a) that the series **is** cointegrated and (b) cointegration is of **rank** $r < q$, **we have a well defined problem** and an unambiguous solution, B_r, **even if one of the roots**, whose corresponding characteristic vector is a column of B_r, **is zero.** The reason why this is so is because in this context, **we have no idea** what cointegration implies **beyond the requirement that** $\Pi(I)$ be singular and of rank $r < q$, and the procedure has, as required, produced an estimator of $\Pi(I)$, viz. $\hat{\Gamma}\hat{B}_r'$, which is a $q \times q$ matrix of rank r, **irrespective of whether some** of the characteristic roots in question **are null**. Thus, the "admissibility" of null characteristic roots is not an issue in this problem. The difficulty is that, absent prior information on the **rank and existence of cointegration**, the estimation problem is not well defined, and thus **cannot yield** relevant tests about the existence of cointegration or the magnitude of its rank!

Remark 2. It was the practice, until the advent of the J procedure, for empirical researchers to carry out a Dickey-Fuller (DF) test, or an "augmented" DF test, and then proceed to test for cointegration and estimate the underlying parameters. The problem was that the several tests suggested in Engle and Granger (1987) (EG), and subsequent modifications that produced a number of variants, were either too cumbersome, or not very well argued, or their sampling properties were not available. For whatever reason, the current practice is to employ almost exclusively the J procedure. This is aided by the fact that the relevant test statistics involve the chi-squared distribution, which is well tabulated. An additional factor, no doubt, is that the last stage of maximization procedure, i.e. the minimization of $|S|$ with respect to B involves certain mechanics which are almost identical to some part of the process of obtaining **pairs of canonical variates.** No doubt this similarity adds to the intuitive appeal of the J procedure, even though this similarity is coincidental, and completely irrelevant. The canonical variate problem is the following: given two sets of (zero mean) random vectors, say x^1 and x^2, and each of dimension q, find r **pairs** of linear combinations of x^1 and x^2, respectively, that exhibit maximal correlation, subject to their being **pairwise uncorrelated** (with other pairs of linear combinations), each element of the pair having unit variance. If the linear combinations in question are given by $G_r'x^1$ and $H_r'x^2$, respectively, their constituent vectors, $g_{\cdot i}, h_{\cdot i}$, $i = 1, 2, \ldots, r$, are the solutions to the system

$$\begin{bmatrix} \lambda\Sigma_{11} & -\Sigma_{12} \\ -\Sigma_{21} & \lambda\Sigma_{22} \end{bmatrix} \begin{pmatrix} g \\ h \end{pmatrix} = 0, \tag{30}$$

which leads to

$$\left(\lambda^2\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) g = 0, \quad h = \frac{1}{\lambda}\Sigma_{22}^{-1}\Sigma_{21}g. \tag{31}$$

The first equation in Eq. (31) defines the vectors $g_{\cdot i}$ as the characteristic vectors corresponding to the characteristic roots λ_i^2 of the appropriate determinantal equation. The second equation defines the vectors $h_{\cdot i}$. Moreover, it can be shown that the correlation between the components of each pair may be displayed in the diagonal matrix, Λ_r, below, i.e.

$$EG_r'x^1x^{2'}H_r = \Lambda_r = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r), \tag{32}$$

it being understood that λ_i is the correlation between $g_{\cdot i}'x^1$ and $h_{\cdot i}'x^2$. It is **only** the part associated with the first equation in Eq. (31) that resembles the mechanics of the J procedure. In the case of canonical variates one **rejects vectors** corresponding to zero roots, either because of the second equation in Eq. (31) above, which would involves division by zero, or because the roots of the first equation may be interpreted as the **square of the correlations** between the two components of a pair of canonical variates. In the J procedure, however, **there is no reason inherent in the formulation of the problem** that may lead one to reject characteristic vectors corresponding to zero roots, particularly since the operation involved in the **second equation** of Eq. (31) **does not enter the J procedure.** The latter is **not designed to, nor does it** obtain the canonical variates to which we may transform NP_{-1} and $N\Delta P$. It would be difficult to visualize such an operation unless one were to argue that NP_{-1} represents observations on a **stationary**

process. Thus, the fact that (part of) the estimation mechanics of the J procedure coincide in with (part of) the mechanics of obtaining canonical variates is completely coincidental and irrelevant to the operation, or interpretation of the results obtained in the J context. As noted above, zero roots present identification problems. As in the simultaneous equations literature identification problems may be "solved" only by means of prior restrictions. There is no more justification in rejecting all zero roots than in rejecting only some of them. For example, if in the J context there are $s > 1$ roots we may admit $s_0 < s$, thereby establishing cointegration of rank $q-(s-s_0)$, or admit all of them, thus establishing that the series is $I(0)$, or admit none of them in which case we shall conclude that we have cointegration of rank $q-s$. From the point of view of the **maximum of the LF** there is absolutely **no distinction** among these alternatives. If we claim that we **know** that **cointegration is present,** then we are entitled to take the rank of B to be $q-1$, since this would make $\Pi(I)$ of rank $q-1$. Equally validly, we could take its rank to be one! Unless we know what the rank is **we cannot settle on anything but maximal rank,** which implies degeneracy. It is important to bear in mind that whether **one or more of the relevant roots employed is zero** has **nothing to do** with the question of the rank of the matrix B we have chosen! Thus, testing to see whether some characteristic roots are, or are not equal to zero is very much like dueling with windmills–it does not affect the outcome of anything of any consequence.

The next question of significance is: why are the results from J so much different from those in EG and Dhrymes (1994b). To answer this question let us juxtapose the three models, obtaining

$$\begin{aligned}
D \,:\, (I-L)X_{t\cdot}' &= A(L)\epsilon_{t\cdot}', \quad A(L)=\sum_{j=0}^{k}A_jL^j, \quad A_0=I_q, \\
EG \,:\, (I-L)X_{t\cdot}' &= C(L)\epsilon_{t\cdot}', \quad C(L)=\sum_{j=0}^{\infty}C_jL^j, \quad \sum_{j=0}^{\infty}\| C_j \|<\infty, \\
C_0 &= I_q \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (33) \\
J \,:\, \Pi(L)X_{t\cdot}' &= \epsilon_{t\cdot}', \qquad \Pi_0 = I_q.
\end{aligned}$$

The first two models are basically similar, with the exception that the ostensible greater generality of the EG formulation leads to difficulties with the concept and existence of cointegration in such models. On the other hand, any application of models that derive from the EG formulation will have to rely on a finite number of parameters only, which is the point of the D formulation to begin with. Thus, we may view D as a refinement and correction of the EG formulation in so far as the comparison with the J formulation is concerned. In the first two formulations we may write, without loss of generality

$$A(L) = A(I) + (I - L)A^*(L), \quad C(L) = C(I) + (I - L)C^*(L), \quad (34)$$

which are obtained by the long division of the operators $A(L)$ and $C(L)$, respectively by $-(L-I)$, coupled in the case of $C(L)$ with a limiting argument. In Eq. (34) $A(I), C(I)$ are the **remainders** and $A^*(L), C^*(L)$ the **quotients**. In the EG case further conditions must be satisfied by $C(L)$ is order for $C^*(L)$ to be well defined. A similar decomposition may be made of $\Pi(L)$, thus obtaining

$$\Pi(L) = \Pi(I) + (I - L)\Pi^*(L). \quad (35)$$

The first thing to observe in all three formulations is the role played by the **remainder terms**. **If** $A(I) = 0$, or $C(I) = 0$, the D and EG formulation suggest that the X-process is $I(0)$. **If** $\Pi(I) = 0$, on the other hand, the J formulation suggests that the X-process is $I(1)$, **provided** $|\Pi^*(z)| = 0$, does not have a root on the unit circle, i.e. if z_j is **any root** then $|z_j| \neq 1$, more precisely $|z_j| < 1$.[4] We shall see in a moment what is the consequence of having $|\Pi^*(z_0)| = 0$, for $|z_0| = 1$. **If** $A(I), C(I)$ are **nonsingular**, we **cannot have cointegration** and, in the case of the D formulation at least, $A(L)$ is **an invertible operator.**[5] In either case, the nonsingularity of the remainder matrix, allows the X-process to remain $I(1)$ wihout being cointegrated. **If** $A(I), C(I) \neq 0$, but are (both) **singular**, we have cointegration and the rank of cointegration is simply the dimension of the null space of these matrices, respectively for the D and EG formulations. Similarly, in the case of the J formulation if $\Pi(I)$ is **nonsingular**, $\Pi(L)$ **is invertible**, for the same reasons given above, and we conclude that the X-process is $I(0)$. Finally if $\Pi(I) \neq 0$, but it is **singular**, we have cointegration and the rank of cointegration is the **dimension of the null space of** $\Pi(I)$.

Notice now certain important distinctions:

[4] It is customary in such discussions to rule out a priori the possibility that roots may lie **outside** the unit circle; thus, the case $|z_j| > 1$, the so called explosive case, is **ruled out** on a priori grounds.

[5] Since the **matrix** involved in $A(I)$ is $A(1)$, the condition that $A(I)$ is **nonsingular rules out a real** unit root; it does not necessarily rule out (pairs of) complex roots with unit modulus. This eventuality is generally ignored in practice, and we follow the prevailing custom.

i. $A(I) = 0$ implies the X-sequence is $I(0)$ in the D formulation; $\Pi(I) = 0$ implies the X-sequence is $I(1)$ in the J formulation. Notice that, in an inference context, it is almost impossible to arrive at these conclusions with real world data in the J formulation; in the D formulation this is possible if and only if the process is, indeed, $I(0)$ and the sample is sufficiently large;

ii. $|A(I)| = 0$ implies cointegration in the D context, and similarly $|\Pi(I)| = 0$ implies cointegration in the J formulation and, in either, the rank of cointegration is the dimension of the null space of the matrices corresponding to $A(I)$ and $\Pi(I)$, respectively;

iii. $A(I)$ and $\Pi(I)$ nonsingular (and implicitly assuming there are no complex roots of unit modulus) imply in the D formulation that the X-process is $I(1)$ without cointegration, while in the J process it implies that the X-process is $I(0)$.

Notice that the two formulations are sort of polar opposites, and this is to be expected since one relies on a $MMA(k)$ representation of the differenced X-process, while the other relies on an $MAR(p)$ representation of the X-process. The range of alternatives in the D formulation is: $I(0)$, $I(1)$ with cointegration, $I(1)$ without cointegration. For the J formulation it is: $I(0)$ and $I(1)$ **with cointegration**. The alternative $I(1)$ without cointegration is almost impossible to attain since, in terms of the earlier discussion would entail something like $P'_{-1}N\Delta P = 0$, or more precisely a "positive outcome" to the test of the hypothesis that $B = 0$. The reader should note that in the J estimation context, there is no way in which B could be estimated to be zero, or even insignificantly different from zero.

Remark 3. If one wished to invent a **valid test** in the J formulation, one of the possible approaches may be: estimate consistently $\Pi(I)$, perhaps by least squares, determine the distribution of $\hat{\Pi}(I)'\hat{\Pi}(I)$ and test for the nummber of nonzero (positive) roots of this matrix; if the number so determined is r, this is the cointegration rank. If no roots are determined to be positive one would conclude that $\Pi(I) = 0$ and the X-process is $I(1)$, **without cointegration**; if no root is determined to be zero one would conclude that $\Pi(I)$ is **nonsingular**, and thus the X-sequence is $I(0)$. This will restore to the test the full range of outcomes that a proper test for cointegration should have.

Remark 4. What is it that is so deficient in the J formulation that it leads to such undesirable features? The basic reason is that, except in the strange case that one writes $\Pi(L)$ when one means to write $(I - L)\Pi^*(L)$, the formulation does not permit the X-process to be $I(1)$ **without cointegration**. Note that in the D or EG formulation the characteristic equation of the system of difference equations is

$$0 = |(1-z)I_q| = (1-z)^q, \tag{36}$$

so that the unit root in question is of **multiplicity** q. In the J formulation the characteristic equation is

$$|\Pi(z)| = (1-z)|\Pi_{(1)}(z)| = 0, \tag{37}$$

if $\Pi(z)$ has **only one unit root**; if it has a unit root of multiplicity $r < q$, it may be represented as

$$|\Pi(z)| = (1-z)^r|\Pi_{(r)}(z)|, \tag{38}$$

and the reader should note that $|\Pi_{(1)}(z)|$ is of degree at most $qp - 1$, while the determinant in Eq. (38) is a polynomial of degree at most $qp - r$. Thus, the multiplicity of degree r version of the J formulation implies that the long run behavior of the X-process is essentially that of

$$\Pi_{(r)}(L)(I - L)^{r/q}X'_{t\cdot} \;=\; K(L)\epsilon'_{t\cdot},$$

for an appropriate forcing function, $K(L)\epsilon'_{t\cdot}$, which is $MMA(s)$ with $s < \infty$. While the relationship above is only suggestive, and need not imply that cointegration is synonymous with fractional integration, nonetheless it illuminates the question of why and how the J formulation is inappropriate for testing for the presence of, or for the rank of, cointegration.

References

Bellman, R. (1960), *Introduction to Matrix Analysis*, New York: McGraw Hill.

Dhrymes, P. J. (1970), *Econometrics: Statistical Foundations and Applications*, New York: Harper and Row, corrected edition (1974) New York: Springer Verlag.

——————— (1971), *Distributed Lags: Problems of Estimation and Formulation* (1971), San Francisco: Holden-Day; second edition, Amsterdam: North Holland (1981).

——————— (1978), *Mathematics for Econometrics*, New York: Springer Verlag; second edition (1984).

——————— (1994a), "Topics in Time Series", mimeo.

——————— (1994b), "A Characterization of Cointegration", Discussion Paper No. 698, Department of Economics, Columbia University.

Dickey, D. A., D.W. Jensen and D. Thornton (1991), "A Primer on Cointegration with an Application to the Demand for Money", *Federal Reserve Bank of St. Louis Review*, vol. 73, pp. 58-78.

Engle, R. F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, vol. 55, pp. 251-276.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, vol. 12, pp. 231-254.

——————— (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, vol. 59, pp. 1551-1580.

Johansen, S. and K. Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money", *Oxford Bulletin of Economics and Statistics*, pp. 169-210.

Stock, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, vol. 55, pp. 1035-1056.

Stock, J. H. and M.W. Watson (1988), "Testing for Common Trends", *Journal of the American Statistical Association*, vol. 89, pp. 1097-1107.

Stone, R. A. (1947), "On the Interdependence of Blocks of Transactions", *Journal of the Royal Statistical Society*, supplement, vol. 9, pp. 1-32.

1993-94 Discussion Paper Series
Department of Economics
Columbia University
420 W. 118 St., Room 1022
New York, N.Y., 10027
Librarian: Angie Ng

The following papers are published in the 1993-94 Columbia University Discussion Paper series which runs from November 1 to October 31. Domestic orders for discussion papers are available for purchase at $5.00 (U.S.) each and $140.00 (U.S.) for the series. Foreign orders cost $8.00 (U.S.) for individual paper and $185.00 for the series. To order discussion papers, please send your check or money order payable to Department of Economics, Columbia University to the above address. Please be sure to include the series number for the paper when you place an order.

671. Investment in U.S. Education and Training
Jacob Mincer (Nov. 1993)

672. Freer Trade and the Wages of the Unskilled: Is Marx Striking Again?
Jagdish Bhagwati and Vivek Dehejia

673. Employer Size and Labor Turnover
Todd Idson

674. Less Crime May Be Worse
Brendan O'Flaherty

675. Team Production Effects on Earnings
Todd Idson

676. Language, Employment, and Earnings in the United States: Spanish-English Differentials from 1970 to 1990
David Bloom and Gilles Grenier

677. The Impact of Performance Incentives on Providing Job Training to the Poor: The Job Training to the Poor: The Job Training Partnership Act (JTPA)
Michael Cragg

678. The Demands to Reduce Domestic Diversity among Trading Nations
Jagdish Bhagwati

679. Mass Layoffs and Unemployment
Andrew Caplin and John Leahy

680. The Economics of Adjustment
Andrew Caplin and John Leahy

681. Miracle on Sixth Avenue: Information Externalities and Search
Andrew Caplin and John Leahy

682. Arbitrage, Gains from Trade and Scoial Diversity: A Unified Perspective on Resource Allocation
Graciela Chichilnisky

683. Who should abate carbon emissions?
Graciela Chichilnisky, Geoffrey Heal

684. Believing in Multiple Equilibria
Graciela Chichilnisky

685. Limited Arbitrage, Gains from Trade and Arrow's Theorem
Graciela Chichilnisky

686. International Emission Permits: Equity and Efficiency
Graciela Chichilnisky, Geoffrey Heal and David Starrett

687. Do Vehicle Emissions Testing Program Improve Air Quality?
Matthew Kahn

688. Sources of Real Exchange Rate Fluctuations: How Important Are Nominal Shocks?
Richard Clarida and Jordi Gali

689. Modeling Soviet Agriculture for Assessing Command Economy Policies
Padma Desai and Balbir Sihag

690. The Changing Labor Market Position of Canadian Immigrants
David Bloom, Gilles Grenier and Morley Gunderson

691. Herd Behavior, the " Penguin Effect ", and the Suppression of Informational Diffusion: An Analysis of Informational Externalities and Payoff Interdependency
Jay Pil Choi

692. Shock Therapy and Russia: Was It Tried? Why Did It Fail? What Did It Do? What Now?
Padma Desai

693. Changes in the Structure of Family Income Inequality in the United States and Other Industrial Nations during the 1980s
McKinley L. Blackburn and David E. Bloom

694. Necessarily Welfare-enhancing Customs Unions with Industrialization Constraints: a Proof of the Cooper-Massell-Johnson-Bhagwati Conjecture
Pravin Krishna and Jagdish Bhagwati

695. The Effect of Household Characteristics on Household-Specific Inflation Rates: An Application to Trends in Child Poverty and Educational Real Wage Differentials
Todd Idson and Cynthia Miller

696. Threats to the World Trading System: Income Distribution and the Selfish Hegemon
Jagdish Bhagwati

697. Intraindustry Trade: Issues and Theory
Jagdish Bhagwati and Donald R. Davis

698. A Characterization of Cointegration
Phoebus Dhrymes

699. The Production of Human Capital and the Lifecycle of Earnings: Variation on a Theme
Jacob Mincer

700. The Philosophy of Locational Competition
Ronald Findlay

701. Size, Sunk Costs, and Judge Bowker's Objection to Free Trade
John McLaren

702. Identification and Kullback Information in the GLSEM
Phoebus Dhrymes

703. Patent Litigation as an Information Transmission Mechanism
Jay Pil Choi

693. Changes in the Structure of Family Income Inequality in the United States and Other Industrial Nations during the 1980s
McKinley L. Blackburn and David E. Bloom

694. Necessarily Welfare-enhancing Customs Unions with Industrialization Constraints: a Proof of the Cooper-Massell-Johnson-Bhagwati Conjecture
Pravin Krishna and Jagdish Bhagwati

695. The Effect of Household Characteristics on Household-Specific Inflation Rates: An Application to Trends in Child Poverty and Educational Real Wage Differentials
Todd Idson and Cynthia Miller

696. Threats to the World Trading System: Income Distribution and the Selfish Hegemon
Jagdish Bhagwati

697. Intraindustry Trade: Issues and Theory
Jagdish Bhagwati and Donald R. Davis

698. A Characterization of Cointegration
Phoebus Dhrymes

699. The Production of Human Capital and the Lifecycle of Earnings: Variation on a Theme
Jacob Mincer

700. The Philosophy of Locational Competition
Ronald Findlay

701. Size, Sunk Costs, and Judge Bowker's Objection to Free Trade
John McLaren

702. Identification and Kullback Information in the GLSEM
Phoebus Dhrymes

703. Patent Litigation as an Information Transmission Mechanism
Jay Pil Choi

704. On the Meaning of Certain Cointegration Tests
Phoebus J. Dhrymes