COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 811-R

STATISTICAL INFERENCE IN REGRESSIONS WITH INTEGRATED PROCESSES: PART 1

by

Joon Y. Park & Peter C. B. Phillips

August 1987

STATISTICAL INFERENCE IN REGRESSIONS WITH INTEGRATED PROCESSES: PART 1*

by

Joon Y. Park and Peter C. B. Phillips

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

This paper develops a multivariate regression theory for integrated processes which simplifies and extends much earlier work. Our framework allows for both stochastic and certain deterministic regressors, vector autoregressions and regressors with drift. The main focus of the paper is statistical inference. The presence of nuisance parameters in the asymptotic distributions of regression F-tests is explored and new transformations are introduced to deal with these dependencies. Some specializations of our theory are considered in detail. In models with strictly exogenous regressors we demonstrate the validity of conventional asymptotic theory for appropriately constructed Wald tests. These tests provide a simple and convenient basis for specification robust inferences in this context. Single equation regression tests are also studied in detail. Here it is shown that the asymptotic distribution of the Wald test is a mixture of the chi square of conventional regression theory and the standard unit root theory. The new result accommodates both extremes and intermediate cases.

First draft: November 1986
Revised version: August 1987

---

*We are grateful to two referees for helpful comments. Our thanks also 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 8519595. This paper was completed while Joon Y. Park was an Alfred P. Sloan Doctoral Dissertation Fellow.

1. INTRODUCTION

There has recently been growing interest in the theory of regression amongst time series that are individually well explained by models of the ARIMA type. Such models generate a simple class of nonstationary time series which we generically describe as integrated processes. More specifically, we call a time series $\{X_t\}$ an integrated process of order k (in short, an $I(k)$ process) if the time series of k'th order differences $\{\Delta^k X_t\}$ is stationary (an $I(0)$ process). $I(1)$ processes behave like accumulated sums of stationary innovations and they possess a single unit root. When we run regressions with such time series the asymptotic properties of the regression coefficients, statistical tests and regression diagnostics are very different from those of regressions with stationary time series. Some of these differences have become apparent in recent work by Phillips [15, 16], Phillips and Durlauf [19], Phillips and Ouliaris [28], Stock [23], and Stock and Watson [24].

The aim of the present paper and its sequel (Part 2) is to develop an asymptotic theory of sufficient generality to accommodate most regressions of this type which are important in econometric work. To do so we use a framework of multivariate regression that permits regressors which are both deterministic and stochastic. We allow for vector autoregressions (VAR's) and for VAR's with a fitted constant term, trend and possibly additional exogenous regressors. In our general theory we allow for $I(1)$ processes with or without drift as well as cointegrated regressors.

The main focus of this paper is statistical inference in regressions with $I(1)$ processes. We develop transformations of conventional tests which, in some cases, eliminate and, in others, substantially reduce the nuisance parameter dependencies which occur in the general asymptotic theory. Certain specializations of our theory are studied in detail. In models with strictly exogenous regressors we show that the conventional asymptotic chi-square theory is valid for appropriately constructed Wald tests. The construction simply requires that the usual error variance matrix estimate be replaced by a consistent estimate of the (scaled) spectral density matrix of the errors at the zero frequency. The resulting estimate is then consistent for a wide class of stationary error processes in the regression equation and the inferences drawn from the associated Wald test are specification robust in the same sense as those of White [28] and White and Domowitz [29]. This construction then offers a substantial simplification of the White and Domowitz procedure for the case of regressions with integrated regressors.

We also examine multivariate tests for unit roots and provide a simple characterization of the asymptotic distribution theory in this case. The results generalize the unit root theory in earlier work by Dickey and Fuller [5], Phillips [16] and Phillips and Perron [22]. After our work was completed we learnt of some related research by Tsay and Tiao [26] who study multivariate tests in a similar context allowing also for complex roots.

Single equation regression tests are also studied in detail. Here we show how the parameter dependencies of the general asymptotic theory become one dimensional, reducing to a single parameter which effectively measures the time series correlation between the innovations that drive the regressors and the errors on the regression equation. In this case the limit distribution of the Wald test is simply a mixture of the chi square of conventional asymptotic theory and a multivariate version of the standard unit root theory. The chi square and the standard unit root theory then become polar cases included within this more general framework at the limits of the domain of the correlation coefficient.

The plan of the paper is as follows. The models and some preliminary theory are discussed in Section 2. Section 3 provides a development of the necessary asymptotics for least squares regression statistics. Hypothesis testing is studied in detail in Section 4 and various specializations of our theory are examined in Section 5. Some concluding remarks are made in Section 6. Proofs are given in the Appendix.

2. THE MODELS AND PRELIMINARY THEORY

Let $\{y_t\}_1^\infty$ be an n-dimensional multiple time series generated by

$$y_t = Ax_t + u_t \tag{1}$$

where A is an $n \times m$ coefficient matrix and where the m-vector process $\{x_t\}_0^\infty$ satisfies

$$x_t = x_{t-1} + v_t \ . \tag{2}$$

Our results do not depend on the initialization of (2) and we therefore allow x_0 to be any random variable including, of course, a constant. As direct extensions of (1) we shall also consider time series $\{y_t\}$ that are generated by

$$y_t = \mu + Ax_t + u_t \tag*{(1)$'$}$$

and

$$y_t = \mu + \theta t + Ax_t + u_t \ . \qquad (1)''$$

We define $w_t' = (u_t', v_t')$ and we require that the partial sum process $S_t = \Sigma_1^t w_j$ constructed from the innovation sequence $\{w_t\}_1^\infty$ satisfies a multivariate invariance principle. More specifically, if for $r \in [0,1]$ we define

$$X_T(r) = T^{-1/2} S_{[Tr]}$$

then we require

$$X_T(r) \Rightarrow B(r) \ , \quad \text{as} \quad T \to \infty \ . \qquad (3)$$

Here, T denotes the sample size, the symbol " $\Rightarrow$ " signifies weak convergence of the associated probability measures and $B(r)' = (\overset{n}{B_1}(r)', \overset{m}{B_2}(r)')$ is (n+m)-vector Brownian motion with covariance matrix

$$\Omega = \begin{array}{ll} \begin{bmatrix} \Omega_1 & \Omega_{21}' \\ \Omega_{21} & \Omega_2 \end{bmatrix} & \begin{array}{l} n \\ m \end{array} \end{array} \qquad (4)$$

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

$$= \Sigma + \Lambda + \Lambda'$$

where

$$\Sigma = \begin{array}{ll} \begin{bmatrix} \Sigma_1 & \Sigma_{21}' \\ \Sigma_{21} & \Sigma_2 \end{bmatrix} & \begin{array}{l} n \\ m \end{array} \end{array} = \lim_{T\to\infty} T^{-1} \Sigma_1^T E(w_t w_t') \qquad (5)$$

and

$$\Lambda = \begin{array}{cc} \left[\begin{array}{cc} \Lambda_1 & \Lambda_{12} \\ \Lambda_{21} & \Lambda_2 \end{array}\right] & \begin{array}{c} n \\ m \end{array} \end{array} = \lim_{T\to\infty} T^{-1}\Sigma_{t=2}^{T}\Sigma_{j=1}^{t-1}E(w_j w_t') \ . \tag{6}$$

The notation introduced in (3)-(6) will be used repeatedly throughout the paper. In particular, $B_1(r)$ and $B_2(r)$ denote, unless otherwise stated, two vector Brownian motions which are, respectively, n and m dimensional with covariance matrices Ω_1 and Ω_2, which we assume to be positive definite. It will often be convenient to write these and other stochastic processes on [0,1] without the argument. Thus, we shall frequently use B, B_1 and B_2 in place of $B(r)$, $B_1(r)$ and $B_2(r)$.

Multivariate invariance principles such as (3) have recently been proved by Eberlain [8] and Phillips and Durlauf [19]. They apply for a very wide class of innovation sequences $\{w_t\}$ that are weakly dependent and possibly heterogeneously distributed. Following Hall and Heyde [11, p. 146], they may also be shown to apply to a large class of linear processes including those generated by all stationary and invertible ARMA models.

In addition to the multivariate invariance principle (3), our asymptotic theory of regression relies upon the weak convergence of certain sample covariance matrices to matrix stochastic integrals of the form $\int_0^1 BdB'$ More specifically,

$$T^{-1}\Sigma_1^T S_{t-1}w_t' \Rightarrow \int_0^1 BdB' + \Lambda \ . \tag{3$'$}$$

This type of weak convergence has recently been shown by Phillips [17, 18] to hold for time series generated by linear processes and by Strasser [25] and Chan and Wei [4] for martingale difference sequences. For the purposes of this paper it will be convenient to assume that both (3) and (3)$'$ hold and that $\{w_t\}$ is strictly stationary and ergodic with finite fourth order moments, although we add that many of our results continue to hold for more general weakly dependent time series. Note that the time series $\{x_t\}$ is then integrated of order one in the terminology of Box and Jenkins [3]. With the assumption of stationarity (5) and (6) reduce to

$$\Sigma = E(w_1 w_1')$$

and

$$\Lambda = \Sigma_2^{\infty} E(w_1 w_j') \ .$$

Also, if the series defining Λ is absolutely convergent then $\{w_t\}$ has a continuous spectral density matrix $f_{ww}(\lambda)$ and (4) may be written simply in the form

$$\Omega = 2\pi f_{ww}(0) \ .$$

Our model (1) may be regarded as a multivariate equations system in which the regressors x_t are driven by a quite general integrated process of order one, such as a vector ARIMA model. We presume none of the common exogeneity conditions and allow for contemporaneous correlation of the form $Ex_t u_t' \neq 0$. The model for which $\{x_t\}$ is strictly exogenous will be considered as a special case later in the paper.

The model defined by (1) and (2) clearly implies that the time series $\{x_t\}$ and $\{y_t\}$ are cointegrated in the sense of Engle and Granger [9]. The nonsingularity requirement for Ω_1 and Ω_2 in (4), however, prevents $\{x_t\}$ or $\{y_t\}$ from being individually cointegrated, as shown in Phillips [15]. Our theory therefore does not apply, for example, to VAR systems with unit roots which include many lags. But it should be noted that our model does not assume invertibility of Ω and does include first order VAR systems with simple unit roots.

The following result is very useful in our subsequent theory. Here and elsewhere in the paper all limits apply as $T \to \infty$. Integrals (such as $\int_0^1 B$) are understood to be taken with respect to Lebesgue measure (that is $\int_0^1 B(r)dr$) when otherwise unspecified. These economies substantially simplify some of our subsequent formulae.

<u>LEMMA 2.1</u>

(a) $T^{-3/2}\Sigma_1^T x_t \Rightarrow \int_0^1 B_2$

(b) $T^{-5/2}\Sigma_1^T t x_t \Rightarrow \int_0^1 r B_2$

(c) $T^{-2}\Sigma_1^T x_t x_t' \Rightarrow \int_0^1 B_2 B_2'$

(d) $T^{-3/2}\Sigma_1^T t u_t \Rightarrow \int_0^1 r dB_1$

(e) $T^{-1}\Sigma_1^T x_t u_t' \Rightarrow \int_0^1 B_2 dB_1' + \Delta_{21}$,

where

$$\Delta_{21} = \Sigma_{21} + \Lambda_{21} .$$

Joint weak convergence of (a) *through* (e) *also applies.*

3. LEAST SQUARES ESTIMATION

We shall consider three multiple least squares regressions which correspond, respectively, to (1), (1)′ and (1)″:

$$y_t = \hat{A}x_t + \hat{u}_t \; ; \tag{7}$$

$$y_t = \bar{\mu} + \bar{A}x_t + \bar{u}_t \; ; \tag{8}$$

$$y_t = \tilde{\mu} + \tilde{\theta}t + \tilde{A}x_t + \tilde{u}_t \; . \tag{9}$$

The inclusion of the constant term and the time trend in (8) and (9), of course, are also equivalent to demeaning and detrending the series $\{x_t\}$ and $\{y_t\}$ prior to regression (7) in their effect on estimation of the coefficient matrix A . We should also note that the coefficient matrix $\tilde{A}$ in (9) (and also $\tilde{\mu}$) is invariant with respect to the introduction of a nonzero drift in the process generating the time series $\{x_t\}$, i.e., they do not depend on π in

$$x_t = \pi + x_{t-1} + v_t \tag{2'}$$

when (2) is replaced by (2)′. This is especially important in our multiple regression context since under (2)′

$$T^{-3}\Sigma_1^T x_t x_t' \xrightarrow{p} (1/3)\pi\pi' \tag{10}$$

which is singular if $m > 1$. This complicates the limiting distribution theory of $\hat{A}$ and will be discussed in detail below.

Let

$$\bar{A}_1 = (\bar{\mu},\ \bar{A})\ ,\quad \tilde{A}_2 = (\tilde{\mu},\ \tilde{\theta},\ \tilde{A}) \tag{11}$$

in (8) and (9) with analogous definitions of A_1 and A_2 . Define $x_t^{1\prime} = (1,\ x_t')$, $x_t^{2\prime} = (1,\ t,\ x_t')$ and, given a sample of size T , define

$$X' = (x_1,\ \ldots,\ x_T)\ ,\quad X_1' = (x_1^1,\ \ldots,\ x_T^1)\ ,\quad X_2' = (x_1^2,\ \ldots,\ x_T^2)\ . \tag{12}$$

We have:

$$\hat{A} = Y'X(X'X)^{-1}\ ,\quad \bar{A}_1 = Y'X_1(X_1'X_1)^{-1}\ ,\quad \tilde{A}_2 = Y'X_2(X_2X_2)^{-1}\ .$$

The least squares estimator of the covariance matrix is given for each regression equation by

$$\hat{\Sigma}_1 = \frac{1}{T}Y'(X - P_X)Y\ ,\quad \bar{\Sigma}_1 = \frac{1}{T}Y'(I - P_{X_1})Y\ ,\quad \tilde{\Sigma}_2 = \frac{1}{T}Y'(I - P_{X_2})Y\ ,$$

where $P_D = D(D'D)^{-1}D'$ for any matrix D of full column rank (with probability one, if it is random).

The following theorems characterize the asymptotic behavior of the least squares estimators in regression equations (7)-(9). To represent the asymptotics concisely, we first define a functional

$$f(B,\ M,\ E) = \left(\int_0^1 dBM' + E'\right)\left(\int_0^1 MM'\right)^{-1} \tag{13}$$

where B is a vector Brownian motion, M is a stochastic process of continuous sample paths such that $\int_0^1 MM' > 0$ a.s., and E is a matrix (which may be random) of conformable dimension. Especially, we will consider functionals of the form $f(B_1,\ M(B_2),\ E)$, where M is a function of B_2 .

Such functionals are important since the limiting distributions of all of the least squares estimators in (7)-(9) can be represented in this form.

THEOREM 3.1

$$T(\hat{A}-A) \Rightarrow f(B_1, B_2, \Delta_{21}) \ .$$

THEOREM 3.2

(a) $T(\bar{A}-A) \Rightarrow f(B_1, B_2^*, \Delta_{21})$

(b) $\sqrt{T}(\bar{\mu}-\mu) \Rightarrow f(B_1, P_1, \delta_1)$

where

$$B_2^*(r) = B_2(r) - \int_0^1 B_2 \tag{14}$$

$$P_1(r) = 1 - \left[\int_0^1 B_2'\left(\int_0^1 B_2 B_2'\right)^{-1}\right] B_2(r) \tag{15}$$

and

$$\delta_1 = -\left[\int_0^1 B_2'\left(\int_0^1 B_2 B_2'\right)^{-1}\right]\Delta_{21} \ .$$

THEOREM 3.3

(a) $T(\tilde{A}-A) \Rightarrow f(B_1, B_2^{**}, \Delta_{21})$

(b) $\sqrt{T}(\tilde{\mu}-\mu) \Rightarrow f(B_1, P_2, \delta_2)$

(c) $T^{3/2}(\tilde{\theta}-\theta) \Rightarrow f(B_1, P_3, \delta_3)$

where

$$B_2^{**}(r) = B_2(r) - 4\left(\int_0^1 B_2 - (3/2)\int_0^1 sB_2\right) + 6r\left(\int_0^1 B_2 - 2\int_0^1 sB_2\right) \tag{16}$$

$$P_2(r) = 1 - \tfrac{3}{2}r - \left[\left(\int_0^1 B_2' - \tfrac{3}{2}\int_0^1 sB_2'\right)\left(\int_0^1 B_2B_2' - 3\int_0^1 sB_2\int_0^1 sB_2'\right)^{-1}\right] \tag{17}$$

$$\cdot\left[B_2(r) - 3r\int_0^1 sB_2\right]$$

$$P_3(r) = r - \tfrac{1}{2} - \left[\left(\int_0^1 sB_2' - \tfrac{1}{2}\int_0^1 B_2'\right)\left(\int_0^1 B_2B_2' - \int_0^1 B_2\int_0^1 B_2'\right)^{-1}\right] \tag{18}$$

$$\cdot(B_2(r) - \int_0^1 B_2)$$

and

$$\delta_2 = -\left[\left(\int_0^1 B_2' - \tfrac{3}{2}\int_0^1 sB_2'\right)\left(\int_0^1 B_2B_2' - 3\int_0^1 sB_2\int_0^1 sB_2'\right)^{-1}\right]\Delta_{21}$$

$$\delta_3 = -\left[\left(\int_0^1 sB_2' - \tfrac{1}{2}\int_0^1 B_2'\right)\left(\int_0^1 B_2B_2' - \int_0^1 B_2\int_0^1 B_2'\right)^{-1}\right]\Delta_{21}$$

The results in Theorems 3.1 and 3.2 were earlier obtained by Phillips and Durlauf [19], though their expressions for the limiting distributions are slightly different from those given here. The advantage of the present formulation is that it demonstrates how all of these cases including those of Theorem 3.3 fall simply within the same overall framework.

Theorems 3.1-3.3 give explicit asymptotic results for the least squares estimators in (7)-(9). The limiting distributions are expressed in terms of a simple but rather general functional that is easy to analyze and extremely useful for our study. In fact, many of the results in this paper can be deduced quite easily from this representation, which will become more apparent as we proceed.

Inclusion of a constant term or a time trend in regressions such as (1) does affect the limiting distribution of the least squares estimator of A . Similarly, demeaning or detrending prior to regression also changes the asymptotic behavior of the estimator. This has been widely recognized in the literature, especially in autoregressive models with unit roots. It is interesting to observe from our own theory that the resulting change in the asymptotics involves simply the replacement of B_2 with B_2^* or B_2^{**} for M in the functional f in (13), where B_2^* and B_2^{**} are "demeaned" and "detrended" Brownian motions. Explicitly, B_2^* and B_2^{**} are simply the residuals from the continuous time regressions

$$B_2(r) = \hat{\alpha}_0 + B_2^*(r)$$

$$B_2(r) = \tilde{\alpha}_0 + \tilde{\alpha}_1 r + B_2^{**}(r)$$

where $\hat{\alpha}_0$ and $(\tilde{\alpha}_0, \tilde{\alpha}_1)$ minimize the (continuous time) least squares criteria

$$\int_0^1 \| B_2(r) - \alpha \|^2 dr \ ,$$

$$\int_0^1 \| B_2(r) - \alpha_0 - \alpha_1 r \|^2 dr \ ,$$

respectively. Thus, the effect of demeaning or detrending in regressions with integrated processes carries through in exactly the same fashion to the asymptotics.

All the least squares estimators of the regression equations (7)-(9) are consistent. There is no bias resulting from the correlation between regressors and regression errors, such as the simultaneous equations and measurement error bias. See Phillips and Durlauf [19] for a detailed discussion of this point. Especially, the least squares estimators of A, the parameter matrix of main interest, converges at the rate of $O_p(T^{-1})$ in all three regressions. Therefore, the demeaning or detrending implied by (8) or (9) does not affect consistency. This implies, of course, that Andrews' [1] consistency result for $\hat{A}$ in (7) with more aberrant regression errors is extended and also holds for $\bar{A}$ and $\tilde{A}$.

THEOREM 3.4

(a) $\hat{\Sigma}_1, \bar{\Sigma}_1, \tilde{\Sigma}_1 \xrightarrow{p} \Sigma_1$

Moreover, if we let $\xi_t = u_t \otimes u_t - E(u_t \otimes u_t)$ *and assume* $\{\xi_t\}_1^\infty$ *is a weakly stationary process for which the invariance principle* (3) *holds, then*

(b) $\sqrt{T}(\hat{\Sigma}_1 - \Sigma_1),\ \sqrt{T}(\bar{\Sigma}_1 - \Sigma_1),\ \sqrt{T}(\tilde{\Sigma}_1,\ \Sigma_1) \Rightarrow N(0,V)$

where

$$V = P_D(\Sigma_{k=0}^{\infty}\Psi_k)P_D$$

$$\Psi_0 = E(u_t u_t' \otimes u_t u_t') - \text{vec}(\Sigma_1)\text{vec}(\Sigma_1)'$$

$$\Psi_k = \Phi_k + \Phi_k' \ ; \quad (k = 1,\ 2,\ \ldots\ )$$

$$\Phi_k = E(u_t u_{t+k}' \otimes u_t u_{t+k}') - \text{vec}(\Sigma_1)\text{vec}(\Sigma_1)'$$

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

Theorem 3.4 generalizes the results in Phillips and Durlauf [19], Theorem 4.2 and Theorem 6.1(g) to regressions with a time trend. Consistency of $\hat{\Sigma}_1$, $\bar{\Sigma}_1$ and $\tilde{\Sigma}_1$, each of which is obtained from the corresponding regression equation in (7)-(9), is a natural consequence of that of the least squares estimators given in Theorems 3.1-3.3. The limiting distributions of these estimators of Σ_1 are normal with an identical covariance matrix that depends on the fourth order cumulant sequence of the innovation sequence $\{u_t\}$. We will show in the proof that the effect of demeaning or detrending on the estimation of error covariance matrix is, in fact, at most of order $O_p(T^{-1})$ asymptotically.

We now consider the case where some of the regressors have nonzero drifts, i.e., that $\{x_t\}$ is driven instead by (2)′ with $\pi \neq 0$. In the VAR system where $x_t = y_{t-1}$, this is obviously equivalent to the introduction of a nonzero intercept in the regression equation. It is useful in this case to write (2)′ as

$$x_t = \pi t + x_t^0$$

where $x_t^0 = \Sigma_1^t v_j + x_0$ is a driftless random walk with initial condition x_0. Now the process $\{x_t\}$ is driven by two components, one involving a deterministic linear trend and the other a stochastic trend. The deterministic component of $\{x_t\}$ apparently dominates asymptotic behavior since the stochastic trend is of order $O_p(t^{1/2})$. Thus $\{x_t\}$ behaves asymptotically as if it were πt and this gives the result stated earlier in (10). The limiting singularity of the sample moment matrix of x_t when $m > 1$ is due to the fact that the leading time trend term fails to introduce sufficient asymptotic variation across the component variables. The situation here is closely related to the concept of cointegration. Note, in particular, that any vector α that is orthogonal to π annihilates the deterministic trend in x_t leaving $\alpha' x_t = \alpha' x_t^0$, which behaves like a random walk without drift. Thus, α takes on the role of a cointegrating vector reducing by an order of magnitude, viz. $O_p(t^{1/2})$, the variability of x_t.

The effects of this apparent degeneracy are also similar to those that occur in the case of cointegration amongst processes which are integrated of order one.

To develop a complete asymptotic theory in this case we first transform the regression equations (7) and (8) as

$$y_t = \hat{c}_1 z_{1t} + \hat{C}_2 z_{2t} + \hat{u}_t \tag{7'}$$

$$y_t = \bar{\mu} + \bar{c}_1 z_{1t} + \bar{C}_2 z_{2t} + \bar{u}_t \tag{8'}$$

where $z_{1t} = h_1' x_t$ for $h_1 = (\pi'\pi)^{-1/2}\pi$, $z_{2t} = H_2' x_t$ and $H = (h_1, H_2)$ is an orthogonal matrix of order m. Upon transformation, the deterministic trend of $\{x_t\}$ is now concentrated in $\{z_{1t}\}$, while $\{z_{2t}\}$ embodies the stochastic trend and is a driftless random walk of dimension $(m-1)$.

The asymptotic results for the least squares estimators of the parameters in (7)′ and (8)′ can be obtained analogously with Theorems 3.1-3.3. The next two theorems characterize the limiting distributions of these estimators and the corresponding estimators of A in terms of the functional f in (13) and $\underline{B} = (B_1, \underline{B}_2)$, an $n+(m-1)$-vector Brownian motion with covariance matrix

$$\begin{array}{cc} \begin{bmatrix} \Omega_1 & \Omega_{21}' H_2 \\ H_2' \Omega_{21} & H_2' \Omega_2 H_2 \end{bmatrix} & \begin{array}{l} n \\ m-1 \end{array} \end{array} .$$

THEOREM 3.5

(a) $T^{3/2}(\pi'\pi)^{1/2}(\hat{c}_1 - c_1) \Rightarrow f(B_1, \underline{P}_4, \underline{\delta}_4)$;

(b) $T(\hat{C}_2 - C_2) \Rightarrow f(B_1, \underline{B}_2^+, \underline{\Delta}_{21})$;

where $\underline{\Delta}_{21} = H_2'\Delta_{21}$,

$$\underline{B}_2^+(r) = \underline{B}_2(r) - 3r\int_0^1 s\underline{B}_2 ,$$

$$\underline{P}_4(r) = r - \left[\int_0^1 s\underline{B}_2'\left(\int_0^1 \underline{B}_2\underline{B}_2'\right)^{-1}\right]\underline{B}_2(r) ,$$

$$\underline{\delta}_4 = -\left[\int_0^1 s\underline{B}_2'\left(\int_0^1 \underline{B}_2\underline{B}_2'\right)^{-1}\right]\underline{\Delta}_{21} ;$$

(c) $T(\hat{A}-A) \Rightarrow f(B_1, \underline{B}_2^+, \underline{\Delta}_{21})H_2'$.

THEOREM 3.6

(a) $T^{1/2}(\bar{\mu}-\mu) \Rightarrow f(B_1, \underline{P}_2, \underline{\delta}_2)$;

(b) $T^{3/2}(\pi'\pi)^{1/2}(\bar{c}_1 - c_1) \Rightarrow f(B_1, \underline{P}_3, \underline{\delta}_3)$;

(c) $T(\bar{C}_2 - C_2) \Rightarrow f(B_1, \underline{B}_2^{**}, \underline{\Delta}_{21})$;

where $\underline{\Delta}_{21} = H_2'\Delta_2$, *and* $\underline{B}_2^{**}$, $\underline{P}_2$, $\underline{P}_3$, $\underline{\delta}_2$ *and* $\underline{\delta}_3$ *are defined from* $\underline{B}_2$ *and* $\underline{\Delta}_{21}$ just *as their counterparts in Theorem 3.3.*

(d) $T(\bar{A}-A) \Rightarrow f(B_1, \underline{B}_2^{**}, \underline{\Delta}_{21})H_2'$.

Theorems 3.5 and 3.6 provide explicit asymptotic results for regressions (7)', (8)' and, hence, (7) and (8). The limiting distributions are not normal if $m > 1$, and are somewhat similar to the previous case of driftless regressors with the extra time trend in the regression equation (see Theorem 3.3).

Interestingly, both estimators $\hat{A}$ and $\bar{A}$ are $O_p(T^{-1})$ consistent even when $\{x_t\}$ has a nonzero drift and its sample moment matrix becomes singular asymptotically. This parallels our previous consistency result when $\{x_t\}$ is driftless. Thus, as far as consistency is concerned, the presence of a drift in regressors is innocuous. All linear combinations of $\hat{A}$ and $\bar{A}$ are $O_p(T^{-1})$ consistent other than $\hat{A}\pi$ and $\bar{A}\pi$ which are $O_p(T^{-3/2})$ consistent.

More importantly, we observe that the limiting distributions of $T(\hat{A}-A)$ and $T(\bar{A}-A)$ are singular and non normal. The singularity of the limit distribution when $m > 1$ arises because of the singularity of the limit of the sample moment matrix (10). In addition to (10) we also have:

$$T^2(\Sigma_1^T x_t x_t')^{-1} \Rightarrow H_2(\int_0^1 \underline{B}_2\underline{B}_2')^{-1}H_2'$$

whose rank is $m-1$. Since

$$T(\hat{A}-A) = (T^{-1}\Sigma_1^T u_t x_t')(T^{-2}\Sigma_1^T x_t x_t')^{-1}$$

this explains why the support of the limit distribution of $T(\hat{A}-A)$ is the range of $I_n \otimes H_2$, a proper subspace of R^{nm} . Similar considerations apply in the case of $T(\bar{A}-A)$. Results (c) and (d) of Theorems 3.5 and 3.6 give the asymptotic theory for all linear combinations of the matrices $T(\hat{A}-A)$ and $T(\bar{A}-A)$. Note that only $T(\hat{A}-A)\pi$ and $T(\bar{A}-A)\pi$ are degenerate in the limit and the asymptotic theory for these vectors is given in part (b) of Theorems 3.5 and 3.6 upon restandardization.

These results should help to clarify the effects on the asymptotic theory of the presence of a drift in the regressor process. As stated earlier when $m > 1$ the limiting distributions of the coefficient estimates are all non normal. When $m = 1$, as is well known, both of $T^{3/2}(\hat{A}-A)$ and $T^{3/2}(\bar{A}-A)$ are asymptotically normal. The limiting distributions are easily obtained from Theorem 3.5(a) and Theorem 3.6(b) with the convention $\underline{B}_2 = 0$. We have $\underline{P}_4(r) = r$, $\underline{\delta}_4 = 0$, and

$$T^{3/2}(\hat{A}-A) \Rightarrow \left[(1/\pi)\int_0^1 rB_1\right]\left[\int_0^1 r^2\right]^{-1}$$
$$= N\left[0, \frac{3}{\pi^2}\Omega_1\right] .$$

Similarly, $\underline{P}_3(r) = r - 1/2$, $\underline{\delta}_3 = 0$ and

$$T^{3/2}(\bar{A}-A) \Rightarrow N\left[0, \frac{12}{\pi^2}\Omega_1\right] .$$

Notice that Ω_1, not Σ_1, appears in the covariance matrices since we allow the u_t's to be serially correlated. These results provide a substantial generalization of recent work by West [27]. West studied the case of a single nonstationary regressor $(m = 1)$ with nonzero drift and argued that conventional normal asymptotics applied. Our theory demonstrates how specialized his results really are.

The estimators $\tilde{\mu}$ and $\tilde{A}$ are invariant with respect to π in (2)', and the asymptotic results for these estimators given in Theorem 3.3 remain valid with the introduction of a nonzero drift in the time series $\{x_t\}$. The estimator $\tilde{\theta}$, however, does depend on π. In fact, it is easy to show that

$$(\tilde{\theta}-\theta) = (\tilde{A}-A)\pi + o_p(T^{-3/2})$$

and the limiting distribution of $T(\tilde{\theta}-\theta)$ can be readily reduced from Theorem 3.3(a). Once again, the limiting distributions are non normal.

4. HYPOTHESIS TESTING

We shall consider linear hypotheses that involve coefficients in the multiple regression equations (7)-(9) when the data are generated by (1), (1)', (1)" and (2). Our approach follows the constructions in Phillips [16] but is complicated by the presence of nuisance parameter matrices, rather than scalars as in [16]. These matrices arise naturally in the general multivariate setting studied here. Later, in Section 5, we show how earlier results follow as simple specializations of our formulae.

The null hypothesis for the coefficient matrix A, which is of primary interest, is written explicitly as

$$R \text{ vec } A = r \tag{19}$$

where R and r are known $q \times nm$, $q \times 1$ matrices, respectively, and R is of full rank q. Other hypotheses of interest are

$$\mu = \mu_0 \tag{20}$$

and

$$\theta = \theta_0 . \tag{21}$$

We commonly employ the Wald test for the hypotheses (19)-(21). Let

$$(X'X)^{-1} = M_0 \tag{22}$$

$$(X_1'X_1)^{-1} = M_1 = \begin{array}{c} \begin{bmatrix} m_{11} & \\ & M_{12} \end{bmatrix} \end{array} \begin{array}{l} 1 \\ m \end{array} \tag{23}$$

$$(X_2'X_2)^{-1} = M_2 = \begin{array}{c} \begin{bmatrix} m_{21} & \\ & M_{22} \end{bmatrix} \end{array} \begin{array}{l} 1 \\ m+1 \end{array} = \begin{array}{c} \begin{bmatrix} m_{21} & & \\ & m_{22} & \\ & & M_{23} \end{bmatrix} \end{array} \begin{array}{l} 1 \\ 1 \\ m \end{array} \tag{24}$$

where X_1 and X_2 are given by (12). We will not need to be specific about off block diagonal entries in (23) and (24). In (7) the Wald statistic to test (19) is

$$F(\hat{A}) = (R \text{ vec } \hat{A} - r)'\left[R(\hat{\Sigma}_1 \otimes M_0)R'\right]^{-1}(R \text{ vec } \hat{A} - r) \ . \tag{25}$$

To test (19) and (20) in regression equation (8), the corresponding statistics are given by

$$F(\bar{A}) = (R \text{ vec } \bar{A} - r)'\left[R(\bar{\Sigma}_1 \otimes M_{12})R'\right]^{-1}(R \text{ vec } \bar{A} - r) \tag{26}$$

and

$$F(\bar{\mu}) = m_{11}^{-1}(\bar{\mu} - \mu_0)'\bar{\Sigma}_1^{-1}(\bar{\mu} - \mu_0) \ . \tag{27}$$

Similarly, the tests of (19)-(21) in (9) can be based on

$$F(\tilde{A}) = (R \text{ vec } \tilde{A} - r)'\left[R(\tilde{\Sigma}_1 \otimes M_{23})R'\right]^{-1}(R \text{ vec } \tilde{A} - r) \tag{28}$$

$$F(\tilde{\mu}) = m_{21}^{-1}(\tilde{\mu} - \mu_0)'\tilde{\Sigma}_1^{-1}(\tilde{\mu} - \mu_0) \tag{29}$$

$$F(\tilde{\theta}) = m_{22}^{-1}(\tilde{\theta} - \theta_0)'\tilde{\Sigma}_1^{-1}(\tilde{\theta} - \theta_0) \ . \tag{30}$$

Moreover, we may want to perform joint tests and we accordingly define

$$\bar{F} = (R_1 \text{ vec } \bar{A}_1 - r_1)'\left[R_1(\bar{\Sigma}_1 \otimes M_1)R_1'\right]^{-1}(R_1 \text{ vec } \bar{A}_1 - r_1) \tag{31}$$

$$\tilde{F} = (R_2 \text{ vec } \tilde{A}_2 - r_2)'\left[R_2(\tilde{\Sigma}_1 \otimes M_2)R_2'\right]^{-1}(R_2 \text{ vec } \tilde{A}_2 - r_2) \tag{32}$$

where

$$R_1 = \begin{bmatrix} I_n & \\ & RK_{nm} \end{bmatrix} K_{(m+1)n} \,, \quad r_1 = \begin{bmatrix} \mu_0 \\ r \end{bmatrix}$$

$$R_2 = \begin{bmatrix} I_{2n} & \\ & RK_{nm} \end{bmatrix} K_{(m+2)n} \,, \quad r_2 = \begin{bmatrix} \mu_0 \\ \theta_0 \\ r \end{bmatrix}$$

Here, K_{mn} denotes the commutation matrix of order $mn \times mn$, and A_1 and A_2 are defined in (11). Finally, it may be of some interest to consider the joint test of (19) and (21) in (9). The Wald statistic for this test is given as

$$\tilde{F}_1 = (R_1 \text{ vec } \tilde{A}_3 - r_3)'\left[R_1(\tilde{\Sigma}_1 \otimes M_{22})R_1'\right]^{-1}(R_1 \text{ vec } \tilde{A}_3 - r_3) \tag{33}$$

where $\tilde{A}_3 = (\tilde{\theta}, \tilde{A})$ and $r_3' = (\theta_0', r')$.

Although the "F"-statistics studied above are constructed in the standard fashion, their construction is not the most appropriate in our time series set up. We allow for serially correlated innovations $\{u_t\}$ and the use of estimators of Σ_1 in the tests only accounts for contributions from the variance of u_t and does not generally allow for its serial covariance properties. One way of reducing unnecessary dependencies on nuisance parameters in the limit distributions of the tests is to properly allow for this serial correlation.

We are thus led to modify the Wald statistics by replacing consistent estimators of Σ_1 with consistent estimators of Ω_1. Such variance estimators seem much more natural in our set up where weakly dependent innovations play an important role in the generality of the theory. Consistent estimation of these nuisance parameters will be discussed later in this section. We denote by

$$G(\hat{A}),\ G(\bar{A}),\ G(\bar{\mu}),\ \bar{G},\ G(\tilde{A}),\ G(\tilde{\mu}),\ G(\tilde{\theta}),\ \tilde{G},\ \tilde{G}_1 \tag{34}$$

the new statistics, each of which is obtained from the corresponding "F"-statistic in (25)-(33) by the replacement of $\hat{\Sigma}_1$, $\bar{\Sigma}_1$ or $\tilde{\Sigma}_1$ with $\tilde{\Omega}_1$, a consistent estimator of Ω_1.

Our subsequent theory centers on these G-statistics rather than the standard Wald statistics. The corresponding theory for the "F"-statistics is easily obtained in the same way but will be omitted since these statistics are less useful. However, if $\{u_t\}$ is a white noise, or a martingale difference sequence, the two statistics are asymptotically equivalent, and in this sense G-statistics are more robust in their construction to assumptions about the innovation sequence. Also observe that if $n = 1$ then the statistics are simply proportional and $G = (\sigma_1^2/\omega_1^2)F$ asymptotically.

Before presenting our main result on the asymptotic behavior of the G-statistics we introduce a functional g_R defined by

$$g_R(B,M,E) = \left(\int_0^1 dB \otimes M + e\right)'\left[I \otimes \left(\int_0^1 MM'\right)^{-1}\right]R' \tag{35}$$

$$\cdot\left\{R\left[\Omega \otimes \left(\int_0^1 MM'\right)^{-1}\right]R'\right\}^{-1} R\left[I \otimes \left(\int_0^1 MM'\right)^{-1}\right]\left(\int_0^1 dB \otimes M + e\right)$$

$$= \mathrm{vec}(f(B,M,E))'R'\left\{R\left[\Omega \otimes \left(\int_0^1 MM'\right)^{-1}\right]R'\right\}^{-1} R\ \mathrm{vec}(f(B,M,E))$$

where B , M and E are as specified for the functional f in (13), $e = \mathrm{vec}\ E'$, Ω is the covariance matrix of B , and R is a constant matrix of an appropriate dimension so that (35) is well defined. We also set $g(B,M,E) = g_I(B,M,E)$.

We now have

THEOREM 4.1

(a) $G(\hat{A}) \Rightarrow g_R(B_1,\ B_2,\ \Delta_{21})$

(b) $G(\bar{A}) \Rightarrow g_R(B_1,\ B_2^*,\ \Delta_{21})$, $G(\bar{\mu}) \Rightarrow g(B_1,\ P_1,\ \delta_1)$,

$\bar{G} \Rightarrow g_R(B_1,\ B_2,\ \Delta_{21}) + g(B_1,\ P_1,\ \delta_1)$

(c) $G(\tilde{A}) \Rightarrow g_R(B_1,\ B_2^{**},\ \Delta_{21})$, $G(\tilde{\mu}) \Rightarrow g(B_1,\ P_2,\ \delta_2)$,

$G(\tilde{\theta}) \Rightarrow g(B_1,\ P_3,\ \delta_3)$,

$\tilde{G} \Rightarrow g_R(B_1,\ B_2,\ \Delta_{21}) + g(B_1,\ P_1,\ \delta_1) + g(B_1,\ P_3,\ \delta_3)$

$\tilde{G}_1 \Rightarrow g_R(B_1,\ B_2^*,\ \Delta_{21}) + g(B_1,\ P_3,\ \delta_3)$

where notation for the arguments in these functionals is defined in (14)-(18).

The limiting distributions of the "F" and G-statistics depend, in general, on the nuisance parameters in quite a complicated manner. This parameter dependency is, in some sense, intrinsic to the theory of regression with integrated processes and is "the rule rather than the exception" similar to finite sample distribution theory, as noted earlier by Phillips and Durlauf [19].

More specifically, Theorem 4.1 shows that the limiting distributions of our test statistics depend not only on the covariance structure of the Brownian motion B but also on the nuisance parameter matrix Δ_{21}. Between these two, the latter is the more problematic from the practical point of view and makes standard testing procedures virtually inapplicable in a regression with integrated processes as considered in this paper. In some special but most interesting cases that we shall explore in Section 5 the dependency of the asymptotics on the covariance matrix of the limiting Brownian motion either disappears or reduces to a parsimonious and manageable form.

In general, however, it is necessary to construct new statistics whose limiting distributions are not dependent on Δ_{21}. Accordingly we now consider some transformations of the G-statistics for testing simple null hypotheses of the form (19) with $R = I$. To do so, we first need consistent estimates of Ω_1 and Δ_{21}. Consistent estimation of Ω_1 is discussed in Phillips and Durlauf [19]. In what follows we shall use $\tilde{\Omega}_1$ to denote a member of the class of consistent positive semi definite estimates of Ω_1 constructed, of course, from the least squares residuals from regressions (7)-(9). Consistent estimates of Λ_{21} can be obtained in a similar way and we shall simply use by $\tilde{\Lambda}_{21}$ to denote a consistent estimate of Λ_{21}.

We now define:

$$H(\hat{A}) = G(\hat{A}) - 2T \operatorname{tr} \tilde{\Omega}_1^{-1}(\hat{A}-A)\tilde{\Delta}_{21} + T^2 \operatorname{tr} \tilde{\Omega}_1^{-1}\tilde{\Delta}_{21}'\left(\Sigma_1^T x_t x_t'\right)^{-1}\tilde{\Delta}_{21} \qquad (36)$$

and for (8) with the constant term

$$H(\bar{A}) = G(\bar{A}) - 2T \text{ tr } \bar{\Omega}_1^{-1}(\bar{A}-A)\bar{\Delta}_{21} + T^2 \text{tr } \bar{\Omega}_1^{-1}\bar{\Delta}_{21}'\left(\Sigma_1^T x_t^* x_t^{*\prime}\right)^{-1}\bar{\Delta}_{21} \qquad (37)$$

$$= G(\bar{A}) + a \ , \quad \text{say,}$$

$$H(\bar{\mu}) = G(\bar{\mu}) - 2\sqrt{T}\ \bar{\delta}_1\bar{\Omega}_1^{-1}(\bar{\mu}-\mu) + T\bar{\delta}_1\bar{\Omega}_1^{-1}\bar{\delta}_1' RSS_1^{-1} \qquad (38)$$

$$\bar{H} = \bar{G} + a \ . \qquad (39)$$

Similarly, for the regression (9) with the time trend, we define:

$$H(\tilde{A}) = G(\tilde{A}) - 2T \text{ tr } \tilde{\Omega}_1^{-1}(\tilde{A}-A)\tilde{\Delta}_{21} + T^2 \text{tr } \tilde{\Omega}_1^{-1}\tilde{\Delta}_{21}'\left(\Sigma_1 x_t^{**} x_t^{**\prime}\right)^{-1}\tilde{\Delta}_{21} \qquad (40)$$

$$= G(\tilde{A}) + b \ , \quad \text{say}$$

$$H(\tilde{\mu}) = G(\tilde{\mu}) - 2\sqrt{T}\ \tilde{\delta}_2\tilde{\Omega}_1^{-1}(\tilde{\mu}-\mu) + T\tilde{\delta}_2\tilde{\Omega}_1^{-1}\tilde{\delta}_2' RSS_2^{-1} \qquad (41)$$

$$H(\tilde{\theta}) = G(\tilde{\theta}) - 2T^{3/2}\tilde{\delta}_3\tilde{\Omega}_1^{-1}(\tilde{\theta}-\theta) + T^3\tilde{\delta}_3\tilde{\Omega}_1^{-1}\tilde{\delta}_3 RSS_3^{-1} \qquad (42)$$

$$\tilde{H} = \tilde{G} + b \qquad (43)$$

$$\tilde{H}_1 = \tilde{G}_1 + b \ . \qquad (44)$$

Here x_t^* and x_t^{**} are, respectively, deviations of x_t from the sample mean and the fitted time trend analogous to the demeaned and detrended Brownian motions B^* and B^{**} . Also, for computational convenience we define the least squares regression equations

(i) $1 = x_t'\hat{\beta}_1 + e_t$

(ii) $1 = \hat{\theta}t + x_t'\hat{\beta}_2 + e_t$

(iii) $t = \hat{\mu} + x_t'\hat{\beta}_3 + e_t$

for a sample of size T . we let RSS_j be the residual sum of squares from the j'th regression $(j = 1, 2, 3)$ and let $\tilde{\delta}_j = -T^{1/2}\hat{\beta}_j'\tilde{\Delta}_{21}$ $(j = 1, 2)$

and $\tilde{\delta}_3 = -T^{-1/2}\hat{\beta}_3'\tilde{\Delta}_{21}$. This notation is used in (37)-(42) and will be used frequently in what follows.

The limiting distributions of the H-statistics, by construction, do not depend on Δ_{21} (and hence, of course, on the δ_i's). They can be conveniently represented in terms of a functional h defined from g in (35) by

$$h(B,M) = g(B,M,0) - g_I(B,M,0) \ . \tag{45}$$

We thus have

<u>THEOREM 4.2</u>

(a) $H(\hat{A}) \Rightarrow h(B_1,\ B_2)$

(b) $H(\bar{A}) \Rightarrow h(B_1,\ B_2^*)$, $H(\bar{\mu}) \Rightarrow h(B_1,\ P_1)$

$\bar{H} \Rightarrow h(B_1,\ 1) + h(B_1,\ B_2^*)$

(c) $H(\tilde{A}) \Rightarrow h(B_1,\ B_2^{**})$, $H(\tilde{\mu}) \Rightarrow h(B_1,\ P_2)$, $H(\tilde{\theta}) \Rightarrow h(B_1,\ P_3)$,

$\tilde{H} \Rightarrow h(B_1,\ 1) + h(B_1,\ r^*) + h(B_1,\ B_2^{**})$

$\tilde{H}_1 \Rightarrow h(B_1,\ r^*) + h(B_1,\ B_2^{**})$

where $r^* = r - 1/2$ *and other notation is defined in* (14)-(18). *Moreover, the results in* (c), *except for* $H(\tilde{\theta})$ *remain valid if* $\{x_t\}$ *is generated by* (2)'.

If $\Delta_{21} = 0$, then the untransformed G-statistics have the same limiting distributions as the corresponding transformed H-statistics. This would clearly occur when the regressors are strictly exogenous and their driving process is generated independently of the regression errors. It may be worth noting that strict exogeneity, however, is not required. We also have asymptotic equivalence of the H- and G-statistics when the regressors are lagged integrated variables whose driving process is only contemporaneously correlated with the regression errors. The VAR system driven by a white noise innovation is one such model.

5. SPECIALIZATIONS

5.1. Multiple Regressions with Strictly Exogenous Regressors

We now consider the regression models (7)-(9) when the regressors are strictly exogenous; that is, when $\{x_t\}$ is driven by a process $\{v_t\}$ which is generated independently from the regression error process $\{u_t\}$. If $\{x_t\}$ is strictly exogenous, then the nuisance parameter Δ_{21} vanishes (and, hence, so do all δ_i's) and B_1 becomes independent of B_2 (and also, of course, B_2^*, B_2^{**} and P_1 through P_3) since $\Omega_{21} = 0$. Notice that in our formulation this never occurs in VAR systems, where $\Omega_{21} = \Omega_1$.

Define a functional k by

$$k(B,M) = \int_0^1 dBM' \left(\int_0^1 MM'\right)^{-1/2} \tag{46}$$

where B and M are specified as for the functional f in (13) and observe:

<u>LEMMA 5.1</u>. *If* B *is a Brownian motion with covariance matrix* Ω *and* M *is a* p*-vector process independent of* B, *then*

$$k(B,M) \equiv N(0, \Omega \otimes I_p) .$$

The following theorem, which characterizes the asymptotic behavior of the least squares estimators when the regressors are strictly exogenous, is an immediate consequence of the above lemma.

THEOREM 5.2 *If* $\Omega_{21} = \Delta_{21} = 0$ *in* (4) *and* (6), *then*

(a) $(\hat{A}-A)M_0^{-1/2}$, $(\bar{A}-A)M_{12}^{-1/2}$, $(\tilde{A}-A)M_{23}^{-1/2} \Rightarrow N(0, \Omega_1 \otimes I_m)$

(b) $(\bar{\mu}-\mu)m_{11}^{-1/2}$, $(\tilde{\mu}-\mu)m_{21}^{-1/2}$, $(\tilde{\theta}-\theta)m_{22}^{-1/2} \Rightarrow N(0, \Omega_1)$

where notation is defined in (22)-(24).

Theorem 5.2 shows, in fact, that upon appropriate standardization the conventional asymptotic theory applies to regressions with integrated regressors which are strictly exogenous. The asymptotic normality given in the above theorem has earlier been discovered by Krämer [12] for a special case where $\{x_t\}$ is a scalar ARIMA (p,1,q) process and $\{u_t\}$ is an AR process with a finite order. The strict exogeneity condition is somewhat stronger than the usual condition that is imposed in the stationary and ergodic case, i.e., no contemporaneous correlation between the regressors and the regression errors. If, however, both $\{u_t\}$ and $\{v_t\}$ are square integrable martingale differences and $\Omega = \Sigma$, then it is rather obvious that $E(u_t v_t') = 0$ or $\Sigma_{12} = 0$ is sufficient to ensure that Theorem 5.2 holds.

The results given in Theorem 5.2 can, of course, be written in a form which is more compatible with classical regression theory. For example, we have

$$\hat{A} - A \sim N(0, \Omega_1 \otimes (X'X)^{-1})$$

conditionally on a realization of x_t $(t = 1, \ldots, T)$. Similar results hold for the other estimators. Here, $X'X = O_p(T^2)$ and $T^{-2}X'X$ converges (weakly) to a stochastic matrix, in contrast to the standard case where $T^{-1}X'X$ converges (or is assumed to converge) to a constant matrix as T tends to infinity. Given the above conditional normality, however, it is not difficult to see that a quadratic form in an appropriate metric yields asymptotic chi-squared criteria, just as in the classical regression theory.

To study the asymptotic behavior of the G-statistics in this case, it is convenient to define the following functional h_R, which generalizes h in (45):

$$h_R(B,M) = g_R(B,M,0)$$

where g_R is given by (35). It is now easily deduced from Lemma 5.1 that

<u>COROLLARY 5.3</u> *Let* M_1 *be independent of* B *and rank* $R_1 = q_1$. *Then*

(a) $h_{R_1}(B, M_1) = \chi^2_{q_1}$.

Moreover, if M_2 *is also independent of* B , *if*

$$\int_0^1 M_1 M_2' = 0 \quad \text{a.s.},$$

and if rank $R_2 = q_2$, *then*

(b) $h_{R_1}(B, M_1) + h_{R_2}(B, M_2) = \chi^2_{q_1+q_2}$.

The following important results follow directly from this corollary.

<u>THEOREM 5.4</u> *If* $\Omega_{21} = \Delta_{21} = 0$, *then the limiting distributions of the G-statistics are chi-square, with degrees of freedom given by the number of restrictions for each test.*

The above theorem obviously holds for the H-statistics, and also for the "F"-statistics if the regression errors are martingale differences. A major application of Theorem 5.4 is to specification robust inferences. Observe that the construction of the G-statistics, as earlier discussed, not only accounts for contributions from the variance of the equation error but also for its serial covariance properties in general. In fact, these statistics make a nonparametric serial correlation error correction through the estimation of Ω_1 rather than Σ_1. That is, they account for serial correlation in the errors without being specific about its precise form. However, as shown in Theorem 5.4, the resulting G-statistics all have asymptotic chi-square distributions, just as if there were no serial correlation in the regression errors. Moreover, the limit distributions are identical to those of the conventional regression F-statistics when there are serially independent errors. In this sense the G-statistics enable us to make inferences that are robust to the specification of the regression errors. Note that they play precisely the same role as the statistics introduced by White and Domowitz [29]. However, in our case with integrated regressors it is no longer necessary to estimate weighted moment matrices (such as $X'VX/T$ where V is the covariance matrix of T consecutive regression errors) as in the White and Domowitz approach. In fact, as shown by Phillips and Park [21], weighted moment matrices such as $X'VX/T$ behave asymptotically like $\sigma_v^2 X'X/T$ (taking the case of a scalar regression) and consistent estimation of the asymptotic covariance matrix in this case simply requires a consistent estimate of σ_v^2 (the scaled spectrum of the regression errors). This provides an important simplification of the form of specification robust test statistics such as the G-statistics. In effect, the robust statistics are identical in form to the conventional Wald tests and involve only the replacement of conventional error variance matrix estimates by (zero frequency) spectral estimates which account for possible serial correlation.

Another interesting application of Theorem 5.4 is to the distribution of tests in the presence of spurious detrending (see Durlauf and Phillips [7] for a discussion of relevant empirical examples). Theorem 5.4 shows that spurious detrending does not affect, at least asymptotically, the nominal size of a test as long as the regressors are strictly exogenous. But spurious detrending does affect the size of tests based on an asymptotic chi-squared criterion when the regressors and the regression errors are correlated $(\Omega_{21} \neq 0)$. In fact, as we show below, the magnitude of the bias that results from detrending depends critically on the extent of the correlation between $\{u_t\}$ and $\{v_t\}$.

5.2. First Order VAR System with Simple Unit Roots

As noted earlier, our previous results in Sections 3 and 4 are directly applicable to the first order VAR system where each component time series in the system has a single unit root. We now suppose the n-vector process $\{y_t\}_0^\infty$ is generated recursively from the (possibly random) initial condition y_0 by

$$y_t = y_{t-1} + u_t$$

and we consider the least squares regressions (7)-(9) with $x_t = y_{t-1}$. The true parameter values for μ and θ are assumed to be zero. If $\{y_t\}$ is driven with a nonzero drift, the results in Theorems 3.4-3.5 apply. In this section we unify and extend the results obtained in earlier work (including those in the papers by Phillips [16], Phillips and Durlauf [19] and Phillips and Perron [22]) using our general results on regressions with integrated processes.

In the VAR system above, it is easy to see that B_2 reduces to B_1 and Δ_{21} to Λ_1 in our earlier asymptotic theory. Hence, the limiting distributions of the least squares estimators and the "F"- and G-statistics in this system are readily obtained from Theorems 3.1-3.3 and Theorem 4.1 simply by replacing B_2 with B_1 and Δ_{21} with Λ_1 . For the test of unit roots we set $R = I$ and $r = \text{vec } I$ in (19).

When (u_t) is a martingale difference sequence, $\Lambda_1 = 0$ and the limiting distributions of the "F" and G-statistics are given by Theorem 4.2 with $B_2 = B_1$. These asymptotic distributions are free of nuisance parameters, as is easily seen from:

LEMMA 5.5. *For* $M(B_1) = B_1$ *and for the various functionals* $M(\cdot)$ *introduced in* (14)-(18),

$$h(B_1, M(B_1)) = h(W, M(W))$$

where W *denotes* n-*vector standard Brownian motion (i.e., Brownian motion with covariance matrix* $= I_n$).

The "F"- and G-statistics are not useful for hypotheses testing in the general case where (u_t) is serially correlated and $\Lambda_1 \neq 0$ since the asymptotic distributions of these statistics depend on Λ_1 . In this case the transformations given in (36)-(44) come into play. This type of transformation was first introduced by Phillips [16] for testing the presence of a unit root in scalar time series; it was extended to VAR models by Phillips and Durlauf [19]; and Phillips and Perron [22] also use the technique in regressions with a constant term and time trend for univariate models.

Theorem 4.2 together with Lemma 5.5 show that the H-statistics are asymptotically invariant within a very wide class of weakly dependent and possibly heterogeneously distributed innovation $\{u_t\}$. The transformed H-statistics thus provide robust tests for the existence of unit roots as well as other, possibly joint, tests that may include the constant term and the coefficient of the time trend. Note that these tests include regression based tests for cointegration, as discussed in Phillips and Ouliaris [20]. It is also worth noting that the H-statistics from the regression (9), with the exception of $H(\tilde{\theta})$, are all invariant with respect to the introduction of a nonzero drift in $\{y_t\}$.

The distribution of $h(W, M(W))$ for $M(W) = W$ and for the various functionals $M(\cdot)$ in (14)-(18) are completely specified by n, the number of variables in the VAR system. For $n \leq 15$, some of these distributions (approximations with $T = 500$) are tabulated in Ouliaris [14]. When $n = 1$, $h(W,W)$, $h(W, W^*)$ and $h(W, W^{**})$ are, respectively, the limiting distributions of the squares of Dickey and Fuller's τ, τ_μ and τ_τ statistics, which are tabulated in Fuller [10, p. 373]. Finally, the tails of the distributions get thicker, regardless of n, as we move from $h(W,W)$ to $h(W, W^*)$ to $h(W, W^{**})$. This implies that spurious demeaning and detrending would lead to progressively greater over rejection of the null hypothesis of unit roots, if the decision were based on the chi-square table (rather than the Dickey-Fuller or Ouliaris tables).

5.3. Hypothesis Testing in a Single Equation

In the case of general linear restrictions, the asymptotic distributions of the transformed H-statistics are still dependent upon the nuisance parameters determined by the covariance matrix Ω . The multidimensionality of Ω would seem to make it virtually impossible to use the H-statistics in practice. However, the following lemma shows that in the important special case of a single equation regression (where $n = 1$) the parameter dependency becomes one dimensional. The result is also very helpful for understanding the nature of the limiting distributions of the H-statistics.

We set here $\Omega_1 = \omega_1^2$ and $\Omega_{21} = \omega_{21}$.

LEMMA 5.6. *When* $n = 1$, *the distribution of* $h(B_1, M)$ *for* $M = B_2$ *(or any functional of* B_2 *given in* (14)-(18)) *depends only on* m *and*

$$\rho^2 = \omega_{21}'\Omega_2^{-1}\omega_{21}/\omega_1^2 .$$

Moreover,

$$h(B_1, M(B_2)) = \ell(W_1, W_2, \rho)\ell(W_1, W_2, \rho)' \tag{47}$$

where $W = (\overset{1}{W_1}, \overset{m}{W_2})$ *is* (m+1)*-vector standard Brownian motion. Here* ℓ *is defined by*

$$\ell(W_1, W_2, \rho) = (1 - \rho^2)^{1/2} k(W_1, M(W_2)) + \rho k(\overline{W}_2, M(W_2)) \tag{48}$$

where k *is the functional given in* (46) *and*

$$\overline{W}_2(r) = \frac{1}{\sqrt{m}} \Sigma_{j=1}^{m} W_{2j}(r) .$$

Lemma 5.6 shows that the distribution of $h(B_1, M(B_2))$ is a mixture of that of $h(W_1, M(W_2))$ and $h(\bar{W}_2, M(W_2))$, where the weights are determined by ρ , the multiple correlation coefficient of B_1 and B_2 . Interestingly, when $\rho = 0$, then it follows from (47) and (48) that $h(B_1, M(B_2)) = h(W_1, M(W_2)) = \chi^2_m$, corresponding to our strictly exogenous regressors case. On the other hand, when $\rho = 1$ we have $h(B_1, M(B_2)) = h(\bar{W}_2, M(W_2))$, as in the case of Lemma 5.5. When $\rho = 1$ and $m = 1$ this reduces to the limit distribution of the square of the t-ratio statistic in the random walk model.

The above observations make it plain that the special cases that we have considered in this section comprise two polar cases, one involving the chi-square distribution of conventional regression theory and the other a multivariate generalization of standard unit root theory. We see from Lemma 5.6 that the general theory may be regarded as a simple mixture of these two extremes.

It is also interesting to note that the representation (47) provides a definitive answer to the effect of spurious demeaning or detrending on the true size of testing procedure based (nominally) on chi-square tables. As $\rho^2 \to 1$, the distributions of $h(B_1, B_2)$, $h(B_1, B_2^*)$ and $h(B_1, B_2^{**})$ involve mixtures that attach less weight to the chi-square distribution (that delivers the nominal size of the test) and more weight to the unit root distribution.

The implications of Lemma 5.6 have as much practical importance as they do theoretical. In particular the asymptotic distributions of the H-statistics are completely specified, for a given number of regressors, by the limiting multiple correlation coefficient ρ between the processes driving the regressors and the regression errors. These distributions are invariant up to ρ for a wide class of regression models with integrated regressors and they may be used in such models as a basis for statistical inference.

6. CONCLUDING REMARKS

The framework of analysis developed in Section 3 of this paper helps to simplify and extend a good deal of earlier work in this field. A key element in this simplification is the functional defined in (13). Through the use of this functional it is possible to represent all the major asymptotic distributions of interest in the same form. Differences in these distributions then arise simply as a consequence of differences in the arguments of the functional. As is clear from the examples of Section 3, these arguments are themselves readily interpreted in terms of the underlying regression equation. Thus, when the regressors include a constant and a time trend, the argument of the functional corresponding to the regressor is simply a detrended Brownian motion.

The results of Section 3 also provide a fairly general asymptotic theory for cases where there is a drift in a VAR or a drift in the regressor variable. In such cases the limiting distribution theory is, in general, degenerate. However, contrary to the conclusion of West [27], it is also generally non normal. Only when there is a single nonstationary regressor with nonzero drift do the usual normal asymptotics that are associated with deterministic trends apply. Our theory includes the latter situation as a special case of more generally applicable results.

APPENDIX: MATHEMATICAL PROOFS

1. Proof of Lemma 2.1. Proofs of (a), (c) and (e) can be found in Phillips an Durlauf [19]. The remaining parts (b) and (d) are trivial extensions of results in Phillips and Perron [22] to the multivariate case. Note that joint weak convergence of (a)-(e) may also be established by writing the vector of sample moments as a functional of $X_T(r)$ up to an error of $o_p(T^{-1/2})$.

2. Proof of Theorem 3.1. See Theorem 4.1 of Phillips and Durlauf [19].

3. Proof of Theorem 3.2. Part (a) is easily deduced from Theorem 4.1(a) and Theorem 6.1(a) of Phillips and Durlauf [19]. To prove (b), we first write

$$T^{1/2}(\bar{\mu}-\mu) = \left\{T^{-1/2}\Sigma u_t - (T^{-1}\Sigma u_t x_t')\left[T^{-2}\Sigma x_t x_t'\right]^{-1}(T^{-3/2}\Sigma x_t)\right\}$$
$$\cdot\left\{1 - (T^{-3/2}\Sigma x_t')\left[T^{-2}\Sigma x_t x_t'\right]^{-1}(T^{-3/2}\Sigma x_t)\right\}^{-1}$$

Now Lemma 2.1 and the continuous mapping theorem (Billingsley [2], Corollary 1, p. 31) yield (b). Notice also that

$$\int_0^1 P_1^2 = 1 - \int_0^1 B_2'\left[\int_0^1 B_2 B_2'\right]^{-1}\int_0^1 B_2$$

as is required for the stated formula.

4. Proof of Theorem 3.3. The stated results are obtained as in the proof of Theorem 3.2 by using Lemma 2.1 and the continuous mapping theorem.

5. Proof of Theorem 3.4. See Phillips and Durlauf [19]: the result for $\hat{\Sigma}_1$ is their Theorem 4.2 and the proofs for $\bar{\Sigma}_1$ and $\tilde{\Sigma}_1$ are entirely analogous to their Theorem 6.1(g).

6. Proof of Theorem 3.5. Since

$$z_{1t} = (\pi'\pi)^{1/2}t + h_1'x_t^0$$

$$z_{2t} = H_2'x_t^0$$

we have, as in Lemma 2.1,

$$T^{-3/2}\Sigma_1^T z_{2t} \Rightarrow \int_0^1 \underline{B}_2$$

$$T^{-5/2}\Sigma_1^T z_{1t}z_{2t} \Rightarrow (\pi'\pi)^{1/2}\int_0^1 t\underline{B}_2$$

$$T^{-2}\Sigma_1^T z_{2t}z_{2t}' \Rightarrow \int_0^1 \underline{B}_2\underline{B}_2'$$

$$T^{-3/2}\Sigma_1 z_{1t}u_t \Rightarrow (\pi'\pi)^{1/2}\int_0^1 t\,dB_1$$

$$T^{-1}\Sigma_1^T z_{2t}u_t' \Rightarrow \int_0^1 \underline{B}_2 dB_1' + \Delta_{21}$$

where the notation is defined in the text. The stated results now follow in a straightforward manner from the regression representations of $T^{3/2}(\hat{c}_1 - c_1)$ and $T(\hat{C}_2 - C_2)$.

7. Proof of Theorem 3.6. The proof of (a), (b) and (c) is entirely analogous to that of Theorem 3.2. Part (d) follows as in the proof of Theorem 3.5.

8. Proof of Theorem 4.1. The limiting null distributions of the G-statistics for a single set of hypotheses in (19)-(21) are easily deduced from the following results and Theorems 3.1-3.3 by application of the continuous mapping theorem. Note that whenever necessary joint weak convergence of the relevant quantities may be established simply by writing these quantities as a continuous functional of the random element $X_T(r)$ up to an error of $o_p(1)$. It is simple to show that:

$$
\begin{aligned}
T^2 M_0 &\Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1} \\
T^2 M_{12} &\Rightarrow \left(\int_0^1 B_2^* \, B_2^{*\prime}\right)^{-1} \\
T^2 M_{23} &\Rightarrow \left(\int_0^1 B_2^{**} B_2^{**\prime}\right)^{-1} \\
T m_{11} &\Rightarrow \left(\int_0^1 P_1^2\right)^{-1} \\
T m_{21} &\Rightarrow \left(\int_0^1 P_2^2\right)^{-1} \\
T^3 m_{22} &\Rightarrow \left(\int_0^1 P_3^2\right)^{-1}
\end{aligned}
\tag{A1}
$$

To prove the stated results for $\bar{G}$, $\tilde{G}$ and $\tilde{G}_1$, we use the following lemma where the notation is not relevant to that of the remainder of the paper.

LEMMA A.1. *Consider the classical linear regression models*

$$\text{(I)} \quad y = X_1\beta_1 + X_2\beta_2 + u$$

and

$$\text{(II)} \quad y_1 = X_1\beta_1 + u \ , \quad y_1 = y - X_2\beta_2^0$$

with a known covariance matrix for u . *We assume without loss of generality (by transforming the models if necessary), that* $Euu' = I$.

Let F_1 , F_2 *and* F *be the Wald statistics for the following hypotheses:*

F_1 : $R\beta_1^0 = r$ *in* (II)

F_2 : $\beta_2 = \beta_2^0$ *in* (I)

F : $R\beta_1^0 = r$ *and* $\beta_2 = \beta_2^0$ *in* (I)

Then under the null hypothesis $\beta_2 = \beta_2^0$,

$$F = F_1 + F_2 \ . \tag{A2}$$

Proof of Lemma A.1. Define the following least squares estimators:

$\hat{\beta}_1$, $\tilde{\beta}_1$: β_1 in (II), unrestricted and restricted to $R\tilde{\beta}_1 = r$ respectively; and

$\bar{\beta}_2$: β_2 in (I).

We also let $X = (X_1, X_2)$.

It is well known that

$$F = SS_H - SS$$

where SS_H and SS are, respectively, the restricted and unrestricted sum of squares from regression (I). Similarly, we have for (II)

$$F_1 = (y_1 - X_1\tilde{\beta}_1)'(y_1 - X_1\tilde{\beta}_1) - u'(I - P_{X_1})u \ .$$

However, if $\beta_2 = \beta_2^0$, $y_1 = y - X_2\beta_2^0$ and

$$\begin{aligned} SS_H &= (y - X_1\tilde{\beta}_1 - X_2\beta_2)'(y - X_1\tilde{\beta}_1 - X_2\beta_2^0) \\ &= (y_1 - X_1\tilde{\beta}_1)'(y_1 - X_1\tilde{\beta}_1) \\ &= F_1 + u'(I - P_{X_1})u \end{aligned}$$

whereas,

$$SS = u'(I - P_X)u \ .$$

Finally, notice that under the null hypothesis of $\beta_2 = \beta_2^0$

$$\begin{aligned} F_2 &= (\bar{\beta}_2 - \beta_2^0)'[X'(I - P_{X_1})X]^{-1}(\beta_2 - \beta_2^0) \\ &= u'(P_X - P_{X_1})u \end{aligned}$$

which completes the proof of Lemma A.1. □

It can be easily seen that the above result is valid also for the standard multivariate system by vectorization. If the covariance matrix is not known and has to be estimated, (A2) is not strictly true, but the equality continues to hold as long as we use a common estimate for the covariance matrix. Further, (A2) holds asymptotically, in general, if the estimated covariances from (I) and (II) have a common limit in probability.

The result for $\bar{G}$ now can be easily obtained from a trivial application of Lemma A.1 to $G(\hat{A})$ and $G(\bar{\mu})$. Moreover, it is not difficult to see that (A5) remains true when there are additional variables in the regressions, the coefficients of which are unrestricted. This implies that $\tilde{G}_1$ is just the sum of $G(\bar{A})$ and $G(\tilde{\theta})$, as is stated. For $\tilde{G}$, we only need to compare the regression equations (8) and (9) to get $\tilde{G} = \bar{G} + G(\tilde{\theta})$, which yields the stated result.

9. <u>Proof of Theorem 4.2</u>. All the stated results follow directly by applying the continuous mapping theorem. In particular

$$T^{1/2}\hat{\beta}_1 \Rightarrow \left(\int_0^1 B_2 B_2\right)^{-1} \int_0^1 B_2$$

$$T^{1/2}\hat{\beta}_2 \Rightarrow \left(\int_0^1 B_2 B_2' - 3\int_0^1 sB_2 \int_0^1 sB_2'\right)^{-1} \left(\int_0^1 B_2 - \tfrac{3}{2}\int_0^1 sB_2\right)$$

$$T^{-1/2}\hat{\beta}_2 \Rightarrow \left(\int_0^1 B_2 B_2' - \int_0^1 B_2 \int_0^1 B_2'\right)^{-1} \left(\int_0^1 sB_2 - \tfrac{1}{2}\int_0^1 B_2\right)$$

and therefore, $\tilde{\delta}_j \Rightarrow \delta_j$ for $j = 1, 2, 3$. For the joint tests, again apply Lemma A.1 and notice that in this case with $R = I$ we have

$$\bar{G} \Rightarrow g(B_1, 1, 0) + g(B_1, B_2^*, \Delta_{21})$$

$$\tilde{G} \Rightarrow g(B_1, 1, 0) + g(B_1, r^*, 0) + g(B_1, B_2^{**}, \Delta_{21})$$

$$\bar{G} \Rightarrow g(B_1, r^*, 0) + g(B_1, B_2^{**}, \Delta_{21}) .$$

10. Proof of Lemma 5.1. In what follows we use the symbol " $.|_M$ " to signify the conditional distribution given a realization of $M(r)$. Write

$$\mathrm{vec}(k(B,M)) = \left\{I \otimes \left(\int_0^1 MM'\right)^{-1/2}\right\}\int_0^1 dB \otimes M$$

and recall that

$$E(dBdB') = \Omega dr$$

where dr is Lebesgue measure on R . Since B is Gaussian and independent of M we have

$$\int_0^1 dB \otimes M\Big|_M = N\left(0,\ \Omega \otimes \int_0^1 MM'\right)$$

and

$$k(B,M)\Big|_M = N(0,\ \Omega \otimes I_p) \ .$$

However, since the latter distribution does not depend on realizations of $M(r)$, it is also the unconditional distribution, giving the result as stated.

11. Proof of Theorem 5.2. If $\Delta_{21} = 0$, it is easily deduced from (A1) and the continuous mapping theorem that the limiting distribution of each of the given statistics can be written in terms of $k(B_1, M)$, where M is B_2 or a function of B_2 explicitly given in (14)-(18) for each statistic. Since $B = (B_1, B_2)$ is Gaussian, $\Omega_{21} = 0$ implies that B_1 is independent of B_2 , and therefore of any functional of B_2 . Now a direct application of Lemma 5.1 yields the stated asymptotic normality.

12. Proof of Corollary 5.3. By the argument used in the proof of Lemma 5.1 we deduce that

$$\left\{R_1\left[\Omega \otimes \left(\int_0^1 M_2 M_1'\right)^{-1}\right]R_1'\right\}^{-1/2} R_1\left\{I \otimes \left(\int_0^1 M_1 M_1'\right)^{-1}\right\}\int_0^1 dB \otimes M_1 \equiv N(0, I_{q_1})$$

leading directly to part (a). For (b), note that

$$E\left\{\left(\int_0^1 dB \otimes M_1\right)\left(\int_1^1 dB' \otimes M_2'\right)\middle| M\right\} = \Omega \otimes \int_0^1 M_1 M_2' = 0$$

where $M = (M_1, M_2)$. The stated result now follows immediately.

13. Proof of Theorem 5.4. When $\Delta_{21} = 0$, the limiting distributions of the G-statistics in Theorem 4.1 become of the form $h_R(B_1, M)$, where M is B_2 or a function of B_2 given correspondingly for each test. Since $B = (B_1, B_2)$ is Gaussian, $\Omega_{21} = 0$ implies that B_1 is independent of B_2 and, hence, of any function of B_2 . The results for the G-statistics which test a single set of hypotheses now follow immediately from Corollary 5.3(a). To get the stated result for the joint test case, we simply apply part (b) of the corollary, making use of the fact that:

$$\int_0^1 B_2 P_j = 0 \quad (j = 1, 2, 3)$$

$$\int_0^1 B_2^* P_3 = 0$$

and

$$\int_0^1 P_1 P_3 = 0 .$$

14. Proof of Lemma 5.5. Write

$$h(B_1, M) = tr\left\{\Omega_1^{-1/2}\int_0^1 dB_1 M'\left[\int_0^1 MM'\right]^{-1}\int_0^1 M dB_1'\Omega_1^{-1/2}\right\} . \tag{A3}$$

Now transform $B_1 \to \Omega_1^{-1/2}B_1$ and notice that

$$M(B_1)'\left[\int_0^1 M(B_1)M(B_1)'\right]^{-1}M(B_1) = M(W)'\left[\int_0^1 M(W)M(W)'\right]^{-1}M(W)$$

for $M(B_1) = B_1$ or the various functionals of B_1 in (14)-(18).

15. Proof of Lemma 5.6. Let

$$B_1 = \omega_1\left[(1 - \rho^2)^{1/2}W_1 + r'W_2\right] , \quad r = \Omega_2^{-1/2}\omega_{21}/\omega_1$$

$$B_2 = \Omega_2^{1/2}W_2 .$$

We have from (A3) that

$$h(B_1, M(B_2)) = \int_0^1\left[(1 - \rho^2)^{1/2}dW_1 + r'dW_2\right]M(W_2)' \tag{A4}$$

$$\cdot\left[\int_0^1 M(W_2)M(W_2)'\right]^{-1}\int_0^1 M(W_2)\left[(1 - \rho^2)^{1/2}dW_1 + r'dW_2\right]$$

To show that the distribution of $h(B_1, M(B_2))$ depends only on ρ and not on r , suppose we have two sets of parameters (ρ, r_1) and (ρ, r_2) . Then by the Vinograd theorem (Muirhead [13], Theorem A.9.5, p. 589), there exists an orthogonal matrix H such that

$$r_2 = Hr_1 .$$

It is, however, not difficult to check that the expression in (A4) is invariant with respect to the transformation

$$W_2 \to H'W_2$$

for $M(W_2) = W_2$ and all functionals $M(\cdot)$ implied by (14)-(18). This proves the assertion. To get the expression given in Lemma 5.6, we simply take

$$r = \frac{\rho}{\sqrt{m}}(1, \ldots, 1)' .$$

REFERENCES

[1] Andrews, D. W. K. Least squares regression with integrated or dynamic regressors under weak error assumptions. *Econometric Theory* 3 (1987): 98-116.

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

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

[4] Chan, N. H. and C. Z. Wei. Limiting distributions of least square estimates of unstable autoregressive processes. *Annals of Statistics* (1988): forthcoming.

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

[6] __________. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49 (1981): 1057-1072.

[7] Durlauf, S. N. and P. C. B. Phillips. Trends versus random walks in time series analysis. Cowles Foundation Discussion Paper No. 788, Yale University, 1986.

[8] Eberlain, E. On strong invariance principles under dependence assumptions. *Annals of Probability* 14 (1986): 260-270.

[9] Engle, R. F. and C. W. J. Granger. Cointegration and error correction: Representation, estimation and testing. *Econometrica* 55 (1987): 251-276.

[10] Fuller, W. A. *Introduction to Statistical Time Series.* New York: John Wiley, 1976.

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

[12] Krämer, W. Least squares regression when the independent variable follows an ARIMA process. *Journal of the American Statistical Association* 81 (1986): 150-154.

[13] Muirhead, R. J. *Aspects of Multivariate Statistical Theory.* New York: John Wiley, 1982.

[14] Ouliaris, S. Testing for unit roots and cointegration in multiple time series. Yale Doctoral Dissertation (in preparation), 1987.

[15] Phillips, P. C. B. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33 (1986): 311-340.

[16] __________. Time series regression with a unit root. *Econometrica* 55 (1987): 277-301.

[17] __________. Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$. *Journal of Multivariate Analysis* (1988): forthcoming.

[18] __________. Weak convergence of sample covariance matrix to stochastic integrals via martingale approximations. Cowles Foundation Discussion Paper No. 846, Yale University, 1987.

[19] __________ and S. N. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies* 53 (1986): 473-496.

[20] __________ and S. Ouliaris. Testing for Cointegration. Cowles Foundation Discussion Paper No. 809, Yale University, 1986.

[21] __________ and J. Y. Park. Asymptotic equivalence of OLS and GLS in regressions with integrated regressors. Cowles Foundation Discussion Paper No. 802, Yale University, *Journal of Americal Statistical Association* (1988): forthcoming.

[22] __________ and Pierre Perron. Testing for a unit root in time series regression. Cowles Foundation Discussion Paper No. 795, Yale University, 1986.

[23] Stock, J. Asymptotic properties of least squares estimators of cointegrating vectors. Mimeo, Harvard University, 1985.

[24] __________ and M. Watson. Testing for common trends. Mimeo, Harvard University, 1986.

[25] Strasser, H. [1986] Martingale difference arrays and stochastic integrals. *Probability Theory and Related Fields* 72, 83-98.

[26] Tsay, R. S. and G. C. Tiao [1980]. Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. University of Chicago, Technical Report #54.

[27] West, K. D. Asymptotic normality, when regressors have a unit root. Princeton Discussion Paper in Economics #110, Princeton University, 1986.

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

[29] __________ and I. Domowitz. Nonlinear regression with dependent observations. *Econometrica* 52 (1984): 143-162.