COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER 768

MULTIPLE TIME SERIES REGRESSION WITH INTEGRATED PROCESSES

by

P. C. B. Phillips and S. N. Durlauf

September 1985

MULTIPLE TIME SERIES REGRESSION WITH INTEGRATED PROCESSES*

by

P. C. B. Phillips and S. N. Durlauf

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

This paper develops a general asymptotic theory of regression for processes which are integrated of order one. The theory includes vector autoregressions and multivariate regressions amongst integrated processes that are driven by innovation sequences which allow for a wide class of weak dependence and heterogeneity. The models studied cover cointegrated systems such as those advanced recently in [15] and quite general linear simultaneous equations systems with contemporaneous regressor-error correlation and serially correlated errors. Problems of statistical testing in vector autoregressions and multivariate regressions with integrated processes are also studied. It is shown that the asymptotic theory for conventional tests involves major departures from classical theory and raises new and important issues of the presence of nuisance parameters in the limiting distribution theory.

September, 1985

*Our thanks go to Glena Ames for her skill and effort in typing the manuscript of this paper and to the NSF for research support under Grant No. SES 8218792.

1. INTRODUCTION

Unlike many of the time series encountered in the natural sciences, economic time series frequently exhibit characteristics that are widely believed to be intrinsically nonstationary. For example, real macroeconomic variables such as output and consumption typically display a strong secular or growth component as well as cyclical behavior; and many financial series like common stock prices behave in general as if they had no fixed mean. Recognizing these typical characteristics of economic time series, econometricians have devoted attention to the problem of describing and modeling nonstationarity. In the 1960's important contributions in the area were made by Granger, Hatanaka and their associates in [3], [12], [13], [20]. Later, following the influential work of Box and Jenkins [4], attention shifted to the role of integrated processes in modeling economic time series. While undoubtedly restricting the class of nonstationary models, integrated processes of the ARIMA type have been found to produce highly satisfactory representations of many observed time series in economics. Quite recently, Nelson and Plosser [29] have published a detailed empirical study of historical economic time series for the USA. These authors provide some convincing evidence that macroeconomic time series normally thought to be stationary about a time trend are better described as integrated processes with drift.

Amongst the latest research in this field have been the studies of cointegration by Granger and his associates in [15]. Two time series are said to be cointegrated if some linear combination of the series has a lower order of integration than the individual series. These authors argue that the notion of (steady state) equilibrium in economics implies the existence of such relationships. Thus, a classical economist's view of the interaction of money growth and price movements would require these series to move closely together over time even if the series themselves are integrated and are individually well described by a model such as a random walk. Empirical support is found in [15] for the cointegration of several macroeconomic time series.

Against this background of empirically motivated research there is a growing need for a general theory of statistical inference for time series regression with integrated processes. A step towards the achievement of this goal has been taken by the first author in recent work [33, 34, 35]. The central idea of [33] is to use a theory of weak convergence in function spaces that allows us to work with integrated processes that are driven by quite general weakly dependent and possibly heterogeneously distributed innovations. The theory developed in [33] is univariate; [34] deals with bivariate regressions; and [35] develops a general theory of asymptotic expansions for vector autoregressions with integrated processes.

The present paper seeks to provide a formal analytical framework for research in this field by the development of a theory of multiple time series regression with integrated processes. Our analysis derives from the use of a multivariate functional central limit theorem for sums of dependent random variables. This theorem opens the way to an asymptotic theory for vector autoregressions and more general multivariate regressions that permit a wide variety of error processes. In this sense the paper builds on the methods and the results of [33], [34] and [35].

Our organization of the paper is as follows. The multivariate functional limit theorems that we use are stated and discussed in Section 2. Proofs of these and other theorems of the paper are given in the Mathematical Appendix. Section 3 studies the statistical properties of vector autoregressions with processes that are integrated of order one. Our analysis includes the asymptotic distribution of the usual error covariance matrix estimator and we also examine the asymptotic properties of the conventional F test for unit roots. Section 4 deals with more general multivariate regressions amongst integrated processes and develops an asymptotic theory for such regressions. This theory includes the estimation of cointegrated systems as a special case and should therefore be useful in such applications. Section 5 examines problems of hypothesis testing in multiple regressions with integrated processes. Here the asymptotic theory that we develop displays striking departures from the classical theory and, in particular, raises new and important problems of the presence of nuisance parameter matrices in the limiting distribution theory. Section 6 demonstrates extensions of this theory to the case of fitted drift vectors. Some conclusions are given in Section 7. Proofs and additional technical material are presented in the Mathematical Appendix to the paper.

2. INVARIANCE PRINCIPLE FOR MULTIPLE TIME SERIES

Our starting point will be a multivariate generalization of the functional central limit theorem or invariance principle due to McLeish [26] that is used in Phillips [33].

We consider a sequence of $n \times 1$ random vectors $\{u_t\}$ defined on a probability space $(\Omega, \mathbf{B}, P)$ such that

(1) $$E(u_t) = 0 \quad \forall t .$$

We define the vector partial sums:

(2) $$S_t = \Sigma_1^t u_j$$

and the vector functionals

(3) $$X_T(t) = \frac{1}{\sqrt{T}} \Sigma^{-1/2} S_{[Tt]} = \frac{1}{\sqrt{T}} \Sigma^{-1/2} S_{j-1}$$

$$\frac{j-1}{T} \leq t < \frac{j}{T}$$

and

(4) $$X_T(1) = \frac{1}{\sqrt{T}} \Sigma^{-1/2} S_T$$

where $\Sigma^{-1/2}$ is a positive definite matrix defined below.

Note that

(5) $$X_T(t) \in D[0,1]^n = D[0,1] \times \ldots \times D[0,1]$$

the product metric space of all real valued vector functions on [0,1] that are right continuous at each element of [0,1] and possess finite left limits.

We endow each $D[0,1]$ with the Skorohod metric, denoted by d, whose definition and properties may be found in the Mathematical Appendix. This metric renders $D[0,1]$ a complete, separable metric space. Separability does not occur when the uniform metric is employed. As will be apparent in the Mathematical Appendix, separability of the underlying coordinate spaces is useful when working with product spaces.

For the product topology $D[0,1]^n$ we therefore choose the metric

(6) $$d'(x,y) = \sup_i d(x_i(t), y_i(t)) \ , \quad d = \text{Skorohod metric.}$$

This choice of product metric implies, given separability of $D[0,1]$, that the σ-algebra for the product space is equivalent to the σ-algebra generated by the measurable rectangles defined on the underlying coordinate spaces.

Finally, we define the following measures of temporal dependence for the $\{u_t\}$ sequence. For σ-algebras F and G, define

$$(7) \qquad \varphi(F,G) = \sup_{\{F\in F, G\in G, P(F)>0\}} |P(G|F) - P(G)|$$

$$\alpha(F,G) = \sup_{\{F\in F, G\in G\}} |P(F,G) - P(F)P(G)| \ .$$

We further define F_a^b as the σ-algebra generated by $\{u_a, \ldots, u_b\}$ and R_a^b as the σ-algebra generated by $\{S_b - S_{a-1}, \forall a \leq b\}$. Temporal dependence for the $\{u_t\}$ sequence may thus be defined by the two measures:

$$(8) \qquad \varphi_m = \sup_N \sup_{j\geq N+M} \varphi(F_1^N, R_{N+M}^j)$$

$$\alpha_m = \sup_N \sup_{j\geq N+M} \alpha(F_1^N, R_{N+M}^j) \ .$$

These measures are weaker than uniform or strong mixing requirements placed upon F_a^b alone. φ_m and α_m generalize the univariate measures defined by McLeish [26]. Following [33] we say that ϕ_m (or α_m) is of size $-p$ if ϕ_m (or α_m) $= O(m^{-p-\varepsilon})$ for some $\varepsilon > o$ as $m \uparrow \infty$.

Our multivariate invariance principle may now be stated.

THEOREM 2.1. If $\{u_t\}_1^\infty$ *is a sequence of random* $n \times 1$ *vectors satisfying* (1) *and*

(a) $E(T^{-1}S_TS_T') \to \Sigma$ *a positive definite matrix;*

(b) $\{u_{it}^2\}$ *is uniformly integrable for all* $i = 1, \ldots, n$*;*

(c) $\sup_t (E|u_{it}|^\beta) < \infty$ *for some* $2 \leq \beta < \infty$ *and all* $i = 1, \ldots, n$*;*

(d) $E(T^{-1}(S_{k+T} - S_k)(S_{k+T} - S_k)') \to \Sigma$ *as min* $(k,T) \uparrow \infty$*;*

(e) *either* φ_m *is of size* $-\beta/(2\beta-2)$ *or* $\beta > 2$ *and* α_m *is of size* $-\beta/(\beta-2)$*;*

then $X_T(t) \Rightarrow W(t)$ *as* $t \to \infty$.

The notation $\Rightarrow$ in the above theorem signifies weak convergence of the associated probability measures [2]. $W(t)$ is a multivariate Wiener process. Each element of $W(t)$ is a univariate Wiener process and the elements of $W(t)$ are independent. Finally, $W(t) \in C[0,1]^n$ almost surely, where $C[0,1]$ is the space of continuous functions defined on $[0,1]$. The proof of Theorem 2.1 may be found in the Mathematical Appendix.

The convergence properties of the theorem permit the $\{u_t\}$ process to possess a high degree of temporal dependence and moderate heteroskedasticity. As will be clear in Sections 3 and 4 below, the multivariate invariance principle will permit the derivation of regression asymptotics for a wide class of error processes.

When $\{u_t\}$ is stationary, the requirements of the theorem may be relaxed. We thus have:

THEOREM 2.2. If $\{u_t\}$ *is a weakly stationary sequence of random* $n \times 1$ *vectors satisfying* (1) *and*

(a) $E|u_{i1}|^\beta < \infty$ *for some* $2 \leq \beta < \infty$

(b) *either* $\Sigma_{n=1}^{\infty}\varphi^{1-1/\beta} < \infty$ *or* $\beta > 2$ *and* $\Sigma_{n=1}^{\infty}\alpha_n^{1-2/\beta} < \infty$

then

$$\Sigma = \lim_{T\to\infty} E(T^{-1}S_TS_T') \tag{9}$$

and if Σ *is positive definite, then* $X_T(t) \Rightarrow W(t)$ *as* $T \uparrow \infty$.

This theorem generalized the stationary version of the McLeish invariance principle. The method of proof mirrors the proof of Theorem 2.1 and is therefore not reported.

The conditions of Theorem 2.2 are different from those of Theorem 2.1 in two respects. First, the convergence of $E(T^{-1}S_TS_T')$ is a conclusion and not an assumption of the theorem. The cost of eliminating this assumption is a tightening of the mixing assumption (b). If mixing assumption (e) in Theorem 2.1 were employed, and (9) added as an assumption, then the two theorems would be identical as the uniform integrability assumption (a) of Theorem 2.1 is automatically fulfilled by the assumption of stationarity and the stated moment condition.

3. VECTOR AUTOREGRESSIONS WITH INTEGRATED PROCESSES

Let $\{y_t\}_1^\infty$ be a multiple $(n \times 1)$ time series generated in discrete time according to:

(10a) $\quad y_t = Ay_{t-1} + u_t\ ; \quad t = 1, 2, \ldots$

(10b) $\quad A = I_n\ ,$

where $\{u_t\}_1^\infty$ satisfies the requirements of either Theorem 2.1 or 2.2 in the previous section. To complete the specification of (10a) we add either of the commonly proposed initial conditions (cf. [33]):

(11a) $\quad y_0 = c\ ,$ a constant; or

(11b) $\quad y_0 =$ random with a certain specified distribution.

Note that (10) allows for quite general vector ARMA specifications such that

(12) $\quad (1-L)A(L)y_t = B(L)v_t$

where $A(L)$ and $B(L)$ are finite order matrix polynomials in the lag operator L. We need only require $A(L)$ to be stable and then $u_t = A(L)^{-1}B(L)v_t$ will satisfy the weak dependence and heterogeneity assumptions of Theorem 2.1 under very general conditions on the innovation sequence $\{v_t\}$. For example, if $\{v_t\}$ is an i.i.d. Gaussian sequence, then the autocorrelation sequence of $\{u_t\}$ decays exponentially. Since this latter sequence bounds the strong mixing coefficient numbers [21], the mixing decay rate conditions of Theorem 2.1 are clearly satisfied. For nonGaussian sequences, similar results hold provided the density of v_t is absolutely continuous [5].[1]

The following lemma is useful in the derivation of our main results.

LEMMA 3.1. If $\{u_t\}_1^\infty$ *satisfies the conditions of Theorem* 2.1 *and* $\{y_t\}_1^\infty$ *is generated by* (10) *then:*

(a) $T^{-3/2}\Sigma_1^T y_t \Rightarrow \Sigma^{1/2}\int_0^1 W(t)dt$

(b) $T^{-2}\Sigma_1^T y_t y_t' \Rightarrow \Sigma^{1/2}\int_0^1 W(t)W(t)'dt\Sigma^{1/2}$

(c) $T^{-2}\Sigma_1^T(y_t - \bar{y})(y_t - \bar{y})' \Rightarrow \Sigma^{1/2}\{\int_0^1 W(t)W(t)'dt - \int_0^1 W(t)dt\int_0^1 W(t)'dt\}\Sigma^{1/2}$

(d) $T^{-1}\Sigma_1^T y_{t-1}u_t' \Rightarrow \frac{1}{2}\{\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_u\}$

where

$$\Sigma_u = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(u_t u_t') \tag{13}$$

$$\Sigma = \lim_{T\to\infty} E(T^{-1}S_T S_T') \tag{14}$$

and $W(t)$ *is a vector Wiener process on* $C[0,1]^n$. *Moreover,* (a)-(b) *continue to hold whether the initial conditions are given by* (11a) *or* (11b).

This Lemma generalizes to vector processes the asymptotic theory for sample moments of integrated processes given earlier in [33] and [34].

---

[1] The results in [5] demonstrate that for an error sequence with absolutely continuous densities, stable autoregressive transformations of the sequence will be strong mixing. However, the order of the strong mixing coefficient cannot, in general, be determined.

Its results enable us to develop a simple asymptotic theory for vector autoregressions. Consider, in particular, the matrix of regression coefficients

$$(15) \qquad A^* = Y'Y_{-1}(Y_{-1}'Y_{-1})^{-1} \; ; \quad Y' = [y_1, \ldots, y_T] \; , \; Y_{-1}' = [y_1, \ldots, y_{T-1}]$$

from a first order autoregression of y_t on y_{t-1}. The associated error covariance matrix estimator is:

$$(16) \qquad \Sigma^* = T^{-1}Y'(I - P_{Y_{-1}})Y \; ; \quad P_{Y_{-1}} = Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}' \; .$$

Our next theorems provide the asymptotic distribution theory for these least squares regression estimates.

THEOREM 3.2. If $\{u_t\}_1^\infty$ *satisfies the conditions of either Theorem* 2.1 *or* 2.2 *and if* $\{y_t\}_1^\infty$ *is generated by* (10) *then as* $T \uparrow \infty$:

(a) $T(A^* - I) \Rightarrow \frac{1}{2}\{\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_u\}\{\Sigma^{1/2}\int_0^1 W(t)W(t)'dt\Sigma^{1/2}\}^{-1}$;

(b) $A^* \underset{p}{\to} A$;

(c) $\Sigma^* \underset{p}{\to} \Sigma_u$.

THEOREM 3.3. Let $v_t = (u_{it}^2 - E(u_{it}^2))_{n\times 1}$. *If* $\{v_t\}_1^\infty$ *satisfies the conditions of Theorem* 2.2 *then as* $T \uparrow \infty$ $T^{1/2}\mathrm{vec}(\Sigma^* - \Sigma_u) \Rightarrow N(0,V)$ *where*

$$(17) \qquad V = P_D\Sigma_{k=0}^\infty\{\Psi_k - (\mathrm{vec}\ \Sigma_u)(\mathrm{vec}\ \Sigma_u)'\}P_D \; ;$$

$$(18) \qquad \Psi_k = E(u_t u_{t+k}' \otimes u_t u_{t+k}') \; ; \quad k = 0, 1, 2, \ldots \; ;$$

$$(19) \qquad P_D = D(D'D)^{-1}D' \; ;$$

and D *is the* $n^2 \times n(n+1)/2$ *duplication matrix for which* $\text{vec}\, S = Ds$ *for any symmetric matrix* S *with* s *denoting the vector of its nonredundant elements* [27].

Theorem 3.2 generalizes Theorems 3.1 and 3.2 of [33] to vector processes which are integrated of order one. In contrast to stable vector autoregressions (in which the latent roots of A in (10) have modulus less than unity), Theorem 3.2 shows that simple least squares regression yields consistent estimates even in the presence of substantial serial correlation. Moreover, the asymptotic distribution of the least squares regression coefficient matrix has the same general form for a wide variety of different error processes. These general results certainly permit serious misspecifications in the vector autoregressive (VAR) framework that has recently become so popular in empirical econometric work [8, 24]. The latter work proceeds under the assumption of white noise innovation sequences and makes use of conventional asymptotic theory based on the original theory of Mann and Wald [28]. This assumption of conventional theory proves to be inadequate even if the underlying system is vector ARMA rather than VAR; and the asymptotic theory breaks down due to the resulting misspecification as well as the nonstationarity of the series. By contrast, Theorem 3.2 allows for nonstationarity in the underlying processes, a quite general weak dependence in the errors that certainly admits underlying vector ARMA structures and, finally, heterogeneous error variances that permit moderate heteroskedasticity. Thus, Theorem 3.2 is an important first step in the development of a robust asymptotic theory for nonstationary VAR's.

Theorem 3.3 provides an associated asymptotic theory for the error covariance matrix estimator Σ^*. Unlike the regression coefficient matrix

A^* , the asymptotic distribution of Σ^* is normal with a covariance matrix that depends on the joint cumulant sequence of the 4^{th} order of the innovation process $\{u_t\}$. The limiting normal distribution is explained by the observation that, since $A^* \underset{p}{\to} I$, the residuals from this regression are asymptotically stationary and, in effect, provide consistent estimates of the innovation process. Conventional normal asymptotics are therefore to be expected for the sample moments of such a sequence.

It is of some interest to consider test statistics based on the matrix of regression coefficients A^* . Of primary interest will be the hypothesis

$$H_0 : A = I \ . \tag{20}$$

We shall first examine the distribution of the Wald statistic for testing H_0 :

$$F = \mathrm{tr}[(A^* - I)'\Sigma^{*-1}(A^* - I)Y_{-1}'Y_{-1}] \ . \tag{21}$$

THEOREM 3.4. If $\{y_t\}_1^\infty$ *is generated by* (10) *and if* $\{u_t\}_1^\infty$ *satisfies the conditions of either Theorem* 2.1 *or* 2.2 *then as* $T \to \infty$:

$$F \Rightarrow (1/4)\mathrm{tr}\Big\{[W(1)W(1)' - \Sigma^{-1/2}\Sigma_u\Sigma^{-1/2}]\ {}^{1/2}\Sigma_u^{-1}\Sigma^{1/2}[W(1)(1)' - \Sigma^{-1/2}\Sigma_u\Sigma^{-1/2}] \cdot [\textstyle\int_0^1 W(t)W(t)'dt]^{-1}\Big\} \ . \tag{22}$$

Note that when $n = 1$ (22) reduces to

$$F \Rightarrow \frac{(1/4)(W(1)^2\sigma^2 - \sigma_u^2)^2}{\sigma^2\sigma_u^2\int_0^1 W(t)^2 dt} \tag{23}$$

which is the square of the result established in [33] for the t ratio statistic in the univariate regression.

We also note that when $\Sigma = \Sigma_u$, as in the case of a white noise innovation sequence $\{u_t\}_1^\infty$, (22) becomes

$$(24) \quad F \Rightarrow (1/4)\mathrm{tr}\left\{[W(1)W(1)' - I][W(1)W(1)' - I][\int_0^1 W(t)W(t)'dt]^{-1}\right\}$$

generalizing the tests of Dickey and Fuller [7] and Evans and Savin [10].

When $\Sigma \neq \Sigma_u$ we need to construct consistent estimates of these matrices. Since $T^{-1}\Sigma_1^T u_t u_t' \to \Sigma_u$ a.s. as $T \uparrow \infty$ we have the simple estimator

$$(25) \qquad S_u = T^{-1}\Sigma_1^T(y_t - y_{t-1})(y_t - y_{t-1})'$$

which is consistent for Σ_u under the null hypothesis (20).

Consistent estimation of $\Sigma = \lim_{T\to\infty} E(T^{-1}S_T S_T')$ proceeds along the lines developed in [32] for the univariate case. We define

$$(26) \quad \Sigma_T = \mathrm{var}(T^{-1/2}S_T)$$
$$= T^{-1}\Sigma_1^T E(u_t u_t') + T^{-1}\sum_{\tau=1}^{T-1}\sum_{t=\tau+1}^{T} E(u_t u_{t-\tau}' + u_{t-\tau}u')$$

and the approximant

$$(27) \quad \Sigma_{T,\ell} = T^{-1}\Sigma_1^T E(u_t u_t') + T^{-1}\sum_{\tau=1}^{\ell}\sum_{t=\tau+1}^{T} E(u_t u_{t-\tau}' + u_{t-\tau}u_t') \ .$$

As in [33] we call ℓ the lag truncation number.

LEMMA 3.5. If the sequence $\{u_t\}_1^\infty$ *satisfies:*

(a) $E(u_t) = 0$ *all* t ;

(b) $\sup_t E|u_{it}|^{2+2\eta} < \infty$, *for some* $\eta > 0$ *and all* $i = 1, \ldots, n$;

(c) *either* φ_m *is of size* -2 *or* α_m *is of size* $-(2+2\eta)/\eta$;

and if $\ell \uparrow \infty$ *as* $T \uparrow \infty$ *then*

$$\Sigma_T - \Sigma_{T,\ell} \to 0$$

as $T \uparrow \infty$.

We now define

$$(28) \quad S_{T,\ell} = T^{-1}\Sigma_1^T u_t u_t' + T^{-1}\Sigma_{\tau=1}^{\ell}\Sigma_{t=\tau+1}^{T}(u_t u_{t-\tau}' + u_{t-\tau}u_t') \ .$$

The following result, which generalizes Theorem 5.2 of [33], establishes that under suitable conditions on the rate at which $\ell \uparrow \infty$ as $T \uparrow \infty$ we may consistently estimate ℓ by sequentially estimating $\Sigma_{T,\ell}$.

THEOREM 3.6. If the sequence $\{u_t\}_1^\infty$ *satisfies the conditions of either Theorem* 2.1 *or* 2.2 *and if*

(a) $\sup_t E|u_{it}|^{4(r+\delta)} < \infty$ *for some* $\delta > 0$ *and* $r > 1$ *and all* $i = 1, \ldots, n$;

(b) *either* φ_m *is of size* -2 *or* α_m *is of size* $-2(r+\delta)/(r+\delta-1)$ *with* $r > 1$ *and* $\delta > 0$ *as in* (a);

(c) $\ell \uparrow \infty$ *as* $T \uparrow \infty$ *such that* $\ell = o(T^{1/4})$ *then* $S_{T,\ell} \underset{p}{\to} \Sigma$ *as* $T \uparrow \infty$.

The consistent estimates S_u and $S_{T\ell}$ may be used to construct a new test statistic whose limiting distribution under the null hypothesis (20) is free of nuisance parameters. We define:

$$(29)\quad F_s = F - \tfrac{1}{4}\mathrm{tr}\left\{\left[\tfrac{1}{T}y_Ty_T'(S_u^{-1} - S_{T\ell}^{-1})y_Ty_T' + (S_u - S_{T\ell})\right]\left(\tfrac{1}{T^2}Y_{-1}'Y_{-1}\right)^{-1}\right\} .$$

The limiting distribution of F_s is given by:

THEOREM 3.7. If the conditions of Theorem 3.6 *are satisfied then as* $T \uparrow \infty$:

$$(30)\quad F_s \Rightarrow (1/4)\mathrm{tr}\left\{[W(1)W(1)' - 1][W(1)W(1)' - I][\textstyle\int_0^1 W(t)W(t)'dt]^{-1}\right\} .$$

under the null hypothesis that $A = I$ *in* (10a).

Note that the limiting distribution of the modified Wald statistic F_s as given by (30) is the same as that of F itself when the innovation sequence $\{u_t\}_1^\infty$ is white noise (cf. (26) and (27) above). Moreover, in the univariate case $(n = 1)$ the limiting distribution of F_s is the limiting distribution of the square of the modified t ratio statistic z_t in [33] (see equation (22) and Theorem 6.1 of [33]).

The limiting distribution of F_s is represented in (30) as a functional of the multivariate Wiener process $W(t)$ on $C[0,1]^n$. This distribution is presented tabulated only for the univariate case $n = 1$ (see [6] and [7]). It is of some interest, therefore, to try to develop alternative test procedures which rely on conventional tabulated distributions.

In place of F_s we introduce a new statistic defined by

$$(31)\quad G = \mathrm{tr}\left\{[T(A^*-I) - (1/2)(T^{-1}y_Ty_T' - S_u)(T^{-2}Y_{-1}'Y_{-1})^{-1}]' \right.$$
$$\left. [T(A^*-I) - (1/2)(Ty_Ty_T' - S_u)(T^{-2}Y_{-1}'Y_{-1})^{-1}]\right\} + T^{-2}y_T'S_T^{-1}y_T .$$

THEOREM 3.8. If the sequence $\{u_t\}_1^\infty$ *satisfies the conditions of Theorem* 2.1 *or* 2.2 *and if the null hypothesis* (20) *is true then* $G \Rightarrow \chi_n^2$ *as* $T \uparrow \infty$.

In the univariate case $(A = a\ ,\ S_u = s_u\ ,\ S_{T\ell} = s_{T\ell})$ the statistic G reduces to:

$$(32)\quad G = (T^{-2}\Sigma_1^T y_{t-1}^2)^{-1}\left[T^{-1}\Sigma y_{t-1}(y_t - y_{t-1}) - (1/2)(y_T^2/T - s_u)\right] + y_T^2/Ts_{T\ell}\ .$$

For stable alternatives $(a < 1)$ $T^{-1/2}y_T \underset{p}{\rightarrow} 0$ as $T \rightarrow \infty$ and the first term of G diverges. Thus, the distribution of G diverges as $T \rightarrow \infty$ and the test is consistent against stable alternatives. When the true model is explosive $(a > 1)$ the sample moments that appear in the construction of G all diverge with the sample size and the test is again consistent.

For the G statistic, we recommend that the $S_{T\ell}$ estimate be constructed from the unconstrained OLS residuals. If the errors in the model fulfill the requirements of Theorem 2.1, then $S_{T\ell}$ will converge to a non degenerate limit. Thus, if the $\{y_T\}_1^\infty$ sequence follows an explosive path the quadratic form $T^{-1}y_T'S_{T\ell}^{-1}y_T$ will asymptotically diverge to infinity.

The intuition behind the G statistic is straightforward. The null hypothesis that $A = I$ imposes well defined rates of convergence on the sample moments of $\{y_t\}_1^\infty$. By constructing a statistic which explodes when these rates of convergence are violated, one thereby ensures test consistency.

4. MULTIPLE TIME SERIES REGRESSIONS ON INTEGRATED PROCESSES

We shall examine the multiple regression equation

$$(33)\qquad y_t = Ax_t + u_t\ ;\quad (t = 1, 2, \ldots)$$

where

$$(34)\qquad x_t = x_{t-1} + v_t\ ;\quad (t = 1, 2, \ldots)\ ,\quad x_0 = \text{constant or random.}$$

In (33) A is an $n \times m$ matrix of unknown coefficients. We shall require the sequence $\{(u_t, v_t)\}_1^\infty$ of joint innovations on (33) and (34) to satisfy the general conditions of Theorem 2.1. This allows for a wide range of serial dependence and simultaneity as well as nonstationarity in the system (33) and (34). Thus (34) may be regarded as the reduced form of a simultaneous equations system in which the exogenous variables x_t are driven by a quite general integrated process of order one, such as a vector ARIMA model, and none of the common exogeneity conditions need necessarily apply since we allow for contemporaneous correlations of the form $E(x_t u_t') \neq 0$.

Define the least squares regression estimators

$$A^* = Y'X(X'X)^{-1} \tag{35}$$

and

$$\Omega^* = T^{-1}Y'(I - P_X)Y \tag{36}$$

where $X' = [x_1, \ldots, x_T]$.

In the theory that follows it will be convenient to set $z_t' = (u_t', v_t')$ and define

$$\Omega_z = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(z_t z_t') \tag{37}$$

$$\Omega = \lim_{T\to\infty} E(T^{-1}P_T P_T') \tag{38}$$

where $P_T = \Sigma_1^T z_t$. We shall employ block partitions of matrices such as Σ in (14) and use the notation

$$\Omega = \begin{array}{cc} n & m \\ \left[\begin{array}{cc} [\Omega]_{11} & [\Omega]_{12} \\ [\Omega]_{21} & [\Omega]_{22} \end{array}\right] & \begin{array}{c} n \\ m \end{array} \end{array}$$

to signify the component submatrices.

THEOREM 4.1. If the sequence $\{z_t\}_1^\infty$ *satisfies the conditions of Theorem* 2.1 *or* 2.2 *then as* $T \uparrow \infty$

(a) $T(A^*-A) \Rightarrow \frac{1}{2}[\Omega^{1/2}V(1)V(1)'\Omega^{1/2} + \Omega_z]_{12} \cdot \{[\Omega^{1/2}\int_0^1 V(t)V(t)'dt\Omega^{1/2}]_{22}\}^{-1}$

(b) $A^* \xrightarrow[p]{} A$

(c) $\Omega \xrightarrow[p]{} [\Omega_z]_{11}$

where $V(t)$ *is a multivariate Wiener process in* $C[0,1]^{m+n}$.

Theorem 4.1 shows that least squares regression is consistent in multivariate regression where the regressors are contemporaneously correlated with the errors and where both errors and regressors may be jointly determined by quite general time series processes. The central requirement of the result is that the regressors follow an integrated process such as (34). Note that this implies that y_t is also an integrated process. In the special case where the sequence $\{u_t\}_1^\infty$ is stationary the processes y_t and x_t are cointegrated in the sense of Granger and Engle [15]. Thus, Theorem 4.1 provides the correct asymptotic theory for linear regression amongst cointegrated variates.

It is an interesting consequence of Theorem 4.1 that there is no asymptotic simultaneous equations bias or measurement error bias in regressions such as (33) when the regressors form an integrated process. This result, which seems never before to have appeared in the literature, has quite a simple intuitive explanation. In the usual theory of time series regression amongst stationary, ergodic processes sample moment matrices converge to constant matrices. The bias that arises from contemporaneous correlation between the regressors and the errors is determined by such a sample moment matrix, which converges to a constant non zero matrix in the stationary, ergodic case. But when the regressor process is integrated of order one and hence nonstationary and nonergodic, as it is above, the usual sample moment matrix of the regressors diverges while the matrix of sample moments between the regressors and the errors converges weakly to a random matrix. Upon appropriate renormalization (as is clear in the proof of Theorem 4.1) the sample moment matrix of the regressors also converges weakly to a random matrix. But the signal that comes from the observed sample variation of the regressors is stronger by an order of magnitude (in the sample size) than the sample correlation of the errors and the regressors. It is this fact which eliminates the simultaneous equations and measurement error bias, at least asymptotically, for integrated processes. A related phenomenon occurs in the case of simultaneous equations with trending regressors as has been known for some time, although simple illustrations seem only recently to have appeared in the literature [23]. The examples that are reported in [23] involve only white noise errors. They may, in fact, be extended using the methods of this paper to include quite general weakly dependent and heterogeneously distributed innovations.

THEOREM 4.2. Let $w_t = (u_{it}^2 - E(u_{it}^2))_{n\times 1}$. *If* $\{w_t\}_1^\infty$ *and* $\{z_t\}_1^\infty$ *satisfy the conditions of Theorem* 2.2, *then as* $T \uparrow \infty$

$$T^{1/2}\mathrm{vec}(\Omega^* - \Omega_u) \Rightarrow N(0,V)$$

where

$$\Omega_u = [\Omega_z]_{11} = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(u_t u_t') \ ; \tag{39}$$

$$V = P_D \Sigma_{k=0}^\infty \{\Psi_k - (\mathrm{vec}\ \Omega_u)(\mathrm{vec}\ \Omega_u)'\} P_D \ ; \tag{40}$$

and

$$\Psi_k = E(u_t u_{t+k}' \otimes u_t u_{t+k}') \ ; \quad k = 0,\ 1,\ 2,\ \ldots \tag{41}$$

Theorem 4.2 gives the asymptotic theory for the estimated error covariance matrix in the regression (33). As with the nonstationary VAR (cf. Theorem 3.3) the asymptotic distribution of Ω^* is normal with a covariance matrix that depends on the fourth order cumulant sequence of the innovation process $\{u_t\}$. Note also that the limiting distribution given in Theorem 4.2 is equivalent to that of the error covariance matrix from the VAR given in Theorem 3.3. Thus, for these sample regression characteristics at least, the asymptotic theory in a multiple regression corresponds to that of a vector autoregression, as indeed it does in the stationary case.

5. HYPOTHESIS TESTING IN MULTIPLE REGRESSION WITH INTEGRATED PROCESSES

We shall consider linear hypotheses that involve the elements of the coefficient matrix A in the multiple regression (33). Thus, the null and alternative hypotheses have the form:

$$H_0 : R \operatorname{vec} A = r , \tag{42a}$$

$$H_1 : R \operatorname{vec} A \neq r \tag{42b}$$

where R and r are known $q \times nm$, $q \times 1$ matrices, respectively, and R has full rank q .

To test H_0 we commonly employ the Wald statistic:

$$F = (R \operatorname{vec} A^* - r)'[R(\Omega^* \otimes (X'X)^{-1}R']^{-1}(R \operatorname{vec} A^* - r) . \tag{43}$$

Our next result gives the asymptotic distribution of F in the present nonstationary case.

THEOREM 5.1. If the null hypothesis (42a) *is true and if the sequence* $\{z_t\}_1^\infty$ *satisfies the conditions of Theorem* 2.1 *or* 2.2 *then as* $T \to \infty$:

$$\begin{aligned} F \Rightarrow\ & (1/4)\operatorname{vec}([\Omega^{1/2}V(1)V(1)'\Omega^{1/2} + \Omega_z]_{12})'\{I \otimes [\Omega^{1/2}\textstyle\int_0^1 V(t)V(t)'dt\Omega^{1/2}]_{22}^{-1}\}R' \\ & \cdot[R\{\Omega_u \otimes [\Omega^{1/2}\textstyle\int_0^1 V(t)V(t)'dt\Omega^{1/2}]_{22}^{-1}\}R']^{-1} \\ & \cdot R\{I \otimes [\Omega^{1/2}\textstyle\int_0^1 V(t)V(t)'dt\Omega^{1/2}]_{22}^{-1}\}\operatorname{vec}([\Omega^{1/2}V(1)V(1)'\Omega^{1/2} + \Omega_z]_{12}) \end{aligned} \tag{44}$$

In spite of its complication the asymptotic result (36) is very interesting. It suggests that, in general, the limiting distribution of F depends on the nuisance parameter matrices Ω and Ω_z. Even when $\Omega = \Omega_z$, as in the case of a white noise innovation sequence $\{z_t\}_1^\infty$, the limiting distribution (44) is still dependent on the matrix of nuisance parameters Ω. The situation is thereby akin to that of finite sample distribution theory where the presence of nuisance parameters is the rule rather than the exception.[1]

Theorem 5.1 shows how very special the usual theory of inference for regression models really is. It is a particular feature of the limiting normal distribution theory that the nuisance parameters are concentrated in the covariance matrix of the asymptotic distribution. Since this matrix may be consistently estimated from the data, a quadratic form in an appropriate metric yields the usual asymptotic chi-squared criterion and removes the nuisance parameters. The nonnormality of the limiting distribution theory in the present case eliminates the possibility of such a simple transformation. In particular, we recall from Theorem 4.1 that $T(R\,\mathrm{vec}\,A^* - r)$ has a nonnormal and asymmetric limiting distribution under the null hypothesis (42a). Moreover, the conventional metric that is embodied in the matrix of the quadratic form of the Wald statistic (44) no longer delivers a relevant measure of distance according to which departures of the vector $T(R\,\mathrm{vec}\,A^* - r)$ from zero may reasonably be measured. More precisely, the limit of $\{R(\Omega^* \otimes (X'X/T^2)^{-1}R'\}^{-1}$ is now a random matrix. It is no longer reasonable to think of it as a precision matrix on the support, R^q, of the limiting distribution of $T(R\,\mathrm{vec}\,A^* - r)$. Thus, the Wald statistic F,

---

[1]The concluding section of [32] contains a detailed discussion of this point.

as given by (36), lacks its usual rationale.

The situation described above is not changed by the use of other test criteria such as the likelihood ratio (LR) or Lagrange multiplier (LM) tests. In fact, although we shall not prove it here, the Wald, the LR and the LM test are equivalent asymptotically in the present context, just as they are in the more conventional setting. In each case the limiting distribution theory is given by the result of Theorem 5.1.

6. MODELS WITH FITTED DRIFT VECTORS

The results in Sections 3-5 may be extended quite easily to models with fitted means or fitted means and time trends. Here we shall report the main results for vector autoregressions with a fitted drift. A more complete treatment in the scaler case is available in the doctoral dissertation [31].

Consider the vector autoregression:

$$y_t = \hat{\mu} + \hat{A}y_{t-1} + \hat{u}_t \tag{45}$$

where $\hat{\mu} = \bar{y} - \hat{A}\bar{y}_{-1}$ and $\hat{A} = [\Sigma_1^T(y_t - \bar{y})(y_{t-1} - \bar{y}_{-1})']$ covariance matrix estimator is now $\hat{\Sigma}_u = T^{-1}Y'(I - P_z)Y$ where the regressor matrix $Z = [i \vdots Y_{-1}]$ and i is the $T \times 1$ sum vector.

THEOREM 6.1. If the conditions of Theorem 3.2 *are satisfied then as* $T \uparrow \infty$:

a. $T(\hat{A}-I) \Rightarrow \left[\frac{1}{2}\{\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_u\} - \Sigma^{1/2}W(1)\int_0^1 W(t)'dt\Sigma^{1/2}\right]\left[\Sigma^{1/2}\{\int_0^1 W(t)W(t)'dt - \int_0^1 W(t)dt\int_0^1 W(t)'dt\}\Sigma^{1/2}\right]^{-1} = \xi$;

b. $\hat{A} \underset{p}{\to} I$;

c. $\hat{\mu} \underset{p}{\to} 0$;

d. $T^{1/2}\hat{\mu} \Rightarrow \Sigma^{1/2}W(1) - \xi\Sigma^{1/2}\int_0^1 W(t)dt$;

e. $\hat{\Sigma}_u \underset{p}{\rightarrow} \Sigma_u$.

Moreover, if the conditions of Theorem 3.3 *are satisfied then*

f. $T^{1/2}\text{vec}(\hat{\Sigma} - \Sigma_u) \Rightarrow N(0,V)$

where the asymptotic covariance matrix V *is given by* (40) *above.*

For testing the unit roots hypothesis (10b) we could use extensions of the Wald statistic F defined by (17) for case of the known zero drift model. This would lead to distributional results such as those given by Theorems 3.4 and 3.7 with attendant problems of nuisance parameter matrices and tabulation. An alternative, which we consider here, is to extend the test statistic G defined by (31) to the case of a fitted dirft vector. Specifically, we define:

$$(46)\quad G_\mu = \text{tr}\Big[\Big\{T(\hat{A}-I) - [(1/2)(T^{-1}y_Ty_T' - S_u) - T^{-1}y_T\bar{y}'][T^{-2}\Sigma_1^T(y_t-\bar{y})(y_t-\bar{y})']^{-1}\Big\}$$
$$\Big\{T(\hat{A}-I) - [(1/2)(T^{-2}y_Ty_T' - S_u) - T^{-1}y_T\bar{y}'][T^{-2}\Sigma_1^T(y_t-\bar{y})(y_t-\bar{y})']\Big\}\Big]^{-1}$$
$$+ T^{-1}y_T'S_{T\ell}^{-1}y_T .$$

THEOREM 6.2. If the conditions of Theorem 3.8 *are satisfied then as* $T \uparrow \infty$ $G_\mu \Rightarrow \chi_n^2$.

This result provides us with a simple χ^2 test of the unit roots hypothesis (20).

7. CONCLUSIONS

Models with integrated regressors and cointegrated variables seem likely to become an important focal point of future macroeconomic theory and debate. At the theoretical level, efficient macroeconomic equilibrium implies that a number of macroeconomic time series must behave as martingales. As argued by Hall [17], the fulfillment of first order conditions for intertemporal utility maximization requires that consumption follows a random walk. Rationality of expectations implies that a sequence of forecasts of a given event form a martingale. On the empirical level, evidence has mounted that a number of time series, ranging from GNP (Nelson and Plosser [30]) to dividends (Marsh and Merton [29]) are integrated processes. Statistical estimation and inference in these models requires a methodology which accounts for the nonstationarity and nonergodicity of the underlying time series.

This paper has developed a general asymptotic theory for regressions with integrated processes. The asymptotic distribution of regression coefficients and covariance matrix estimators in these models have been derived. Least squares estimation procedures produce consistent coefficient estimates which converge at a faster rate $(O_p(T^{-1}))$ than in conventional regression theory. However, these estimates are not asymptotically normal when appropriately centered and scaled. The nonnormality results from the fact that suitably scaled sample moments (and also the hessian) converge weakly to random matrices rather than constant matrices. Our development and employment of a multivariate functional central limit theory has permitted explicit derivation of the nonnormal asymptotics.

We have further analyzed the asymptotic properties of conventional statistical tests in the context of these regressions. Standard tests, such as the Wald test, no longer yield asymptotically distributed χ^2 criteria.

This is because the metric underlying the Wald test is no longer relevant when the limiting distribution of the regression coefficients is nonnormal. Moreover, in multiple equation systems the asymptotic distribution of the usual test criteria are dependent upon the limiting covariance matrix of the contemporaneous innovations across equations and the limiting covariance matrix of the accumulated sums of past innovations. In the case of weakly dependent innovation sequences these limiting covariance matrices are different whereas for i.i.d. innovations they are the same. Interestingly, in both cases the asymptotic distribution of the usual test criteria is parameter dependent. For these reasons, it is important to devise new statistical tests whose asymptotic distributions, at least, are free of nuisance parameters. For the case of testing for unit roots in vector autoregressions we demonstrate the existence of transformations of the Wald statistic which are independent of nuisance parameters. We also develop for this case a new test statistic which is asymptotically χ^2, so that conventional tabulations may be used in testing.

Our asymptotic results are compatible with a wide range of innovation processes and model formulations. They allow for a high degree of serial correlation and moderate heterogeneity in the error processes. Moreover, our conditions permit regressor-error correlation and therefore apply to simultaneous equations systems with integrated processes. As a result, conventional measurement error and simultaneity bias do not generally arise in these models when they are formulated and estimated with integrated processes.

One area of future research lies in the modelling of integrated process regressions when the integrated regressors are themselves cointegrated. The multivariate regression theory in Sections 4-6 presupposes that no cointegration exists amongst the regressors. This requirement is not innocuous. Sims [39], for example, has indicated that in the AR(2):

$$y_t = \alpha y_{t-1} + \beta y_{t-2} + u_t$$

with $\alpha+\beta = 1$ the least squares estimates $\hat{\alpha}$ and $\hat{\beta}$ each possess limiting marginal normal distributions with an $O(T^{-1/2})$ convergence rate. Intuitively, the leverage needed to generate $O(T^{-1})$ convergence requires that the regressors not become asymptotically collinear, as occurs in the Sims example. Research into models that combine integrated, cointegrated and nonintegrated regressors is currently being pursued by the authors.

MATHEMATICAL APPENDIX

Skorohod Metric

The specific metric we use to render D[0,1] separable is:

(A1) $$d(x,y) = \inf_{\varepsilon>0}\{\varepsilon : ||\lambda|| \leq \varepsilon, \sup_t |X(t) - Y(\lambda(t))| \leq \varepsilon\}$$

where λ is any continuous mapping of [0,1] onto itself with $\lambda(0) = 0$, $\lambda(1) = 1$ and

(A2) $$||\lambda|| = \sup_{t \neq s} \left| \log \frac{\lambda(t) - \lambda(s)}{t-s} \right| , \quad t, s \in [0,1] .$$

This metric may be contrasted with the uniform metric

(A3) $$d(x,y) = \sup_t |X(t) - Y(t)| .$$

The two metrics coincide when $X(t)$ and $Y(t)$ are continuous. For D[0,1] , the Skorohod metric differs from the uniform for points of discontinuity. Two functions in the uniform metric are close only if their discontinuities are of approximately equal magnitude and occur at exactly the same values t_i , whereas under the Skorohod metric the discontinuities need not occur at exactly the same t_i .

Following an example of Billingsley [2], one can easily see that under the uniform metric D[0,1] is not separable. Consider functions of the form

$$X_\delta(t) = \begin{cases} 0 & 0 \leq t \leq \delta \\ 1 & \delta \leq t \leq 1 \end{cases}$$

Under the uniform metric $d(x_{\delta_1}, x_{\delta_2}) = 1$ if $\delta_1 \neq \delta_2$. This would imply the existence of an uncountable set of elements mutually separated by a non negligible distance. Since each $x_{\delta_i}(t)$ may be surrounded by an open sphere $S(x_{\delta_i}(t), 1/2)$ which does not intersect an open sphere surrounding another $x_{\delta_j}(t)$, there cannot exist a countable open covering for the $x_\delta(t)$ corresponding to the uncountable open covering generated by $\bigcup_\delta S(x_\delta(t), 1/2)$. The failure of the latter property is equivalent to nonseparability since an uncountable subcover of open spheres of radius 1/2 over all points in the space will not possess a countable subcover.

Proof of Theorem 2.1

The result holds for $n = 1$. In addition, the marginal distributions all converge to univariate Wiener processes, as the requirements of the univariate McLeish invariance principle hold for each element of $X_T(t)$. Thus we need to demonstrate that the convergence of the marginal probability measures implies convergence of the joint probability measure.

The proof of convergence proceeds in two steps. First, we demonstrate that $X_T(t)$ converges to $W(t)$ over a determining subclass in $D[0,1]^n$. A determining class is a class of subsets of $D[0,1]^n$ such that any two measures which coincide over the class must coincide over $D[0,1]^n$. A determining subclass insures equality of two measures over the σ-algebra generated by the class, when they coincide on the class. Thus, if $X_T(t)$ possesses a limiting probability measure, it must equal $W(t)$. Second, we demonstrate that the sequenceof probability measures $P_T(\ )$ associated with $X_T(t)$ is "tight" which will imply from Prohorov's Theorem (see Billingsley [2]) that the limit of $P_T(\ )$ is a properly defined probability measure. By the equality of this limiting probability measure with $W(t)$ over the determining subclass on $D[0,1]^n$, this limiting measure must equal $W(t)$.

We first define the finite dimensional distributions of $X_T(t)$. For each $x \in D[0,1]^n$ define the mapping:

(A4) $\Pi_{t_1 \ldots t_k}(x) = [x(t_1), \ldots, x(t_k)]$ a $k \times n$ matrix.

We may also consider the inverse mapping, since $\Pi_t(x)$ is measurable with respect to the Skorohod topology,

(A5) $\Pi^{-1}_{t_1 \ldots t_k}(H)$, $H \in R^{k \times n}$

where each point in H identifies a vector of values for X at k different values t_i . This inverse mapping is equivalent to considering the distribution of x at k distinct elements of $[0,1]$.

Our first result is Lemma 2.1.

LEMMA 1. *The finite dimensional sets* $\Pi^{-1}_{t_1 \ldots t_k}(H)$ *form a determining class over the subspace of* $D[0,1]^n$ *containing* $X_T(t)$ *and* $W(t)$.

Proof: For $n = 1$, Billingsley verifies the result. For higher n , note that the product of determining classes equal the determining classes of the product space, when the coordinate spaces are separable. Separability is necessary, as Halmos [18] verifies.

Now, this result by itself is insufficient as it means we must consider finite dimensional distributions on $D[0,1]^n$ where the t values may vary across the coordinate subspaces. However, $X_T(t)$ and $W(t)$ are constrained to lie in the subspace of $D[0,1]^n$ where each element of the vector function is evaluated at the same t . Thus the "diagonal" of the product of the determining classes will constitute the determining subclass for $X_T(t)$. The sets $\Pi^{-1}_{t_1 \ldots t_k}(H)$ form a base for the diagonal, which renders finite dimensional distributions of (12) a determining subclass.

LEMMA 2. $X_T(t) \Rightarrow W(t)$ *over the finite dimensional sets* $\Pi^{-1}_{t_1 \ldots t_k}(H)$.

Proof. Since each marginal distribution of $X_T(t)$ converges, for fixed t to a Wiener process, $\lambda' X_T(t)$ will converge to $W(t)$ for all $\lambda'\lambda = 1$. This follows as:

$$\text{(A6)} \qquad \lambda' X_T(t) = \frac{1}{\sqrt{T}} \Sigma_{t=1}^{j-1} \lambda' \Sigma^{-1/2} u_t$$
$$= \frac{1}{\sqrt{T}} \Sigma_{t=1}^{j-1} V_t$$

where $\{V_t\}$ is a scalar sequence which fulfills the requirements of the McLeish invariance principle. The variance of V_t will converge to

$$\text{(A7)} \qquad \sigma^2 = \lim_{T\to\infty} E\left(\frac{1}{T} S_T^2\right)$$
$$= \lambda' \Sigma^{-1/2} E\left(\frac{1}{T} \Sigma_{j=1}^{T} u_j u_j'\right) \Sigma^{-1/2} \lambda$$
$$= 1$$

by assumptions of the theorem.

Since λ was arbitrary, one can invoke the Cramér-Wold device to conclude that

$$\text{(A8)} \qquad X_T(t) \Rightarrow W(t) \quad \text{fixed } t \text{ .}$$

Clearly, by straightforward use of the Cramér-Wold device, and use of independence of $W(t_1 - t_0)$ and $W(t_0)$

$$\text{(A9)} \qquad X_T(t) \to W(t) \quad \text{for any } t_1, \ldots, t_k \text{ ,}$$

which implies that the finite dimensional distributions of $X_T(t)$ and $W(t)$ coincide.

Thus, if $X_T(t)$ has a limiting joint probability measure, this measure will coincide with $W(t)$.

By the Helly selection theorem, the sequence of $P_T(\ )$ probability measures associated with $X_T(t)$ will converge to a measure with every requirement of a probability measure except (possibly) $\lim_{T\to\infty} P_T(D[0,1])^n) = 1$. This last requirement will be fulfilled according to Prohorov's Theorem if $P_T(\ )$ is tight (i.e. there exists a compact set A such that $P_T(A) > 1-\varepsilon \quad \forall T$).

Tightness of $\{P_T(\ )\}$ follows from:

LEMMA 3. The joint probability measure on a separable product space is tight if each marginal probability measure is tight.

Proof. Let $S = S_1 \times \ldots \times S_k$ be a product space. The marginal probability measure P_i is defined by

(A10) $$P_i(A_i) = P(S_1 \times \ldots \times A_i \times S_{i+1} \times \ldots \times S_k) .$$

If each marginal probability measure is tight, then there exist a sequence of sets $A_i \in S_i$ such that

(A11) $$P_i(A_i) > 1 - \varepsilon/2^i .$$

Thus the compact set $A = A_1 \times \ldots \times A_k$ must from (15) and (16) obey

(A12) $$P_T(A) > 1-\varepsilon$$

thus $P_T(\ )$ is tight.

Now, each marginal probability measure for $X_T(t)$ is tight on $D[0,1]$. Thus the joint measure for $X_T(t)$ is also tight, which completes the proof.

Proof of Lemma 3.1

The proof follows the same lines as that of Theorem 3.1 of [33] and Lemma 2.1 of [34]. Thus to prove (a) we have

$$\text{(A13)}\qquad T^{-3/2}\Sigma_1^T y_t = T^{-3/2}\Sigma_{i=1}^T (S_{i-1} + u_i + y_0)$$

$$= \Sigma^{1/2}\Sigma_1^T \int_{(i-1)/T}^{i/T} X_T(t)dt + o_p(1)$$

$$= \Sigma^{1/2}\int_0^1 X_T(t)dt + o_p(t)$$

$$\Rightarrow \Sigma^{1/2}\int_0^1 W(t)dt$$

by Theorem 1.1 and the continuous mapping theorem. The proofs of (b) and (c) are entirely analogous.

To prove (d) we define the following functional on $C[0,1]^n$:

$$\text{(A14)}\quad Z_T(t) = \frac{1}{\sqrt{T}}\Sigma^{-1/2}S_{[Tt]} + \frac{Tt - [Tt]}{\sqrt{T}}\Sigma^{-1/2}u_{[Tt]+1}\ ; \quad (j-1)/T \leq t < j/T$$

$$(j = 1, \ldots, T)$$

$$\text{(A15)}\quad Z_T(1) = \frac{1}{\sqrt{T}}\Sigma^{-1/2}S_T\ .$$

Note that $Z_T(t) = X_T(t) + o_p(1)$ and thus $Z_T(t) \Rightarrow W(t)$ by Theorem 2.1 above and [33, Theorem 4.1]. Moreover, by direct integration (noting that $dZ_T(t) = T^{1/2}\Sigma^{-1/2}u_j dt$ for $(j-1)/T \leq t < j/T$) we find that:

$$\text{(A17)}\quad \int_{(j-1)/T}^{j/T} Z_T(t)dZ_T(t)' - [Z_T(t)Z_T(t)']_{(j-1)/T}^{j/T} - \int_{(j-1)/T}^{j/T} Z_T(t)dZ_T(t)'$$

so that

$$\text{(A18)} \quad \int_{(j-1)/T}^{j/T} Z_T(t)\,dZ_T(t)' = \tfrac{1}{2}[Z_T(t)Z_T(t)']_{(j-1)/T}^{j/T} .$$

Summing (A18) over $j = 1, \ldots, T$ and combining (A16) and (A18) we find:

$$\text{(A19)} \quad T^{-1}\Sigma^{-1/2}(\Sigma_{j=1}^{T} S_{j-1}u_j')\Sigma^{-1/2} = \tfrac{1}{2}Z_T(1)Z_T(1)' - \tfrac{1}{2T}\Sigma^{-1/2}(\Sigma_1^T u_j u_j')\Sigma^{-1/2} .$$

Hence, by the weak convergence of $Z_T(t)$ $W(t)$, the strong law of large numbers for dependent variables [25] and the continuous mapping theorem, we deduce that:

$$T^{-1}\Sigma_1^T y_{t-1}u_t' = T^{-1}\Sigma_{j=1}^T (S_{j-1} + y_0)u_j' \Rightarrow \tfrac{1}{2}\{\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_u\}$$

as required.

Proof of Theorem 3.2. Let $U' = [u_1, \ldots, u_T]$ and we have:

$$\text{(A20)} \qquad A^* = I + U'Y_{-1}(Y_{-1}'Y_{-1})^{-1}$$

Then

$$\text{(A21)} \qquad T(A^*-I) = (T^{-1}U'Y_{-1})(T^{-2}Y_{-1}'Y_{-1})^{-1}$$

$$\Rightarrow \tfrac{1}{2}\{\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_u\}\left\{\Sigma^{1/2}\int_0^1 W(t)W(t)'dt\,\Sigma^{1/2}\right\}^{-1}$$

proving (a). (b) follows directly. To prove (c) note that

$$\text{(A23)} \quad \Sigma^* = T^{-1}U'U - T^{-1}(U'Y_{-1})(Y_{-1}'Y_{-1})^{-1}(Y_{-1}'U) \underset{p}{\to} \Sigma_u$$

as required.

Proof of Theorem 3.3

Write

$$\text{(A24)}\qquad T(\Sigma^* - \Sigma_u) = \sqrt{T}(T^{-1}U'U - \Sigma_u) - \frac{1}{\sqrt{T}}\left(\frac{U'Y_{-1}}{T}\right)\left(\frac{Y'_{-1}Y_{-1}}{T^2}\right)^{-1}\left(\frac{Y'_{-1}U}{T}\right)$$

$$= \sqrt{T}(T^{-1}U'U - \Sigma_u) + o_p(1)$$

$$= \Sigma_1^T(u_tu_t' - \Sigma_u)/\sqrt{T} + o_p(1) \ .$$

By assumption $\{(u_tu_t' - \Sigma_u)\}_1^\infty$ is a weakly stationary sequence of random matrices with zero mean that satisfies the moment and mixing conditions of Theorem 2.2. Define

$$\text{(A25)}\qquad w_t = (D'D)^{-1}D'\text{vec}(u_tu_t' - \Sigma_u) = L\ \text{vec}(u_tu_t' - \Sigma_u)$$

where D is the $n^2 \times n(n+1)/2$ duplication matrix and $L = D^-$ is the $n(n+1)/2 \times n^2$ elimination matrix [27]. By Theorem 2.2

$$\text{(A26)}\qquad T^{-1/2}\Sigma_1^T w_t \Rightarrow N(0,Q)$$

where

$$\text{(A27)}\qquad Q = E(w_1w_1') + \Sigma_{k=2}^\infty\{E(w_1w_k') + E(w_kw_1')\} \ .$$

Now

$$\text{(A28)}\qquad E(w_1w_1') = L\{\Psi_0 - (\text{vec}\ \Sigma_u)(\text{vec}\ \Sigma_u)'\}L'$$

and

$$\text{(A29)}\qquad E(w_1w_k') = L\{\Psi_k - (\text{vec}\ \Sigma_u)(\text{vec}\ \Sigma_u)'\}L'$$

where

(A30) $$\Psi_k = E(u_t u'_{t+k} \times u_t u'_{t+k}) \; ; \quad k = 0, 2, \ldots .$$

It follows that

(A31)
$$\begin{aligned} \sqrt{T}\, \mathrm{vec}(\Sigma^* - \Sigma_u) &= T^{-1/2}\Sigma_1^T \mathrm{vec}(u_t u'_t - \Sigma_u) + o_p(1) \\ &= T^{-1/2}\Sigma_1^T D w_t + o_p(1) \\ &\Rightarrow N(0,\ DQD') \equiv N(0,V) \end{aligned}$$

as required.

Proof of Theorem 3.4

Under the null hypothesis (20) as $T \uparrow \infty$

(A32) $$T\,\mathrm{vec}(A^*-I) \Rightarrow \mathrm{vec}\left\{\frac{1}{2}[\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_u]\left[\Sigma^{1/2}\int_0^1 W(t)W(t)'dt\Sigma^{1/2}\right]^{-1}\right\}$$

and $\Sigma^* \xrightarrow[p]{} \Sigma_u$ by Theorem 3.2. Additionally,

(A33) $$T^{-2}Y'_{-1}Y_{-1} \Rightarrow \Sigma^{1/2}\int_0^1 W(t)W(t)'dt\Sigma^{1/2}$$

by Lemma 3.1. Thus the stated result (22) follows by the continuous mapping theorem.

Proof of Lemma 3.5. The result follows by choosing an arbitrary element of the matrix $\Sigma_T - \Sigma_{T,\ell}$ and proceeding as in Lemma 5.1 of [33].

Proof of Theorem 3.6. The result follows by taking an arbitrary element of the matrix $S_{T,\ell} - \Sigma_{T,\ell}$ and proceeding as in Theorem 5.2 of [33].

Proof of Theorem 3.7

Note that

$$\text{(A34)}\qquad T^{-1/2}y_T = T^{-1/2}\Sigma_1^T u_t + T^{-1/2}y_0$$

$$\Rightarrow \Sigma^{1/2}W(1)$$

by Theorem 2.1. Moreover, under the conditions of the Theorem $S_u \underset{p}{\to} \Sigma_u$ and $S_{T\ell} \underset{p}{\to} \Sigma$ by Theorem 3.6. Also

$$\text{(A35)}\qquad T^{-2}Y_{-1}'Y_{-1} \Rightarrow \Sigma^{1/2}\int_0^1 W(t)W(t)'dt\Sigma^{1/2}$$

as before. Hence by Theorem 3.4 and the continuous mapping theorem

$$\text{(A36)}\quad F_s \Rightarrow \frac{1}{4}\mathrm{tr}\left\{[W(1)W(1)' - I][W(1)W(1)' - I]\left[\int_0^1 W(t)W(t)'dt\right]^{-1}\right\}$$

as required.

Proof of Theorem 3.8

Under the null hypothesis we may write (see (A21) above):

$$\text{(A37)}\quad T(A^*-I)(T^{-1}Y_{-1}'Y_{-1}) = T^{-1}\Sigma_1^T u_t y_{t-1}' = T^{-1}\Sigma_1^T u_t(S_{t-1}+y_0)$$

$$= (1/2)\{\Sigma^{1/2}Z_T(1)Z_T(1)'\Sigma^{1/2} - T^{-1}\Sigma_1^T u_t u_t'\} + T^{-1}\Sigma_1^T u_t y_0$$

where $Z_T(t)$ is defined in (A14) above. Also

$$\text{(A38)}\qquad T^{-1/2}y_T = T^{-1/2}(S_T + y_0) = \Sigma^{1/2}Z_T(1) + T^{-1/2}y_0 \ .$$

Thus

$$T(A^*-I)(T^{-2}Y'_{-1}Y_{-1}) - (1/2)(T^{-1}y_T y'_T - S_u)$$

$$= T^{-1}\Sigma_1^T u_t y'_0 - (1/2)\{T^{-1/2}y_0 Z_T(1)'\Sigma^{1/2} + T^{-1/2}\,{}^{1/2}Z_T(1)y'_0 + T^{-1}y_0 y'_0\} \underset{p}{\to} 0 .$$

Moreover

$$T^{-1/2}S_T^{-1/2}y_T \Rightarrow W(1) = N(0, I_n) .$$

It follows by Theorem 4.1 of [33] and the continuous mapping theorem that as $T \uparrow \infty$ $G \Rightarrow W(1)'W(1) \equiv \chi_n^2$, as required.

Proof of Theorem 4.1

Define

(A39) $$Z_T(t) = \frac{1}{\sqrt{T}}\Omega^{-1/2}p_{j-1} + \frac{Tt - [Tt]}{\sqrt{T}}\Omega^{-1/2}z_j \ ; \quad (j-1)/T \le t < j/T \ , \ (j = 1, \ldots, T)$$

$$Z_T(1) = \frac{1}{\sqrt{T}}\Omega^{-1/2}p_T .$$

By Theorem 2.1 or 2.2 $Z_T(t) \Rightarrow V(t)$ where $V(t)$ is a multivariate Wiener process on $C[0,1]^{m+n}$.

Set $q_0 = 0$ and further define the sequence $\{q_t\}_1^\infty$ by

(A40) $$q_t = q_{t-1} + z_t \ ; \quad t = 1, 2, \ldots .$$

Let $Q' = [q_1, \ldots, q_T]$, $Z' = [z_1, \ldots, z_T]$ and then we have

(A41) $$\begin{aligned} T^{-1}Z'Q &= T^{-1}\Sigma_1^T z_t q'_t \\ &= T^{-1}\Sigma_1^T z_t q'_{t-1} + T^{-1}\Sigma_1^T z_t z'_t \\ &\Rightarrow \left\{\tfrac{1}{2}\Sigma^{1/2}V(1)V(1)'\Sigma^{1/2} - \tfrac{1}{2}\Sigma_z\right\} + \Sigma_z \\ &= \tfrac{1}{2}\left\{\Sigma^{1/2}V(1)V(1)'\Sigma^{1/2} + \Sigma_z\right\} \end{aligned}$$

by arguments entirely analogous to those used in the proof of Lemma 3.1. In a similar way we find

$$T^{-2}Q'Q \Rightarrow \Omega^{1/2}\int_0^1 V(t)V(t)'dt\Omega^{1/2} \ .$$

The stated result (a) now follows since $T^{-1}U'X = [T^{-1}Z'Q]_{12}$, $T^{-2}X'X = [T^{-2}Q'Q]_{22}$ at least up to initial values and

$$T(A^*-A) = (T^{-1}U'X)(T^{-2}X'X)^{-1}$$

(b) follows immediately. To prove (c) we note that

$$\text{(A42)} \qquad \Omega^* = T^{-1}U'U - T^{-1}(T^{-1}U'X)(T^{-2}X'X)^{-1}(T^{-1}X'U)$$

$$\underset{p}{\rightarrow} \lim_{T\to\infty} T^{-1}\Sigma_1^T E(u_t u_t')$$

$$= [\Omega_z]_{11}$$

by the strong law of large numbers for weakly dependent variates [25].

Proof of Theorem 4.2. The proof is entirely analogous to that the Theorem 3.3.

Proof of Theorem 5.1.

Under the null hypothesis (42a) as $T \uparrow \infty$

$$\text{(A43)} \quad T(R\,\mathrm{vec}\,A^* - r) = R\{\mathrm{vec}\ T(A^*-A)\}$$

$$\Rightarrow R\ \mathrm{vec}\ \frac{1}{2}\left\{\Omega^{1/2}V(1)V(1)'\Omega^{1/2} + \Omega_z\right]_{12}[\Omega^{1/2}\int_0^1 V(t)V(t)'dt\Omega^{1/2}]_{22}^{-1}\right\}$$

and $\Omega^* \underset{p}{\rightarrow} [\Omega_z]_{11} = \Omega_u$ by Theorem 4.1. Additionally,

$$\text{(A44)} \qquad T^{-2}X'X \Rightarrow \left[\Omega^{1/2}\int_0^1 V(t)V(t)'dt\Omega^{1/2}\right]_{22} .$$

The stated result follows directly.

Proof of Theorem 6.1

From Lemma 3.1 we know that

$$\text{(A45)} \quad T^{-2}\Sigma_1^T(y_{t-1}-\bar{y}_{-1})(y_{t-1}-\bar{y}_{-1})' \Rightarrow \Sigma^{1/2}\left\{\int_0^1 W(t)W(t)'dt - \int_0^1 W(t)dt\int_0^1 W(t)'dt\right\}\Sigma^{1/2} .$$

Additionally,

$$\text{(A46)} \quad T^{-1}\Sigma_1^T u_t(y_{t-1}-\bar{y}_{-1})' = T^{-1}\Sigma_1^T u_t y_{t-1} - (T^{-1/2}\Sigma_1^T u_t)(T^{-3/2}\Sigma_1^T y'_{t-1})$$

$$\Rightarrow (1/2)\{\Sigma^{1/2}W(1)W(1)'\Sigma^{1/2} - \Sigma_u\} - \Sigma^{1/2}W(1)\int_0^1 W(t)'dt\Sigma^{1/2}$$

by Lemma 3.1 again. Results (a) and (b) follow directly. To prove (c) and (d) note that

$$\text{(A47)} \qquad \hat{\mu} = T^{-1}\Sigma_1^T u_t - T^{1/2}(\hat{A}-I)(T^{-3/2}\Sigma_1^T y_{t-1}) \underset{p}{\rightarrow} 0$$

and

$$\text{(A48)} \qquad T^{1/2}\hat{\mu} \Rightarrow \Sigma^{1/2}W(1) - \xi\Sigma^{1/2}\int_0^1 W(t)dt .$$

To prove (e) and (f) note that

$$\text{(A49)} \qquad \hat{\Sigma} - \Sigma_u = T^{-1}U'U - \Sigma_u - T^{-1}(U'Z)(Z'Z)^{-1}(Z'U)$$

and

$$\text{(A50)} \quad \sqrt{T}(\hat{\Sigma} - \Sigma_u) = \sqrt{T}(T^{-1}U'U - \Sigma_u) - T^{-1/2}(U'ZD_T^{-1/2})(D_T^{-1/2}Z'ZD_T^{-1/2})^{-1}D_T^{-1/2}Z'U .$$

In this expression

$$D_T = \begin{bmatrix} T & 0 \\ 0 & T^2 I_n \end{bmatrix} \tag{A51}$$

and the second term of the expression is $O_p(T^{-1/2})$ as $T \uparrow \infty$. Result (f) now follows as in Theorem 3.3.

Proof of Theorem 6.2

The proof is analogous to that of Theorem 3.8 and is omitted.

REFERENCES

[1] Anderson, T. W. *The Statistical Analysis of Time Series*. John Wiley: New York, 1971.

[2] Billingsley, P. *Convergence of Probability Measures*. John Wiley: New York, 1968.

[3] Brillinger, D. R. and M. Hatanaka. "An harmonic analysis of nonstationary multivariate economic processes," *Econometrica*, 35 (1968), 131-141.

[4] Box, G. E. P. and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden Day: San Francisco, 1976.

[5] Chandra, K. C. "Strong mixing properties of linear stochastic processes," *Journal of Applied Probability*, 11 (1974), 401-408.

[6] Dickey, D. A. and W. A. Fuller. "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74 (1979), 427-431.

[7] Dickey, D. A. and W. A. Fuller. "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, 49 (1981), 1057-1072.

[8] Doan, T., R. B. Litterman and C. Sims. "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3 (1984), 1-100.

[9] Durlauf, Steven Neil. *Estimation and Statistical Inference for Integrated Time Series Models*, Doctoral Dissertation in progress, Yale University, 1986.

[10] Evans, G. B. A. and N. E. Savin. "Testing for unit roots: 1," *Econometrica*, 49 (1981), 753-779.

[11] Evans, G. B. A. and N. E. Savin. "Testing for unit roots: 2," *Econometrica*, 52 (1984), 1241-1269.

[12] Granger, C. W. J. and Oskar Morgenstern. "Spectral analysis of New York Stock Market prices," *Kyklos*, 16 (1963), 1-27.

[13] Granger, C. W. J. and M. Hatanaka. *Spectral Analysis of Economic Time Series*. Princeton: Princeton University Press, 1964.

[14] Granger, C. W. J. "The typical spectral shape of an economic variable," *Econometrica*, 34 (1966), 150-161.

[15] Granger, C. W. J. and R. F. Engle. "Dynamic model specification with equilibrium constraints: Cointegration and error correction," UCSD Discussion Paper No. 85-18, 1985.

[16] Hall, P. and C. C. Heyde. *Martingale Limit Theory and its Applications.* Academic Press: New York, 1980.

[17] Hall, R. E. "Stochastic implication of the life cycle-permanent income hypothesis: Theory and evidence," *Journal of Political Economy*, 86 (1978), 971-987.

[18] Halmos, P. *Measure Theory.* New York: van Nostrand, 1950.

[19] Hannan, E. J. *Multiple Time Series.* New York: Wiley, 1970.

[20] Hatanaka, M. and M. Suzuki. "A theory of the pseudospectrum and its application to nonstationary dynamic econometric models," in *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, M. Shubik, ed. Princeton: Princeton University Press, 1967.

[21] Ibragimov, I. A. and Yu. V. Linnik. *Independent and Stationary Sequence of Random Variables.* Groningen: Wolters-Nordhoff, 1971.

[22] Kelley, J. *General Topology.* New York: van Nostrand, 1955.

[23] Kramer, W. "On the consequences of trend for simultaneous equation estimation," *Economic Letters*, 14 (1984), 23-30.

[24] Litterman, R. B. "Forecasting with Bayesian vector autoregressions: Four years of experience," mimeographed (Federal Reserve Bank of Minneapolis), 1984.

[25] McLeish, D. L. "A maximal inequality and dependent strong laws," *Annals of Probability*, 3 (1975), 829-839.

[26] McLeish, D. L. "Invariance principles for dependent variables," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32 (1975), 165-178.

[27] Magnus, J. R. and H. Neudecker. "The elimination matrix: Some lemmas and applications," *SIAM Journal on Algebraic and Discrete Methods*, 1 (1980), 422-449.

[28] Mann, H. B. and A. Wald. "On the statistical treatment of linear stochastic difference equations," *Econometrica*, 11 (1943), 173-220.

[29] Marsh, Terry A. and Robert C. Merton. "Dividend variability and variance bounds tests for the rationality of stock market prices," Sloan School Working Paper #1584-84, August 1984.

[30] Nelson, Charles R. and Charles I. Plosser. "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10 (1982), 139-162.

[31] Perron, Pierre "Efficient markets and tests of the random walk hypothesis", Yale doctoral dissertation, 1986 (forthcoming).

[32] Phillips, P. C. B. "Exact small sample theory in the simultaneous equations model," Ch. 8 of *Handbook of Econometrics*, Vol. I, M. D. Intriligator and Z. Gricliches (eds.), 1983, pp. 449-516.

[33] Phillips, P. C. B. "Time series regression with unit roots," Cowles Foundation Discussion Paper No. 740, Yale University, 1985.

[34] Phillips, P. C. B. "Understanding spurious regressions in econometrics," Cowles Foundation Discussion Paper No. 757, Yale University, 1985.

[35] Phillips, P.C.B. "Asymptotic expansions in nonstationary vector autoregressions", Cowles Foundation Discussion Paper No. 765, Yale University.

[36] Pollard, D. *Convergence of Stochastic Processes.* New York: Springer-Verlag, 1984.

[37] Rao, M. M. "Asymptotic distribution of an estimator of the boundary parameter of an unstable process," *Annals of Statistics*, 6 (1978), 185-190.

[38] Rozanov, Y. A. *Stationary Random Processes.* San Francisco: Holden Day, 1967.

[39] Sims, C. A. "Least squares estimation of autoregressions with some unit roots," University of Minnesota Discussion Paper No. 78-95, 1978.

[40] White, H. *Asymptotic Theory for Econometricians.* New York: Academic Press, 1984.

[41] White, H. and I. Domowitz. "Nonlinear regression with dependent observations," *Econometrica*, 52 (1984), 143-162.

[42] White, J. S. "The limiting distribution of the serial correlation coefficient in the explosive case," *Annals of Mathematical Statistics*, 29 (1958), 1188-1197.

[43] Yaglom, A. M. "*An introduction to the theory of stationary random functions*". New York: Dover, 1973.