

ASYMPTOTIC PROPERTIES OF LEAST SQUARES ESTIMATORS OF COINTEGRATING VECTORS

BY JAMES H. STOCK

Time series variables that stochastically trend together form a cointegrated system. In such systems, certain linear combinations of contemporaneous values of these variables have a lower order of integration than does each variable considered individually. These linear combinations are given by cointegrating vectors. OLS and NLS estimators of the parameters of a cointegrating vector are shown to converge in probability to their true values at the rate $T^{1-\delta}$ for any positive δ . These estimators can be written asymptotically in terms of relatively simple nonnormal random matrices which do not depend on the parameters of the system. These asymptotic representations form the basis for simple and fast Monte Carlo calculations of the limiting distributions of these estimators. Asymptotic distributions thus computed are tabulated for several cointegrated processes.

KEYWORDS: Cointegration, error correction models, unit roots.

1. INTRODUCTION

IT IS WIDELY OBSERVED that many time series of economic interest follow a nondeterministic trend in levels or in logarithms, but that these variables appear to be stationary after first differencing. At the same time, however, these variables often trend together: certain linear combinations of contemporaneous observations seem to be stationary in the sense that they do not require further differencing to exhibit limited dependence. This article examines the asymptotic properties of least squares estimators of the parameters of these linear combinations. Using the theory of cointegrated process, I show that these estimators have asymptotic properties different from those of least squares estimators in stationary time series models: convergence to their probability limits occurs faster than usual, and their limiting distribution is nonnormal. If the view of variables stochastically trending together (in the formal sense described below) is correct, then these results suggest that a reexamination of statistical inferences based on such common macroeconomic regressions as aggregate consumption on aggregate income is in order.

Casual observation suggests that the phenomenon of economic time series variables stochastically trending together might be widespread. For example, real GNP and Federal tax receipts have both grown sharply in the postwar U.S. economy. However, the ratio of tax receipts to GNP has fluctuated but has remained in the neighborhood of 19 per cent over the past three decades. As a second, somewhat more formal example, it can be argued that the permanent income hypothesis suggests that in the long run there is a unit income elasticity of consumption. If this is so, then one would expect deviations of consumption from the "long run" level—the level implied by the unit elasticity condition—to be a process with a short memory, even though real consumption expenditures

¹ The author thanks Christopher Cavanagh, Robert Engle, Clive Granger, Andrew Harvey, Thomas Rothenberg, Mark Watson, and an anonymous referee for helpful suggestions concerning this research, and Suzanne Sowinski for her careful preparation of the manuscript. The research was supported in part by the National Science Foundation, Grant No. SES-84-08797.

and disposable income have grown considerably over the postwar period. Accordingly, the difference between log consumption and log income might reasonably be modeled as being stationary, although these variables, each taken individually, cannot be.

Recently there has been increased interest in error correction models, a class of models in which it is explicitly assumed that two or more time series variables stochastically trend together. In these models, deviations from a long run "equilibrium condition" (e.g., that the long run income elasticity of consumption is one) feed back into short run dynamics so that the long run relation tends to be maintained. This "error correction" mechanism ensures that the variables possess a common stochastic trend. Imposing the unit elasticity condition, Davidson, Hendry, Srba, and Yeo (1978) estimate an error correction model of aggregate consumption in the postwar United Kingdom.²

Granger (1983), Granger and Weiss (1983), and Engle and Granger (1987) have recently provided a unified framework for the analysis of error correction models and of time series processes in which the variables stochastically trend together. In this framework, a process is called "cointegrated" (CI) if there are one or more linear combinations of contemporaneous time series variables which have a lower order of integration than do any of the individual random variables which comprise the system. Adopting Engle and Granger's (1987) terminology, a vector that reduces the order of integration of the system is referred to as a cointegrating vector. From the perspective of error correction models, these cointegrating vectors describe the long run "equilibrium conditions" to which the variables tend to return.

The cointegrated process/error correction model is summarized in Section 2. Cointegrated processes have strikingly different properties from stationary time series, and the asymptotic theory presented here reflects those differences. With stationary time series, the expectation of a squared error objective function attains its minimum at the true value of an identifiable parameter (at least asymptotically); the objective function has a finite but larger expectation for other values of the parameter. However, this is not true for some squared error objective functions with CI processes: while the linear combinations of contemporaneous values of the data formed using cointegrating vectors are stationary and have a finite second moment, all other linear combinations are nonstationary and have an infinite unconditional second moment. This suggests first that least squares should produce relatively precise estimators of CI vectors, and second that standard asymptotic results will not apply to these estimators.

The results of this paper confirm this intuition. In Section 3, the properties of the ordinary least squares estimator of the identifiable elements of the CI vector obtained from a contemporaneous levels regression are examined. The main result is that, if each element of the vector X_t is integrated of order one, and if

² See Hendry and Richard (1983) for a further discussion of error correction models. One theoretical justification for considering error correction models is reviewed by Salmon (1982), who describes their derivation from linear quadratic economic optimization problems with normally distributed errors.

there is a cointegrating vector such that the linear combination formed using this vector is integrated of order zero, then the OLS estimator of this vector is consistent (subject to an identification condition) and converges in probability faster than $T^{1-\delta}$ for any positive δ . This contrasts sharply with conventional asymptotic results in which the rate of convergence is $T^{1/2}$.

A nonlinear least squares estimator of the CI vector is examined in Section 4. This estimator is obtained by applying nonlinear least squares to a regression of first differences of an element of X_t on lagged first differences of itself and of the other variables in the system, and on lagged levels of X_t , combined using the cointegrating vector. This NLS estimator corresponds to an unconstrained version of Davidson, Hendry, Srba, and Yeo's (1978) estimator.³ In this section, it is shown that the NLS estimator is consistent and converges at the same rate as the OLS estimator.

The behavior of least squares estimators of the parameters describing the short run dynamics of the CI process are discussed in Section 5. Using the results of the previous two sections, I argue that these estimators converge to limiting normal random variables at the usual rate, $T^{1/2}$. Indeed, because of the fast rate of convergence of the estimators of the CI vector, the short run parameter estimators are asymptotically independent of the estimators of the CI vector and their distribution is well approximated by the standard output of OLS packages.

In Section 6, a technique to simulate the distribution of these estimators is described and implemented for several CI systems. This technique is based on asymptotic representations of the OLS and NLS estimators in terms of two matrix-valued random variables which are multivariate generalizations of the numerator and denominator of the standardized OLS estimator of the first autoregressive coefficient in a scalar AR(1) process with a unit root. As White (1958) showed, the limit distribution of this estimator is nonnormal when there is a unit root. Similarly, the limiting distribution of the standardized (by T) least squares estimators of the CI vector will also be nonnormal. Despite this complication, the asymptotic representations greatly simplify the task of approximating the distribution of the estimators using Monte Carlo techniques.

2. COINTEGRATED PROCESSES AND ESTIMATORS OF COINTEGRATING VECTORS

The Model

Engle and Granger (1987) define a N dimensional time series variable X_t to be cointegrated of orders d and b ($CI(d, b)$) if it satisfies two conditions: (i) each component of X_t , when considered individually, is integrated of order zero after differencing d times; and (ii) there exists at least one (and possibly r) "cointegrating vectors" α_i such that the linear combination $\alpha_i' X_t$ is integrated of order zero after differencing $d - b$ times. Note that, for the vector α_i in (ii) to be consistent with criterion (i), if $b > 0$ then the cointegrating vector must have

³ These authors constrain the long run income elasticity of consumption to be one. If bequests are a luxury good, then the long run income elasticity of consumption would be less than one. Whether this is so is an empirical question.

at least two nonzero elements. In this paper, attention is restricted to the special case discussed in the introduction, $d = b = 1$.

Two different representations of cointegrated processes will be useful in the development below. The first involves the Wold moving average representation of the (mean zero) first differences of X_t :

$$(2.1) \quad \Delta X_t = C(L)\varepsilon_t$$

with $\Delta \equiv 1 - L$, where L is the lag operator and $C(L)$ is a matrix of lag polynomials. I shall assume that the errors ε_t have a nonsingular covariance matrix G and finite fourth moments:

$$(2.2) \quad \varepsilon_t \text{ are i.i.d. } (0, G), \quad E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| \leq \mu < \infty \quad (i, j, k, l = 1, \dots, N).$$

Engle and Granger (1987) show that, if X_t is cointegrated, then $C(1)$ is singular. Furthermore, the cointegrating vectors are those linear combinations of the rows of $C(1)$ that yield the zero vector: $\alpha'_i C(1) = 0, i = 1, \dots, r$. Since $C(1)$ is singular, the spectral density matrix of ΔX_t is singular at zero frequency. This singularity is the restatement in the frequency domains of the time domain notion that, as a result of the error correction mechanism, the components of the CI process stochastically trend together.

The second representation of cointegrated processes is in terms of an error correction model (ECM). In a single equation ECM with a single "equilibrium error", the change in one variable (e.g. $\Delta X_{1,t}$) is written as a linear function of lagged changes of X_t plus a lagged weighted difference of levels of X_t . This lagged difference in levels is the "error correction" term of the ECM. From a time series perspective, if $\Delta X_{1,t}$ is not integrated, this error correction term clearly cannot be integrated since it enters into the expression for $\Delta X_{1,t}$. Thus this equilibrium error must be that linear combination of the integrated variables comprising X_t that is itself *not* integrated, i.e. $\alpha'X_t$. Extending this reasoning to multiple equations, with $r = 1$ the ECM representation for ΔX_t is

$$(2.3) \quad A^*(L)\Delta X_t = -\gamma\alpha'X_{t-1} + d(L)\varepsilon_t$$

where $A^*(0) = I_N$, the $N \times N$ identity matrix, γ is a $N \times 1$ vector, and $d(L)$ is a lag polynomial. The intuition leading to (2.3) is formalized in Theorem 1 of Engle and Granger (1987). They prove that if X_t is cointegrated with $r = 1$ then it has the ECM representation (2.3), where γ is such that $C(1)\gamma = 0$ and $d(L)$ is a scalar lag polynomial with no unit roots. When $1 < r < N$, their theorem indicates that (2.3) still obtains with α and γ respectively being the $N \times r$ matrices with columns satisfying $\alpha'_i C(1) = 0$ and $C(1)\gamma_i = 0, i = 0, \dots, r$.

The asymptotic analysis in the following sections proceeds conditional on $\varepsilon_s = 0$ for $s \leq 0$. This assumption facilitates a simple derivation of a useful representation of X_t and the equilibrium errors. Substituting backwards in (2.1) and rearranging,

$$(2.4) \quad X_t = C(1) \sum_{s=1}^t \varepsilon_s + C^*(L)\varepsilon_t,$$

where

$$C^*(L) = \sum_{j=0}^{\infty} \left[\sum_{i=0}^j C_i - C(1) \right] L^j = \sum_{j=0}^{\infty} \left[- \sum_{i=j+1}^{\infty} C_i \right] L^j.$$

Since $\alpha'_i C(1) = 0$, (2.3) and (2.4) imply that

$$(2.5) \quad z_{it} \equiv \alpha'_i X_t = \alpha'_i C^*(L) \varepsilon_t.$$

The Normalized Cointegrating Vector

The cointegrating vectors are identifiable only up to a scale parameter: if $\alpha'_i X_t$ is stationary, then so too is $c\alpha'_i X_t$ for any constant c . I therefore adopt a normalization that extends the natural specification when $N = 2$. For example, if consumption C_t and income Y_t are cointegrated so that $C_t - \theta Y_t$ is stationary, α can be written as $\alpha' = (1 \ 0) + (0 \ -1)\theta$. More generally, let the first element of α_i be 1, and assume that the remaining elements are linear combinations of k unknown identifiable parameters $\theta_{i1}, \dots, \theta_{ik}$, $k \leq N - 1$. Let $e_{(i)}$ denote the N dimensional unit vector with a one in the same place as it appears in α_i and zeros elsewhere. With this normalization, the cointegrating vector can be written

$$(2.6) \quad \alpha_i = e_{(i)} + R_i \theta_i \quad (i = 1, \dots, r)$$

where $\theta'_i = (\theta_{i1} \ \theta_{i2} \ \dots \ \theta_{ik})$ and R_i is a $N \times k$ matrix of known constants.

The Estimators

In the consumption-income example, the ordinary least squares estimator of the identifiable element of the cointegrating vector is simply the slope coefficient estimated by regressing consumption on income. More generally, the OLS estimator $\hat{\theta}$ solves

$$(2.7) \quad \min_{\theta} \sum_{t=1}^T (\alpha'_i X_t)^2, \quad \alpha_i = e_{(i)} + R_i \theta$$

when the normalization (2.6) is adopted. If, for example, $\alpha_i = (1 \ -\theta_{i1} \ \dots \ -\theta_{iN-1})$, so that there were no restrictions on the elements of α_i other than the first, then (2.7) would simply result in the estimator obtained by regressing X_{1t} on the other elements of X_t .

I also study a nonlinear least squares estimator of θ . There are two reasons to examine this estimator. First, it generalizes the estimator used by Davidson, Hendry, Srba, and Yeo (1978).⁴ Second, the OLS estimator uses only contemporaneous values of X_t , thereby foregoing potential gains in efficiency from

⁴ Davidson, Hendry, Srba, and Yeo (1978) estimate the "consumption function,"

$$\Delta_4 c_t = \beta_1(L) \Delta_4 i_t + \beta_2(C_{t-4} - \theta_{i-4}) + \beta_3(L)' X_t + u_t$$

where c_t , i_t , X_t , and u_t are respectively log personal consumption, log disposable income, other variables (price deflators and a dummy), and an error term, and where $\beta_1(L)$ and $\beta_3(L)$ are lag polynomials. They constrain $\theta = 1$; the NLS estimator estimates θ simultaneously with the other parameters of the equation.

simultaneously estimating the parameters of the short run dynamics and the long run “equilibrium condition.”⁵ The NLS estimator is based on a regression of X_t in first difference form. Let

$$\xi_t = (\Delta X'_t \Delta X'_{t-1} \cdots \Delta X'_{t-p+1})',$$

where p is fixed and finite. Also let $\gamma_j, j = 1, \dots, N$, denote the elements of γ in (2.3). The NLS estimators $\hat{\theta}$, $\hat{\gamma}_1$, and $\hat{\beta}$ solve:

$$(2.8) \quad \min_{\theta, \gamma_1, \beta} \sum_{t=p+1}^T (\Delta X_{1t} - \beta' \xi_{t-1} + \gamma_1 \alpha' X_{t-1})^2$$

where β is the vector of parameters corresponding to the first row of $[I - A^*(L)]$ in (2.3).

If there is one cointegrating vector and if θ is exactly identified, then the NLS estimator can be computed directly from the OLS estimators of the coefficients on the lagged levels terms in the unconstrained ECM representation of the equation at hand. For example, if $N = 2$, then the first equation in the unconstrained ECM representation is

$$(2.9) \quad \Delta X_{1t} = \beta' \xi_{t-1} + \delta_1 X_{1t-1} + \delta_2 X_{2t-1} + u_t$$

where δ_1 and δ_2 are scalar parameters. Under the normalization (2.6), $\alpha = (1 - \theta)'$, $\theta = -\delta_2 / \delta_1$, and $\hat{\theta} = -\hat{\delta}_2 / \hat{\delta}_1$, where $\hat{\delta}_1$ and $\hat{\delta}_2$ are the OLS estimators from (2.9). More generally, the NLS estimators can be computed directly from the unconstrained OLS estimators if the dimensionality of θ is $N - 1$.

3. THE OLS ESTIMATOR

Using the normalization (2.6) and the first order conditions arising from (2.7), the OLS estimator can be written,

$$(3.1) \quad \hat{\theta}_i - \theta_i = -V_T^{-1} U_T$$

where

$$(3.2) \quad \begin{aligned} V_T &= T^{-2} \sum_{t=1}^T R'_t X_t X'_t R_t, \\ U_T &= T^{-2} \sum_{t=1}^T R'_t X_t X'_t \alpha_t, \end{aligned}$$

assuming that the inverse of V_T exists. In this section an asymptotic representation of V_T and TU_T is given in terms of Γ_T and Ψ_T , where

$$\begin{aligned} \Gamma_T &= T^{-2} \sum_{t=1}^T Y_t Y'_t, \\ \Psi_T &= T^{-1} \sum_{t=2}^T Y_{t-1} \eta'_t, \end{aligned}$$

⁵ This NLS estimator is generally not the full information estimator, however. The full information estimator would require imposing cross-equation constraints on the CI process and estimating the system subject to these constraints. Engle and Granger (1987) discuss these constraints.

where $Y_t = \sum_{s=1}^t \eta_s$, $\eta_s = G^{-1/2} \varepsilon_s$, and $G^{1/2} G^{1/2'} = G$.⁶ This representation, given in Theorem 1, makes it possible asymptotically to express the OLS estimator as a function of parameters of the CI process and these two random matrices, the distribution of which is functionally independent of the CI process parameters. In Theorem 2, this representation is used to show that, for all $\delta > 0$, $T^{1-\delta}(\hat{\theta}_t - \theta_t)$ converges to zero in probability.

THEOREM 1: *If $\sum_{u=0}^{\infty} |C_u^*| < \infty$, then $V_T - D_1' \Gamma_T D_1 \xrightarrow{p} 0$ and $TU_T - D_1' \Psi_T D_2 - M \xrightarrow{p} 0$, where $D_1' = R_1' C(1) G^{1/2}$, $D_2' = \alpha_1' C^*(1) G^{1/2}$, and $M = D_1' D_2' + R_1' \sum_{j=0}^{\infty} C_j^* G C_j^{*'} \alpha_1$.*

Proofs of the theorems are given in the Appendix.

The definition of the cointegrating vector requires that z_t be integrated of order zero. Theorem 1 places a stronger requirement on the dependence exhibited by these equilibrium errors: that $C^*(L)$ be summable. This condition, however, is not very restrictive. For example, if $C(L)$ has a finite order q , then by (2.4) $C^*(L)$ has order $q - 1$, so the summability condition automatically holds. More generally, if ΔX_t is a vector ARMA process of finite order, then $C^*(L)$ is absolutely summable.⁷ The conditions that $C(L)$ and $C^*(L)$ be absolutely summable are not equivalent, however; for example, if $C_j = (j + 1)^{-2} I$, then $C(L)$ is absolutely summable but $C^*(L)$ is not.

Theorem 1 expresses TU_T as a sum of mean zero random variables plus a vector of constants. Thus the asymptotic distribution of TU_T (if it exists) will in general have a nonzero mean. Given (3.1), this in turn suggests that the OLS estimator will have a bias of order $O(T^{-1})$. The bias in TU_T may initially seem surprising, for standardized consistent least squares estimators in conventional time series models generally have a limit distribution with mean zero. However, this nonzero limiting mean has a simple interpretation. Since X_t and z_t both consist in part of the same shocks ε_t , $\text{cov}(X_t, z_t)$ will be of order $O(1)$. Were X_t stationary, implying a standardizing factor $T^{1/2}$ rather than T , the estimator would therefore be inconsistent. However, X_t follows a stochastic trend, so this bias enters as $O(T^{-1})$. In other words, the right hand variable $R_1' X_t$ is correlated with the errors in the regression resulting from (2.7). Since $R_1' X_t$ is nonstationary, this

⁶ The random variables Γ_T and Ψ_T are the multivariate generalizations of random variables appearing in the representation of the standardized OLS estimator of the first autoregressive parameter when there is a unit root. Let $y_t = \rho y_{t-1} + \eta_t$, η_t i.i.d. $(0, \sigma^2)$. When $\rho = 1$, then the OLS estimator $\hat{\rho}$ of ρ can be written $T(\hat{\rho} - 1) = [T^{-2} \sum_{t=1}^T y_t^2]^{-1} [T^{-1} \sum_{t=2}^T y_{t-1} \eta_t]$ where y_t is a random walk. The numerator of this expression is the scalar version of Ψ_T ; the denominator is the scalar version of Γ_T , minus the asymptotically negligible term y_T^2/T^2 .

⁷ The one-sided filter $C(L)$ is m -summable if $\sum_{j=0}^{\infty} j^m |C_j| < \infty$. If $C(L)$ is 1-summable, then $C^*(L)$ is summable:

$$\sum_{j=0}^{\infty} |C_j^*| = \sum_{j=0}^{\infty} \left| - \sum_{i=j+1}^{\infty} C_i \right| \leq \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} |C_i| = \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} |C_i| = \sum_{i=1}^{\infty} i |C_i| < \infty.$$

If $C(L)$ is the moving average representation polynomial matrix implied by a finite ARMA model with stable roots, then $C(L)$ is m -summable for all finite m , from which it follows that $C^*(L)$ is summable.

correlation results in a nonzero limiting mean rather than in estimator inconsistency.

I now turn to the consistency of the OLS estimator.

THEOREM 2: *If the column space of R_i does not contain a cointegrating vector, then $T^{1-\delta}(\hat{\theta}_i - \theta_i) \xrightarrow{p} 0$ for all $\delta > 0$.*

The asymptotic representation of Theorem 1 provides a basis for examining the limiting distribution of $T(\hat{\theta}_i - \theta_i)$. Since Y_i is a vector random walk with i.i.d. increments, Donsker's Theorem can be used to write the limit of Γ_T and Ψ_T as functionals of Wiener processes. Let $\{B(t), 0 \leq t \leq 1\}$ denote the N dimensional Wiener process, and let \Rightarrow denote weak convergence of the associated distributions (Billingsley (1968)). Then

$$(3.3) \quad \Gamma_T \Rightarrow \int_0^1 B(t)B(t)' dt \equiv \Gamma,$$

$$\Psi_T \Rightarrow \int_0^1 B(t) dB(t)' \equiv \Psi.$$

Invoking Theorem 1, $V_T \rightarrow D_1' \Gamma D_1$ and $TU_T \rightarrow M + D_1' \Psi D_2$. These results imply that the limiting distribution of the OLS estimator is independent of the distribution of ε_i .⁸

The consistency result of Theorem 2 suggests a reinterpretation of the large sample properties often associated with the slope coefficient in a regression of aggregate consumption against aggregate disposable income. Using a simple Keynesian model, Haavelmo (1943) argued that this regression will yield an inconsistent estimator of the long run marginal propensity to consume because of simultaneous equations bias. In contrast, according to Theorem 2, this estimator is consistent. Indeed, because the normalization of the cointegrating vector is inconsequential for Theorem 2, the inverse of the OLS estimator of the slope coefficient in the reverse regression of income on consumption will also be consistent for the long run marginal propensity to consume. This in turn implies that, as the sample size increases, the correlation between income and consumption (or any two cointegrated variables) will approach 1. Of course, in finite samples these two estimates will differ. The size of this difference will depend in large part on the term M , which arises from the correlation between the right hand variables and the error term in the OLS regression. Thus, if consumption

⁸ In the univariate case, i.e., in the problem of the OLS estimator of a unit root, the possibility of using the invariance principle was noted by White (1958). Rao (1978) uses Donsker's Theorem to argue that his representation of the asymptotic distribution of the OLS estimator in the unit roots problem with normal errors is valid even if the errors are nonnormal but have finite variance. Solo (1984) uses a similar approach, expressing the LM statistic for testing for a unit root in terms of these functionals of scalar Brownian motion. Donsker's Theorem and the continuous mapping theorem for univariate processes are presented and proved in Billingsley (1968) and Hall and Heyde (1980). Subsequent to the submission of this paper, these results have been extended to multivariate processes by Phillips and Durlauf (1986) and Chan and Wei (1986).

and income are cointegrated, Haavelmo's simultaneous equations bias vanishes asymptotically but is present in the form of finite sample bias.

4. THE NLS ESTIMATOR

The normal equation for the NLS estimator $\tilde{\theta}_i$ solving (2.8) implies that

$$(4.1) \quad \tilde{\theta}_i - \theta_i = -\tilde{V}_T^{-1} \tilde{W}_T$$

where

$$\tilde{V}_T = T^{-2} \sum_{t=p+1}^T R'_i X_{t-1} X'_{t-1} R_i,$$

$$\tilde{W}_T = T^{-2} \sum_{t=p+1}^T R'_i X_{t-1} (\Delta X_{1t} - \hat{\beta}' \xi_{t-1} + \tilde{\gamma}_1 z_{it-1}) / \tilde{\gamma}_1,$$

assuming that $\tilde{\gamma}_1 \neq 0$. This estimator is motivated by considering a single equation in the ECM formulation of CI processes (2.3). The plan of this section is the same as Section 3. An asymptotic representation for \tilde{V}_T and \tilde{W}_T is given in Theorem 3 in the case that (2.8) is correctly specified, i.e., $d(L) = 1$. Consistency at the rate $T^{1-\delta}$ is proven in Theorem 4 even if $d(L) \neq 1$.

THEOREM 3: *If $d(L) = 1$, $\gamma_1 \neq 0$, and $\sum_{u=0}^{\infty} |C_u^*| < \infty$, then $T\tilde{V}_T - D'_1 \Gamma_T D_1 \xrightarrow{p} 0$ and $T\tilde{W}_T - D'_1 \Psi_i D_3 \xrightarrow{p} 0$, where $D'_3 = e'_1 G^{1/2} \gamma_1^{-1}$ and $e'_1 = (1 \ 0 \ \dots \ 0)$.*

If $d(L) \neq 1$, then the error term $d(L)\varepsilon_t$ will be correlated with the right hand variables. Even so, as long as $d(L)\varepsilon_t$ exhibits limited dependence, the NLS estimator will be consistent.

THEOREM 4: *Assume that $\sum_{u=0}^{\infty} |C_u^*| < \infty$ and that the column space of R_i does not contain a cointegrating vector. If either:*

- (a) $d(L) = 1$ and $\gamma_1 \neq 0$; or
- (b) $\sum |d_i| < \infty$ and $\tilde{\gamma}_1 \xrightarrow{p} \gamma_1^* \neq 0$;

then $T^{1-\delta}(\tilde{\theta}_i - \theta_i) \xrightarrow{p} 0$ for all $\delta > 0$.

Two remarks on these results are in order. First, using the argument sketched in Section 3, under the conditions of Theorem 3, $T\tilde{W}_T \rightarrow D'_1 \Psi D_3$. Second, Theorem 3 indicates that if $d(L) = 1$, the limiting representation for $T\tilde{W}_T$ does not include a constant term like the matrix M appearing in the limiting representation of TU_T in Theorem 1. The reason is simple: in the case of the OLS estimator, $R'_i X_t$ is correlated with the error term in the regression, while in the NLS case with $d(L) = 1$ the right hand variables are uncorrelated with the error, ε_{1t} . This suggests that the NLS estimator might tend to exhibit less bias than the OLS estimator. It might, however, also exhibit a greater spread in its distribution: if γ_1 is small, then since $D_3 = G^{1/2} e_1 / \gamma_1$, the variability of $\tilde{\theta}_i$ will tend to be large.

5. PROPERTIES OF LEAST SQUARES ESTIMATORS OF OTHER ECM COEFFICIENTS

So far I have been concerned with estimating the parameters describing the long run dynamics of the CI process, i.e., the parameters of the cointegrating vector. The task still remains to estimate the parameters describing the short run dynamics of the system which, in the ECM representation with $d(L) = 1$, are the coefficients on z_{t-1} and on lagged first differences of X_t . Two sets of estimators of the parameters describing the short run dynamics of the CI process are implicit in the previous sections. The first is a two-step estimator: in the first step, α is estimated by OLS as in Section 3, and in the second step z_t is replaced by $\hat{z}_t = \hat{\alpha}'X_t$ and the relevant parameters of the ECM representation are estimated equation by equation. The second procedure is the one step NLS procedure of Section 4.

I now examine the distributions of these two estimators of the parameters describing the short run dynamics. Compared to the estimators of θ , the asymptotics of these estimators are simple: both the NLS and the two step OLS estimators have a limiting normal distribution, converging at rate $T^{1/2}$. In addition, the covariance matrix of the limit distribution is estimated consistently by conventional least squares computer packages, and the two step OLS and NLS estimators are asymptotically independent of the respective estimators of the cointegrating vectors. It follows that the two estimators of the parameters describing the short run dynamics are asymptotically equivalent, even though the OLS and NLS estimators of the cointegrating vectors are not.

The second stage of the two step procedure consists of estimating (by OLS) equations of the form,

$$(5.1) \quad \Delta X_{1t} = \beta' \xi_{t-1} - \gamma_1 \hat{z}_{t-1} + \varepsilon_{1t}$$

where it is assumed that there is a single cointegrating vector and that $d(L) = 1$. Letting $\hat{f}_t = (\xi_t' \quad -\hat{z}_t')$, $f_t = (\xi_t' \quad -z_t')$, and $\phi = (\beta' \quad \gamma_1)'$, the estimator $\hat{\phi}$ of ϕ obtained by estimating (5.1) using OLS is:

$$\hat{\phi} = \hat{F}_T^{-1} T^{-1} \sum_{t=p+1}^T \hat{f}_{t-1} \Delta X_{1t}$$

so that

$$(5.2) \quad T^{1/2}(\hat{\phi} - \phi) = \hat{F}_T^{-1} T^{-1/2} \sum_{p+1}^T \hat{f}_{t-1} (f_{t-1} - \hat{f}_{t-1})' \phi + \hat{F}_T^{-1} T^{-1/2} \sum_{p+1}^T \hat{f}_{t-1} \varepsilon_{1t}$$

where $\hat{F}_T = T^{-1} \sum_{p+1}^T \hat{f}_{t-1} \hat{f}_{t-1}'$. Also let $\tilde{f}_t = (\xi_t' - \tilde{z}_t)'$ and $\tilde{F}_T = T^{-1} \sum_{p+1}^T \tilde{f}_{t-1} \tilde{f}_{t-1}'$, where $\tilde{z}_t = \tilde{\alpha}'X_t$ and $\tilde{\alpha}$ is the NLS estimator of α . Using the normal equations from (2.8), the NLS estimator $\tilde{\phi}$ of ϕ can similarly be written:

$$(5.3) \quad T^{1/2}(\tilde{\phi} - \phi) = \tilde{F}_T^{-1} T^{-1/2} \sum_{p+1}^T \tilde{f}_{t-1} (f_{t-1} - \tilde{f}_{t-1})' \phi + \tilde{F}_T^{-1} T^{-1/2} \sum_{p+1}^T \tilde{f}_{t-1} \varepsilon_{1t}$$

Finally, let $\hat{\phi}^*$ denote the OLS estimator of ϕ based on f_t , i.e., the OLS estimator that would be used were α known. Then

$$(5.4) \quad T^{1/2}(\hat{\phi}^* - \phi) = F_T^{-1} T^{-1/2} \sum_{p+1}^T f_{t-1} \varepsilon_{1t}$$

where

$$F_T = T^{-1} \sum_{t=p+1}^T f_{t-1} f'_{t-1}.$$

Turning first to the two step OLS estimator, the scaled difference between $\hat{\phi}$ and $\hat{\phi}^*$ is

$$\begin{aligned} (5.5) \quad T^{1/2}(\hat{\phi} - \hat{\phi}^*) &= \hat{F}_T^{-1} T^{-1/2} \sum_{p+1}^T \hat{f}_{t-1} (f_{t-1} - \hat{f}_{t-1})' \varepsilon_{1t} \\ &\quad + \hat{F}_T^{-1} T^{-1/2} \sum_{p+1}^T (\hat{f}_{t-1} - f_{t-1}) \varepsilon_{1t} \\ &\quad + (\hat{F}_T^{-1} F_T - I) F_T^{-1} T^{-1/2} \sum_{p+1}^T f_{t-1} \varepsilon_{1t}. \end{aligned}$$

Unlike the moment matrices involving X_t , the sums in (5.5) involve either the stationary variables f_t and ΔX_{1t} , or estimates of f_t based on $\hat{\alpha}$. Since $F_T \xrightarrow{p} F = E f_t f'_t$ (a nonsingular covariance matrix) and since $T^{-1/2} \sum_{p+1}^T f_{t-1} \varepsilon_{1t}$ has a limiting distribution, sufficient conditions for $T^{1/2}(\hat{\phi} - \hat{\phi}^*) \xrightarrow{p} 0$ are that:

$$(5.6a) \quad T^{-1/2} \sum_{p+1}^T (\hat{f}_{t-1} - f_{t-1}) \varepsilon_{1t} \xrightarrow{p} 0,$$

$$(5.6b) \quad \hat{F}_T \xrightarrow{p} F,$$

$$(5.6c) \quad T^{-1/2} \sum_{p+1}^T \hat{f}_{t-1} (\hat{f}_{t-1} - f_{t-1})' \xrightarrow{p} 0.$$

(Note that (5.6b) implies that $\hat{F}_T^{-1} F_T \xrightarrow{p} I$.) Since $\hat{f}_t - f_t = (0' z_t - \hat{z}_t)'$, sufficient conditions for (5.6) to hold are that:

$$(5.7a) \quad T^{-1/2} \sum (z_{t-1} - \hat{z}_{t-1}) \varepsilon_{1t} \xrightarrow{p} 0,$$

$$(5.7b) \quad T^{-1/2} \sum (z_{t-1} - \hat{z}_{t-1}) f_{t-1} \xrightarrow{p} 0,$$

$$(5.7c) \quad T^{-1/2} \sum (\hat{z}_t^2 - z_t^2) \xrightarrow{p} 0.$$

Conditions (5.7) obtain by combining the argument in the proof of Theorem 1 that $T^{-1} \sum X_t \Delta X'_{t-j} = O_p(1)$ for fixed j with the result of Theorem 2 that $T^{1-\delta}(\hat{\theta} - \theta) \xrightarrow{p} 0$. Thus $T^{1/2}(\hat{\phi} - \hat{\phi}^*) \xrightarrow{p} 0$. Note that this argument extends directly to the case that (5.1) contains an intercept as well.

The key observations in arguing that $T^{1/2}(\hat{\phi} - \hat{\phi}^*) \xrightarrow{p} 0$ are that certain random matrices not involving $\hat{\alpha}$ are of order $O_p(1)$ and that $T^{1-\delta}(\hat{\theta} - \theta) \xrightarrow{p} 0$. Because of the similarity between the formulas for $\hat{\phi}$ in (5.2) and for $\tilde{\phi}$ in (5.3), the preceding argument can be repeated exactly with \tilde{f}_t , \tilde{F}_T , and $\tilde{\alpha}$ replacing \hat{f}_t , \hat{F}_T , and $\hat{\alpha}$. Thus $T^{1/2}(\tilde{\phi} - \hat{\phi}^*) \xrightarrow{p} 0$ if $T^{1-\delta}(\tilde{\theta} - \theta) \xrightarrow{p} 0$, which in turn follows from Theorem 4. In addition, it follows that $T^{1/2}(\tilde{\phi} - \hat{\phi}) \xrightarrow{p} 0$, so the two estimators of the short run dynamics are asymptotically equivalent.

These results are formal restatements of the economic distinction made between the short run dynamics embodied in ϕ and the long run relation embodied in α . The long run relation can be estimated in such a way that the convergence of these estimators is especially fast. However, the estimated parameters of the short run relation converge at the usual, slower rate. Moreover, because of the fast rate of convergence of the estimator of the CI vector, the distribution of the short run parameter estimator $\hat{\phi}$ is asymptotically well approximated by the distribution of $\hat{\phi}^*$, for which α is assumed known. If $d(L) = 1$, then $\hat{\phi}^*$ (and thus $\hat{\phi}$ and $\tilde{\phi}$) will be consistent. In addition, $\hat{\phi}$ and $\tilde{\phi}$ will be approximately normal, with a covariance matrix $G_{11}F^{-1}$ (where G_{11} denotes the (1, 1) element of G). This covariance matrix is consistently estimated by the usual OLS output from the second stage OLS regression.

6. ASYMPTOTIC DISTRIBUTIONS

Despite the convenient asymptotic representation of the OLS and NLS estimators in terms of the functionals in (3.3), the exact asymptotic distribution of the OLS estimators would be difficult to compute in the course of applied research (cf. Rao (1978), Evans and Savin (1981a)). In this section I propose an alternative approach to the distributional problem which may prove more practical than exact calculation. This approach is based on the representations of Theorems 1 and 3.

In the asymptotic representations, the parameters of the cointegrated process enter as additive and multiplicative factors. Thus it is not feasible to tabulate the distribution of the estimators for all potentially interesting cointegrated processes. In addition, direct simulation of the distribution of the estimators of the cointegrating system by bootstrap or Monte Carlo methods (using the estimated parameters as the basis for generating pseudo-data) would be excessively costly for common use. However, the representations suggest an alternative two step approach to obtaining numerical approximations to the distributions. In the first step, Monte Carlo draws of the random matrices Γ_T and Ψ_T are generated (with sufficiently large T) and stored. In the second step, pairs of (Γ_T, Ψ_T) can be drawn randomly from the many such pairs already created; using the asymptotic representation of $\hat{\theta}$ and $\tilde{\theta}$ and using known or estimated values of D_1, D_2, D_3 , and M , the distribution of the OLS and NLS estimators can then be computed numerically. This technique has the advantage that, once the pairs (Γ_T, Ψ_T) have been created, it is inexpensive to evaluate the distributions of the OLS and NLS estimators for different values of the nuisance parameters.⁹

⁹ An alternative approach to the distributional problem would be to obtain an integral expression for the distribution of the estimator using the limit representation in terms of the limit functionals of Brownian motion. This approach would entail generalizing the results of White (1958), Rao (1978, 1980), and Evans and Savin (1981a, 1981b), who variously derive and evaluate expressions for the limit distribution of the OLS estimator of the first autoregressive parameter when there is a unit root. Our proposal is in keeping with the Monte Carlo approach to the unit root problem taken by, for example, Dickey and Fuller (1979), Dickey, Hasza, and Fuller (1984), and Evans and Savin (1984).

I present two sets of results based on this numerical approach. First, the convergence of the distribution of the OLS estimator to the distribution of its asymptotic representation is investigated for a specific cointegrated process. Second, the asymptotic distributions of the OLS and NLS estimators are tabulated and compared for several cointegrated processes. These asymptotic distributions are studied using a three parameter family of CI processes in their ECM representation. This family is given by (2.3) with $N = 2$, $r = 1$, and

$$(6.1) \quad \begin{aligned} A^*(L) &= (1 - \rho L)I, & G &= I, & d(L) &= 1, \\ \alpha' &= (1 - \theta), & \gamma' &= (\gamma_1 \gamma_2). \end{aligned}$$

The true value of the parameter to be estimated, θ , is taken to be 1. This family of CI processes is thus parameterized by γ_1 , γ_2 , and ρ .¹⁰

Monte Carlo distributions of the OLS estimator and its asymptotic approximation (computed using the true values of D_1 , D_2 , and M) are presented in Table I.¹¹ For the CI process investigated in this table, both the asymptotic distribution and the empirical Monte Carlo distribution of the standardized OLS estimator have a nonzero mean. The resultant bias in small samples can be considerable: for example, for $T = 25$, the expectation of the OLS estimator is .833; the expectation of its asymptotic approximation is .827. Because of standardization at the rate T , this bias rapidly becomes small: for $T = 200$, the OLS estimator has expectation .977. In addition to being biased, the distribution of the OLS estimator is strongly skewed. For $T = 50$, the mean of the standardized OLS estimator is -4.27 , while its median is -3.39 . This skewness persists as T increases.

The results of Table I suggest that the asymptotic approximation performs well for small samples. Although the accuracy of the approximation improves slowly as the sample size increases, there are still discrepancies between the approximate and the true distribution at $T = 200$. Still, the convergence in this case appears rapid enough to make the asymptotic approximations useful.

The asymptotic distribution of the standardized OLS estimator ($T = 200$) is tabulated for various parameter values in Table II. These distributions vary greatly, depending upon the values of the three nuisance parameters of the CI process. For example, the medians of the standardized distributions range from $-.29$ to -13.96 ; the standard deviations range from 1.64 to 21.18. In all cases, however, the median of the distribution is less than zero, and in all cases but one the OLS estimator is biased towards zero.

The asymptotic distribution of the standardized NLS estimator, based on the representation of Theorem 3, is presented in Table III for $T = 200$. When $\gamma_2 = 0$, the distribution of the NLS estimator is considerably less skewed than that of

¹⁰ If $d(L) = 1$ in (2.3), then a simple recursion can be used to compute $C(L)$ in terms of the parameters of the ECM representation. Using (2.1), (2.3), (2.4), and $\alpha' C(1) = 0$, if $A^*(L) = I - \sum_{i=1}^p A_i^* L^i$, then $C_0 = I$ and, for $j \geq 1$, $C_j = \sum_{i=1}^j (A_i^* - \gamma \alpha') C_{j-i}$, where A_i^* is set to zero for $i > p$.

¹¹ Multivariate normal errors were used to generate the cointegrated time series and the random variates Γ_T and Ψ_T . The simulated CI time series and the matrix variates were constructed from different noise vectors.

TABLE I
 EMPIRICAL DISTRIBUTION OF STANDARDIZED OLS ESTIMATOR $T(\hat{\theta} - \theta)$ AND ITS ASYMPTOTIC APPROXIMATION
 MODEL (6.1), $\rho = .25$, $\gamma_1 = .5$, $\gamma_2 = 0$, and $\theta = 1$.^a

Estimator	Probability of a Smaller Value (p)										Moments	
	.05	.10	.25	.50	.75	.90	.95	Mean	Std. Dev.			
$T=25$	$F_{OLS}^{-1}(p)$	-13.05	-10.20	-6.13	-3.11	-1.24	0.40	1.57	-4.17	4.77		
	$F_{asy}^{-1}(p)$	-14.42	-10.61	-6.00	-3.04	-1.15	0.30	1.32	-4.32	5.22		
	$F_{OLS}(F_{asy}^{-1}(p))$.040	.088	.260	.509	.762	.891	.943	—	—		
$T=50$	$F_{OLS}^{-1}(p)$	-13.00	-10.15	-6.07	-3.39	-1.31	0.37	1.40	-4.27	4.86		
	$F_{asy}^{-1}(p)$	-14.17	-10.20	-5.84	-3.26	-1.27	0.19	1.31	-4.28	4.91		
	$F_{OLS}(F_{asy}^{-1}(p))$.039	.098	.226	.513	.755	.891	.946	—	—		
$T=100$	$F_{OLS}^{-1}(p)$	-13.61	-10.05	-6.10	-3.28	-1.37	0.27	1.25	-4.31	4.79		
	$F_{asy}^{-1}(p)$	-14.10	-10.69	-6.29	-3.32	-1.38	0.32	1.22	-4.55	5.32		
	$F_{OLS}(F_{asy}^{-1}(p))$.046	.086	.239	.494	.749	.902	.948	—	—		
$T=200$	$F_{OLS}^{-1}(p)$	-13.70	-10.26	-6.60	-3.50	-1.44	0.13	1.10	-4.56	5.06		
	$F_{asy}^{-1}(p)$	-14.31	-10.93	-6.52	-3.46	-1.36	0.07	1.01	-4.66	5.24		
	$F_{OLS}(F_{asy}^{-1}(p))$.041	.093	.253	.505	.757	.896	.948	—	—		

^a Based on 2000 replications each of $\hat{\theta}$, F_T , and Ψ_T . $T = 25, 50, 100, 200$. The distributions labeled "OLS" refer to Monte Carlo draws of $\hat{\theta}$; the asymptotic distributions were calculated using the two step numerical approach described in the text.

TABLE II
ASYMPTOTIC DISTRIBUTION OF THE STANDARDIZED OLS ESTIMATOR $T(\hat{\theta} - \theta)$ FOR MODEL (6.1), $T = 200$, AND $\theta = 1$.^a

Model Parameters			Probability of a Smaller Value							Moments	
γ_1	γ_2	ρ	.05	.10	.25	.5	.75	.9	.95	Mean	Std. Dev.
.5	0	0	-24.24	-18.07	-10.72	-5.64	-2.60	-0.54	0.33	-7.93	8.47
		.25	-14.31	-10.93	-6.52	-3.46	-1.36	0.07	1.10	-4.66	5.24
		.5	-6.18	-4.77	-2.86	-1.37	-0.18	1.04	1.85	-1.69	2.64
		.75	-2.63	-1.91	-1.13	-0.35	0.52	1.55	2.47	-0.25	1.64
.2	0	0	-59.96	-44.70	-26.52	-13.96	-6.44	-1.34	0.81	-19.60	20.94
		.25	-41.35	-30.95	-18.43	-9.83	-4.32	-0.64	1.57	-13.44	14.61
		.5	-21.69	-16.81	-10.07	-5.23	-1.86	0.52	2.12	-6.92	8.10
		.75	-3.73	-2.63	-1.50	-0.29	1.16	3.06	4.88	0.03	2.70
.5	-0.5	0	-35.38	-27.79	-16.01	-7.09	-2.79	-0.78	-0.03	-11.22	12.40
		.25	-22.16	-17.55	-9.82	-4.36	-1.67	-0.24	0.34	-7.02	7.96
		.5	-13.65	-10.50	-6.01	-2.61	-0.93	-0.05	0.35	-4.21	4.87
		.75	-5.97	-4.79	-2.75	-1.23	-0.51	-0.16	-0.05	-1.95	2.13
.2	-0.2	0	-59.92	-44.27	-25.88	-10.78	-3.41	0.89	2.95	-17.59	21.18
		.25	-42.76	-31.98	-18.51	-7.67	-2.25	0.99	2.60	-12.57	15.34
		.5	-27.86	-20.62	-12.12	-4.98	-1.39	0.70	1.89	-8.12	10.00
		.75	-12.50	-9.83	-5.50	-2.43	-0.91	-0.10	0.27	-3.91	4.48

^a Based on 2000 replications of (I_T, Ψ_T) , $T=200$.

TABLE III
ASYMPTOTIC DISTRIBUTION OF THE STANDARDIZED NLS ESTIMATOR $T(\tilde{\theta} - \theta)$ MODEL (6.1), $T = 200$, $\theta = 1$.^a

Model Parameters			Probability of a Smaller Value							Moments	
γ_1	γ_2	ρ	.05	.10	.25	.5	.75	.9	.95	Mean	Std. Dev.
.5	0	0	-6.61	-4.56	-2.22	0.01	2.36	4.98	7.40	0.16	4.53
		.25	-4.96	-3.42	-1.66	0.01	1.77	3.73	5.55	0.12	3.40
		.5	-3.31	-2.28	-1.11	0.00	1.18	2.49	3.70	0.08	2.27
		.75	-1.65	-1.14	-0.55	0.00	0.59	1.25	1.85	0.04	1.13
.2	0	0	-16.53	-11.40	-5.54	0.02	5.90	12.45	18.51	0.40	11.33
		.25	-12.40	-8.55	-4.15	0.01	4.43	9.34	13.88	0.30	8.49
		.5	-8.27	-5.70	-2.77	0.01	2.95	6.22	9.25	0.20	5.66
		.75	-4.13	-2.85	-1.38	0.00	1.48	3.11	4.63	0.10	2.83
.5	-0.5	0	-4.83	-3.31	-1.27	1.80	6.76	12.86	17.49	3.59	7.34
		.25	-3.62	-2.48	-0.95	1.35	5.07	9.65	13.12	2.70	5.50
		.5	-2.41	-1.65	-0.63	0.90	3.38	6.43	8.74	1.80	3.67
		.75	-1.21	-0.83	-0.32	0.45	1.68	3.22	4.33	0.90	1.83
.2	-0.2	0	-12.07	-8.27	-3.16	4.51	16.89	32.16	43.72	8.99	18.34
		.25	-9.05	-6.20	-2.37	3.38	12.67	24.12	32.79	6.74	13.75
		.5	-6.03	-4.14	-1.58	2.25	8.45	16.08	21.87	4.49	9.17
		.75	-3.05	-2.08	-0.80	1.11	4.22	8.04	10.81	2.23	4.58

^a Based on 2000 replications of (I_T, Ψ_T) , $T = 200$.

the OLS estimator. In addition, in this case the NLS estimator is much closer to being unbiased than is the OLS estimator. However, for nonzero γ_2 and small γ_1 , the standardized NLS estimator is not asymptotically unbiased; indeed, for small ρ , this bias can be substantial, although it typically is not as large in absolute value as is the bias of the OLS estimator. Finally, in most cases the asymptotic distribution of the NLS estimator has a smaller spread (as measured either by the interquartile range or by the standard deviation) than does the OLS estimator.

This numerical approach to computing the distributions of $\hat{\theta}$ and $\tilde{\theta}$ can be used as a basis for statistical inference. For example, to test a hypothesis on an element of θ , say $\theta_1 = \theta_1^0$ vs. $\theta_1 \neq \theta_1^0$, the other elements of θ and the nuisance parameters $C(L)$ and G (and thus D_1 , D_2 , D_3 , and M) could be estimated consistently under the restriction on θ either by two step OLS or by NLS. The asymptotic distribution of $\hat{\theta}_1$ under the null hypothesis could then be computed using the proposed techniques, thereby obtaining the probability under the null of realizing a value of (say) $|\hat{\theta}_1 - \theta_1^0|$ at least as large as that estimated. The construction of confidence intervals for θ_1 would be more difficult, since it would entail computing the critical points of $\hat{\theta}_1 - \theta_1^0$ for many values of θ_1^0 . An instructive approach would be to adopt Kendall and Stuart's (1967, p. 120) suggestion and to "invert" the test statistic graphically.

In summary, these results suggest three conclusions. First, the asymptotic representation theorems provide an accurate and fast means of obtaining numerical approximations to the otherwise complicated distributions of the NLS and OLS estimators. Second, these distributions depend strongly on the other parameters of the CI system. Thus the parameters of the entire CI process must be estimated before inferences can be made about the estimator of the CI vector. Third, the OLS and NLS estimators can be substantially biased, especially for small sample sizes. For the parameterizations considered here, the OLS estimator appears more prone to this bias than does the NLS estimator, although this need not be so in general. Thus researchers estimating the parameters of cointegrated processes should exercise caution when interpreting their estimates without first adjusting for bias.

7. SUMMARY

Two broad conclusions emerge from this analysis. First, if a vector of time series variables is cointegrated, then least squares estimators of the parameters of the cointegrating vector will have a nonnormal limiting distribution which resembles a multivariate generalization of the distribution of the estimated autoregressive coefficient when there is a unit root in a univariate series. This distribution will in general be skewed, and both the OLS and the NLS estimators can have limit distributions with nonzero means after standardization. Furthermore, these estimators converge to their limiting distributions at a fast rate.

Second, these results indicate that inference based on the standard least squares output can be misleading in time series regressions with both lagged differences and levels of the dependent variable appearing as explanatory variables. In this

case, the moment matrix of the levels regressors converges to a limiting random variable, and the distribution of certain regression coefficients will not be well approximated by normality.

These results suggest a simple computational technique for evaluating these distributions. This technique is based on representing the estimators as a function of consistently estimable parameters and two matrix-valued random variables. Once these random matrices have been generated, the distribution of the OLS and NLS estimators is readily computed by Monte Carlo integration based on draws of these random matrices. The computational savings of this approach over a bootstrap or full Monte Carlo procedure is substantial, reducing the number of computations by a factor of T .

On an informal level, there is good reason to suspect that cointegration could be a widespread phenomenon in macroeconomic data. For example, Nelson and Plosser (1982) have presented evidence that many macroeconomic variables seem to follow a stochastic rather than a deterministic trend and that, after differencing, these variables often appear to be stationary. If in addition certain linear combinations of contemporaneous values of these variables are stationary without further differencing, then the system of these variables form a cointegrated process, and the usual distributional approximations will not apply. As a specific example, Engle and Granger (1987) present evidence supporting the application by Davidson et al. (1978) of an ECM model to consumption and income. Their finding suggests that researchers who have regressed aggregate consumption on aggregate income have erred when using the standard errors produced by conventional regression packages. Still, this regression of consumption on income is not spurious: under the assumption of cointegration, the OLS and NLS estimators of the cointegrating vector are consistent and, indeed, converge to their probability limits at a relatively fast rate.

The John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138, U.S.A.

Manuscript received May, 1985; final revision received November, 1986.

APPENDIX

PROOFS OF THEOREMS

PROOF OF THEOREM 1: Let $C^+(L) = C^*(L)G^{1/2}$. From (2.4),

$$X_T = C(1) \sum_{i=1}^T \varepsilon_i + C^*(L)\varepsilon_T = C(1)G^{1/2}Y_T + C^+(L)\eta_T.$$

Thus:

$$\begin{aligned} V_T &= T^{-2} \sum_{i=1}^T R_i' [C(1)G^{1/2}Y_i + C^+(L)\eta_i] [C(1)G^{1/2}Y_i + C^+(L)\eta_i]' R_i \\ &= R_i' C(1)G^{1/2} \left[T^{-2} \sum_{i=1}^T Y_i Y_i' \right] G^{1/2} C(1)' R_i \end{aligned}$$

$$\begin{aligned}
 &+R'_i C(1)G^{1/2} \left[T^{-2} \sum_{i=1}^T Y_i(C^+(L)\eta_i)' \right] R_i \\
 &+R'_i T^{-2} \sum_{i=1}^T (C^+(L)\eta_i) Y_i' G^{1/2} C(1)' R_i \\
 &+R'_i T^{-2} \sum_{i=1}^T (C^+(L)\eta_i)(C^+(L)\eta_i)' R_i
 \end{aligned}$$

so

$$(A.1) \quad V_T - D'_1 \Gamma_T D_2 = T^{-1} (D'_1 H_1 R_i + R'_i H'_1 D_1 + R'_i H_2 R_i)$$

where

$$\begin{aligned}
 H_1 &= T^{-1} \sum_{i=1}^T Y_i(C^+(L)\eta_i)', \\
 H_2 &= T^{-1} \sum_{i=1}^T (C^+(L)\eta_i)(C^+(L)\eta_i)'.
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 TU_T &= T^{-1} \sum_{i=1}^T R'_i [C(1)G^{1/2} Y_i + C^+(L)\eta_i] [C(1)G^{1/2} Y_i + C^+(L)\eta_i]' \alpha_i \\
 &= D'_1 H_1 \alpha_i + R'_i H_2 \alpha_i
 \end{aligned}$$

since $\alpha'_i C(1) = 0$. Proceed by rewriting H_1 , recalling that $\varepsilon_s = 0, s \leq 0$:

$$\begin{aligned}
 (A.2) \quad H_1 &= T^{-1} \sum_{i=1}^T Y_i \left[\sum_{j=0}^{i-1} C_j^+ \eta_{i-j} \right]' \\
 &= T^{-1} \sum_{i=1}^T \sum_{j=0}^{i-1} (Y_i - Y_{i-j}) \eta'_{i-j} C_j^{+'} \\
 &\quad + T^{-1} \sum_{i=1}^T \sum_{j=0}^{T-1} Y_{i-j} \eta'_{i-j} C_j^{+'} \\
 &= H_3 - H_4 + (\Psi_T + H_5) \sum_{j=0}^{T-1} C_j^{+'}
 \end{aligned}$$

where

$$\begin{aligned}
 H_3 &= T^{-1} \sum_{j=0}^{T-1} \zeta_j C_j^{+'}, \quad \zeta_j = \sum_{i=j+1}^T (Y_i - Y_{i-j}) \eta'_{i-j}, \\
 H_4 &= T^{-1} \sum_{j=0}^{T-1} \phi_j C_j^{+'}, \quad \phi_j = \sum_{i=T-j+1}^T Y_i \eta'_i, \\
 H_5 &= T^{-1} \sum_{i=1}^T \eta_i \eta'_i.
 \end{aligned}$$

Thus

$$\begin{aligned}
 (A.3) \quad TU_T &= D'_1 \Psi_T D_2 + D'_1 H_5 D_2 + R'_i H_2 \alpha_i \\
 &\quad + D'_1 (H_3 - H_4) \alpha_i + D'_1 (\Psi_T + H_5) \left[\left(\sum_{j=0}^{T-1} C_j^+ \right) - C^+(1) \right]' \alpha_i
 \end{aligned}$$

since $C^+(1)' \alpha_i = G^{1/2} C^*(1)' \alpha_i = D_2$. The expressions (A.1)-(A.3) are exact, involving no asymptotic approximations. From (A.3), the representation for TU_T stated in the theorem follows if it is shown that: (i) $H_2 \xrightarrow{P} \sum_{j=0}^{\infty} C_j^* G C_j^{*'}; (ii) H_3 \xrightarrow{P} 0; (iii) H_4 \xrightarrow{P} 0; (iv) H_5 \xrightarrow{P} I$. If in addition it is shown that (v) $T^{-\delta} \Psi_T \xrightarrow{P} 0$ for all $\delta > 0$, then, using (A.1), the representation for V_T also obtains.

Since η_i has finite fourth moments and identity covariance matrix, (iv) is immediate. Since $C(L)$ is summable, (i) is also immediate. I therefore turn to (ii), (iii), and (v), thereby completing the proof.

(ii) This will be shown using Chebyshev's inequality. Since $E\eta_s\eta'_t=0, s \neq t, E\xi_j=0$. To show that $H_3 \xrightarrow{p} 0$, consider a typical element of H_3 . The (l, k) element of H_3 is

$$(A.4) \quad (H_3)_{lk} = \sum_{m=1}^N \delta_{lmk} \quad \text{where}$$

$$\delta_{lmk} \equiv T^{-1} \sum_{j=0}^{T-1} (\xi_j)_{lm} (C_j^+)_{km},$$

where $(\xi_j)_{lm}$ is the (l, m) element of ξ_j . It will be shown that, under the stated conditions, each element of the sum (A.4) converges to zero in probability, and thus that $H_3 \xrightarrow{p} 0$. Since $E\xi_j=0$, it is sufficient to show that $\text{var}(\delta_{lmk}) \rightarrow 0$. Now

$$(A.5) \quad \text{var}(\delta_{lmk}) = T^{-2} \sum_{j=0}^{T-1} \sum_{i=0}^{T-1} E(\xi_i)_{lm} (\xi_j)_{lm} (C_j^+)_{km} (C_i^+)_{km}$$

$$\leq T^{-2} \sum_{j=0}^{T-1} \sum_{i=0}^{T-1} |(C_j^+)_{km}| |(C_i^+)_{km}| \times \max_{i,j,l,m} (E(\xi_j)_{lm}^2, E(\xi_i)_{lm}^2).$$

Also

$$E(\xi_i)_{lm}^2 = \sum_{t=j+1}^T \sum_{s=j+1}^T E(Y_{1t} - Y_{1t-j})(Y_{1s} - Y_{1s-j}) \eta_{m,t-j} \eta_{m,s-j}$$

$$= j(T-j)$$

since $E\eta_{mt}\eta_{ls}=1$ if $m=1$ and $t=s$ and $=0$ otherwise. Thus

$$\max_{i,j,l,m} (E(\xi_i)_{lm}^2, E(\xi_j)_{lm}^2) \leq [\max(j, i)][T - \min(j, i)]$$

so by (A.5),

$$\text{var}(\delta_{lmk}) \leq T^{-2} \sum_{j=0}^{T-1} \sum_{i=0}^{T-1} |(C_i^+)_{km}| |(C_j^+)_{km}| [\max(j, i)][T - \min(j, i)]$$

$$\leq 2 \left[\sum_{j=0}^{T-1} (j/T) |(C_j^+)_{km}| \right] \left[\sum_{i=0}^{T-1} (1-i/T) |(C_i^+)_{km}| \right].$$

Since $C^*(L)$ is summable, $C^+(L)$ is summable and is therefore Cesaro summable (Anderson (1971), Lemma 8.3.1). Thus $\sum_{j=0}^{T-1} (1-j/T) |C_j^+| \rightarrow \sum_{j=0}^{\infty} |C_j^+|$ and $\sum_{j=0}^{T-1} (j/T) |C_j^+| \rightarrow 0$, so $\text{var}(\delta_{lmk}) \rightarrow 0$ and $H_3 \xrightarrow{p} 0$.

(iii) This argument proceeds in the same manner as the argument for (ii). First,

$$EH_4 = T^{-1} \sum_{j=0}^{T-1} E\phi_j C_j^{+'} = T^{-1} \sum_{j=0}^{T-1} (j-1) C_j^{+'} \rightarrow 0$$

by Cesaro convergence. Letting $\tau_{lmk} \equiv T^{-1} \sum_{j=0}^{T-1} (\phi_j)_{lm} (C_j^+)_{km}$, by the argument leading to (A.5),

$$\text{var}(\tau_{lmk}) \leq T^{-2} \sum_{j=0}^{T-1} \sum_{i=0}^{T-1} |(C_j^+)_{km}| |(C_i^+)_{km}| \max[\text{var}((\phi_i)_{lm}), \text{var}((\phi_j)_{lm})].$$

By direct calculation,

$$\text{var}[(\phi_j)_{lm}] = \sum_{t=T-j+1}^T \text{var}(Y_{it}\eta_{mt}) = (j-1)\mu_4 + j(T-1) - j(j-1)/2$$

where $\mu_4 = \max_j E(\eta_{jt}^4)$. Consequently

$$\text{var}(\tau_{lmk}) \leq 2 \left[T^{-2} \sum_{j=0}^{T-1} [(j-1)\mu_4 + j(T-1) - j(j-1)/2] |(C_j^+)_{mk}| \right]$$

$$\times \left[\sum_{j=0}^{T-1} |(C_j^+)_{mk}| \right] \rightarrow 0$$

since

$$\sum_{i=0}^{T-1} (j^2/T^2) |C_j^+| \leq \sum_{j=0}^{T-1} (j/T) |C_j^+| \rightarrow 0.$$

(v) This result also obtains using Chebyshev's inequality. Note that $E\Psi_T = 0$ and that $\text{var}(\Psi_{Tij}) = T^{-2} \sum_{t=2}^T \sum_{s=2}^T E(Y_{it-1} Y_{is-1} \eta_{jt} \eta_{js}) = T(T-1)/2T^2$. Thus $\text{var}(T^{-\delta} \Psi_{Tij}) \rightarrow 0$, so $T^{-\delta} \Psi_T \xrightarrow{p} 0$. Q.E.D.

PROOF OF THEOREM 2: Using (3.1), $T^{1-\delta}(\hat{\theta}_i - \theta_i) = -V_T^{-1} T^{-\delta}(TU_T)$. Consequently, $T^{1-\delta}(\hat{\theta}_i - \theta_i) \xrightarrow{p} 0$ if (i) V_T^{-1} has a limiting distribution; and (ii) $T^{1-\delta} U_T \xrightarrow{p} 0$.

(i) Using the representation of Theorem 1, $V_T^{-1} = [D_1' \Gamma_T D_1]^{-1} + o_p(1)$. Thus V_T^{-1} will have a limiting distribution if Γ_T^{-1} has a limiting distribution and if D_1 has full column rank. By Theorem 2.4 of Chan and Wei (1986) or Lemma 3.1 of Phillips and Durlauf (1986), $\Gamma_T \Rightarrow \Gamma$, where Γ is given in (3.3). By the continuous mapping theorem, Γ_T^{-1} thus has the limiting distribution of $\Gamma^{-1} = [\int_0^1 B(t) B(t)' dt]^{-1}$.

The condition that D_1 have full column rank can be restated as requiring that there not exist a vector b such that $b'R_1' C(1) G^{1/2} = 0$. Since G is nonsingular, this is equivalent to requiring that there be no b^* such that $b^* C(1) = 0$, where $b^* = b'R_1'$. If such a b^* were to exist, it would be by definition a cointegrating vector, say α^* . Thus V_T^{-1} has a limiting distribution if there is no α_j , $j=1, \dots, r$, and vector b such that $R_1 b = \alpha_j$. That is, the column space of R_1 cannot contain a cointegrating vector.

(ii) By Theorem 1 and the result (v) obtained in its proof, $T^{-\delta}(TU_T) \xrightarrow{p} 0$. Q.E.D.

PROOF OF THEOREM 3: Since $\tilde{V}_T = V_T - R_1' T^{-2} \sum_{i=1}^{p-1} X_i X_i' R_1 - R_1' T^{-2} X_T X_T' R_1$, and since $T^{-2} \sum_{i=1}^{p-1} X_i X_i' \xrightarrow{p} 0$ and $T^{-2} X_T X_T' \xrightarrow{p} 0$, it follows from Theorem 1 that $T\tilde{V}_T - D_1' \Gamma_T D_1 \xrightarrow{p} 0$. Turning to \tilde{W}_T ,

$$\begin{aligned} T\tilde{W}_T &= (D_1' \Psi_T D_3 + R_1' G^{1/2} [T^{-1} \sum \eta_t (C^*(L) \eta_{t-1})'] D_3) (\gamma_1 / \tilde{\gamma}_1) \\ &\quad - R_1' (T^{-1} \sum X_{t-1} \xi_{t-1}') (\tilde{\beta} - \beta) / \tilde{\gamma}_1 \\ &\quad + R_1' (T^{-1} \sum X_{t-1} z_{t-1}) (\tilde{\gamma}_1 - \gamma_1) / \tilde{\gamma}_1. \end{aligned}$$

Since β , γ_1 , and θ are identifiable and since the objective function (2.8) asymptotically attains a minimum at the true parameter values, $\tilde{\beta}$, $\tilde{\gamma}_1$, and $\tilde{\theta}$ are consistent. Therefore, the theorem obtains if

- (i) $T^{-1} \sum \eta_t (C^*(L) \eta_{t-1})' \rightarrow 0$,
- (ii) $\text{var}[(T^{-1} \sum X_{t-1} \xi_{t-1}')_{im}] = 0(1)$,
- (iii) $\text{var}[(T^{-1} \sum X_{t-1} z_{t-1})_j] = 0(1)$,

for typical elements (l, m) and j of the corresponding random variables.

Condition (i) can be demonstrated to hold if $C^*(L)$ is summable using the same arguments as in the proof of Theorem 1. Conditions (ii) and (iii) also follow from the proof of Theorem 1 upon simple rewriting. For condition (ii), a typical element of ξ_{t-1} is $\Delta X_{jt-k} = \Delta X_{t-k} e_j$, where e_j is the j th canonical unit vector. Now

$$\begin{aligned} T^{-1} \sum X_{t-1} \Delta X_{t-k} e_j &= C(1) G^{1/2} T^{-1} \sum Y_{t-1} (C(L) G^{1/2} \eta_{t-k})' e_j \\ &\quad + T^{-1} \sum (C^+(L) \eta_{t-1}) (C(L) G^{1/2} \eta_{t-k})' e_j. \end{aligned}$$

From the proof of Theorem 1, the first term has variance of order $0(\text{var}(\Psi_{Tjk})) = 0(1)$, and the second term converges in probability to $\sum C_{j+k-1}^* G C_j^*$. Thus (ii) is satisfied. Condition (iii) follows analogously. Q.E.D.

PROOF OF THEOREM 4: Under both conditions (a) and (b), $\tilde{V}_T - D_1' \Gamma_T D_1 \xrightarrow{p} 0$. It follows from the representation (4.1) and result (i) in the proof of Theorem 2 that $T^{1-\delta}(\hat{\theta}_i - \theta_i) \rightarrow 0$ if $T^{-\delta} \tilde{W}_T \xrightarrow{p} 0$. I now demonstrate that $T^{-\delta} \tilde{W}_T \xrightarrow{p} 0$ under both the conditions.

Allowing for general $d(L)$, $T\tilde{W}_T$ can be written as

$$\begin{aligned} T\tilde{W}_T &= [D_1' \Psi_T D_3 + D_1' T^{-1} \sum Y_{t-1} (d^+(L) \eta_{t-1})' D_3] (\gamma_1 / \tilde{\gamma}_1) \\ &\quad + R_1' T^{-1} \sum [C^+(L) \eta_{t-1}] [d(L) \eta_t]' D_3 (\gamma_1 / \tilde{\gamma}_1) \\ &\quad - R_1' (T^{-1} \sum X_{t-1} \xi_{t-1}') (\tilde{\beta} - \beta) / \tilde{\gamma}_1 \\ &\quad + R_1' (T^{-1} \sum X_{t-1} z_{t-1}) (\tilde{\gamma}_1 - \gamma_1) / \tilde{\gamma}_1 \end{aligned}$$

where $d^+(L) = (d(L) - d(0))L^{-1}$. Thus sufficient conditions for $T^{-1-\delta}\tilde{W}_T \xrightarrow{P} 0$ are that:

- (i) $T^{-1-\delta}\Psi_T \xrightarrow{P} 0$,
- (ii) $T^{-1-\delta} \sum Y_{t-1}[d^+(L)\eta_{t-1}]' \xrightarrow{P} 0$,
- (iii) $T^{-1} \sum [C^+(L)\eta_{t-1}][d(L)\eta_t]' \xrightarrow{P} \sum_{j=1}^{\infty} C_{j+1}^* Gd_j < \infty$,
- (iv) $T^{-1-\delta} \sum X_{t-1}\xi'_{t-1} \xrightarrow{P} 0$ and $(\tilde{\beta} - \beta)/\tilde{\gamma}_1 \xrightarrow{P} c_1$,
- (v) $T^{-1-\delta} \sum X_{t-1}z_{t-1} \xrightarrow{P} 0$ and $(\tilde{\gamma}_1 - \gamma_1)/\tilde{\gamma}_1 \xrightarrow{P} c_2$,

where c_1 and c_2 are finite constants.

Condition (i) was shown as condition (v) in the proof of Theorem 1. Condition (iii) follows from the assumed absolute summability of $C^*(L)$ and $d(L)$. The first parts of conditions (iv) and (v) are implied by conditions (ii) and (iii) in the proof of Theorem 3. The second parts of conditions (iv) and (v) hold under both the conditions (a) and (b), but for slightly different reasons. Under (a), $\tilde{\beta} \xrightarrow{P} \beta$ and $\tilde{\gamma}_1 \xrightarrow{P} \gamma_1 \neq 0$, so $c_1 = c_2 = 0$. Under (b), $\tilde{\beta} \xrightarrow{P} \beta^* \neq \beta$ in general and $\tilde{\gamma}_1 \xrightarrow{P} \gamma_1^* \neq \gamma_1$ in general; however, $\gamma_1^* \neq 0$ by assumption, so c_1 and c_2 are finite.

All that remains is to show that condition (ii) is satisfied. This is clearly so under assumption (a), for then $d^+(L) = 0$. Under (b), note that $T^{-1} \sum Y_{t-1}[d^+(L)\eta_{t-1}]'$ has exactly the same form as H_1 in Theorem 1, with $d^+(L)I$ replacing $C^+(L)$, except for asymptotically negligible differences in the limits of the summations. Since $T^{-\delta}H_1 \xrightarrow{P} 0$ if $\sum |C_j^*| < \infty$, it follows immediately that $T^{-1-\delta} \sum Y_{t-1}[d^+(L)\eta_{t-1}]' \xrightarrow{P} 0$ if $\sum |d_j^+| < \infty$. Since $\sum |d_j^+| < \sum |d_j|$, condition (ii) holds under both assumptions (a) and (b). Q.E.D.

REFERENCES

ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series*. New York: Wiley.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

CHAN, N. H., AND C. Z. WEI (1986): "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," manuscript, University of Maryland.

DAVIDSON, J. E. H., D. F. HENDRY, F. SRBA, AND S. YEO (1978): "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdom," *Economic Journal*, 8, 661-692.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

DICKEY, D. A., D. P. HASZA, AND W. A. FULLER (1984): "Testing for Unit Roots in Seasonal Time Series," *Journal of the American Statistical Association*, 79, 355-367.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-Integration and Error-Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

EVANS, G. B. A., AND N. E. SAVIN (1981a): "The Calculation of the Limiting Distribution of the Least Squares Estimator of the Parameter in a Random Walk Model," *Annals of Statistics*, 9, 1114-1118.

— (1981b): "Testing for Unit Roots: I," *Econometrica*, 49, 753-779.

— (1984): "Testing for Unit Roots: II," *Econometrica*, 52, 1241-1260.

FULLER, W. A. (1976): *Introduction to Statistical Time Series*. New York: Wiley.

GRANGER, C. W. J. (1983): "Cointegrated Variables and Error—Correcting Models," University of California—San Diego Discussion Paper 83-13.

GRANGER, C. W. J., AND A. A. WEISS (1983): "Time Series Analysis of Error—Correcting Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press, 255-278.

HAAVELMO, T. (1943): "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1-12.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and its Application*. New York: Academic Press.

HENDRY, D. F., AND J. F. RICHARD (1983): "The Econometric Analysis of Economic Time Series," *International Statistical Review*, 51, 111-163.

- KENDALL, M., AND A. STUART (1979): *The Advanced Theory of Statistics*, Vol. 2, Fourth edition. New York: MacMillan.
- NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139-162.
- PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-496.
- RAO, M. M. (1978): "Asymptotic Distribution of an Estimator of the Boundary Parameter of an Unstable Process," *Annals of Statistics*, 6, 615-629.
- SALMON, M. (1982): "Error Correction Mechanisms," *Economic Journal*, 92, 615-629.
- SOLO, V. (1984): "The Order of Differencing in ARMA Models," *Journal of the American Statistical Society*, 79, 916-921.
- WHITE, J. S. (1958): "Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematic Statistics*, 29, 1188-1197.