Cowles Foundation Paper 711

*Econometrica*, Vol. 56, No. 5 (September, 1988), 1021–1043

# REGRESSION THEORY FOR NEAR-INTEGRATED TIME SERIES

BY P. C. B. PHILLIPS[1]

The concept of a near-integrated vector random process is introduced Such processes help us to work towards a general asymptotic theory of regression for multiple time series in which some series may be integrated processes of the ARIMA type, others may be stable ARMA processes with near unit roots, and yet others may be mildly explosive A limit theory for the sample moments of such time series is developed using weak convergence and is shown to involve simple functionals of a vector diffusion The results suggest finite sample approximations which in the stationary case correspond to conventional central limit theory. The theory is applied to the study of vector autoregressions and cointegrating regressions of the type recently advanced by Engle and Granger (1987). A noncentral limiting distribution theory is derived for some recently proposed multivariate unit root tests This yields some interesting insights into the asymptotic power properties of the various tests. Models with drift and near-integration are also studied The asymptotic theory in this case helps to bridge the gap between the nonnormal asymptotics obtained by Phillips and Durlauf (1986) for regressions with integrated regressors and the normal asymptotics that usually apply in regressions with deterministic regressors.

KEYWORDS: Brownian motion, cointegration, diffusion, near-integration, unit root tests

## 1. INTRODUCTION

MANY OBSERVED TIME SERIES in economics seem to be modeled rather well by integrated processes. The simplest model generating an integrated process is, of course, a random walk; and this is a model that has been widely used in financial and commodity market studies, in theories of rational expectations, and in recent work with aggregate economic time series. More general models of the ARIMA type have also been used frequently in econometric work and have been found to represent very adequately the movements in many different economic series. Moreover, in a recent study Nelson and Plosser (1982) provide substantial empirical evidence that a wide selection of macroeconomic time series for the U.S. are modeled better in terms of integrated processes than as stationary processes about a deterministic trend. In fact, their findings support autoregressive representations with unit roots for all but one of the historical time series in their study.

It is also known that the discriminatory power of statistical tests for the presence of unit roots is generally quite low against the alternative of roots which are close (but not equal) to unity. This is explained by the fact that the distributions of the relevant test statistics in finite samples of data are usually quite similar under the null and the alternative hypotheses in such cases. Thus, strongly autoregressive processes or even mildly explosive processes must often be considered as realistic alternatives in many cases where the statistical tests may actually support the null hypothesis of a unit root.

[1] I am grateful to the editors and two referees for comments on earlier versions of this paper. My thanks also go to Glena Ames for her skill and effort in typing the manuscript of the paper Research support from the NSF is gratefully acknowledged

Time series which possess an autoregressive component with a root close (but not necessarily equal) to unity provide an important general class of processes which we describe as near-integrated. The class may be taken to include stationary time series with a strongly autoregressive component and nonstationary time series with a mildly explosive root as well as integrated processes of the ARIMA type. Thus, the class of near-integrated processes with which this paper is concerned is rather wide.

The simulation studies of Evans and Savin (1981, 1984) gave rise to the interesting finding that the coefficient estimator and the $t$ test in a stationary AR(1) with a root near unity have statistical properties even in moderately large samples ($T = 50, 100$) that are closer to the asymptotic theory for a random walk than they seem to be to the classical asymptotic theory that applies for stationary time series. Similar results also seem to apply when the AR(1) is mildly explosive. In all cases the approach to the strictly correct asymptotic distribution is very slow as the sample size $T \uparrow \infty$. These results suggest that an alternative asymptotic theory may be of value, one which takes into account the fact that the time series under study are near-integrated processes.

The primary object of the present paper is to develop such a theory. We shall work explicitly with multiple time series in which some series may be integrated processes of the ARIMA type, others may be stationary ARMA processes with roots near unity, while yet others may be mildly explosive series. These alternatives are determined by the values assumed by the elements of a certain noncentrality parameter matrix. This matrix occurs in the formulation of the near-integrated process model and enables us to assess the impact on the asymptotic theory of the presence of various forms of near-integration.

The organization of the paper is as follows. Section 2 develops some preliminary notation, assumptions, and theory that are useful throughout the rest of the paper. The concept of a near-integrated system is introduced and examples are given illustrating several interesting special cases. Section 3 develops a limit theory for the sample moments of a near-integrated time series and relates the results to conventional central limit theory for stationary processes. In Section 4 the theory is applied to the study of vector autoregressions and is extended to include regressions that involve cointegrated series. A noncentral limiting distribution theory is derived in Section 5 for the multivariate unit root tests that have been proposed recently by Phillips and Durlauf (1985) and Park and Phillips (1986). These noncentral distributions help in the analysis of the local asymptotic power properties of the various tests. Section 6 shows how the theory may be extended to allow for systems with near unit roots and nonzero drift. The results of this section help to bring together the apparently divergent theories of regression with integrated processes (that leads to nonnormal asymptotics) and regression with deterministic regressors (that leads to conventional normal asymptotics). Section 7 develops an asymptotic theory for multiple regressions with near-integrated time series. The results of this section include the spurious regressions theory given recently by the author (1986a) and a theory for cointegrating regressions of the type that have been advanced by Engle and

Granger (1987). Some conclusions are given in Section 8. The Appendix contains a brief outline of some proofs of results in the text.

## 2 PRELIMINARY THEORY AND DISCUSSION

Let $\{u_t\}_0^\infty$ be a weakly stationary sequence of random $n$-vectors. We introduce the vector of partial sums $S_t = \Sigma_{j=1}^t u_j$ and set $S_0 = 0$. Throughout the paper we assume that $\{u_t\}_0^\infty$ satisfies the following conditions:

(A) $\quad E(u_0) = 0,$

(B) $\quad E\|u_0\|^{\beta+\varepsilon} < \infty \quad \text{for some} \quad \beta > 2,$

(C) $\quad \{u_t\}_0^\infty$ is strong mixing with mixing numbers $\alpha_m$ that satisfy

$$\sum_1^\infty \alpha_m^{1-2/\beta} < \infty.$$

These conditions allow for many weakly dependent time series and include a broad class of data generating mechanisms such as finite order ARMA models under very general conditions (see Withers (1981)). Note that (B) and (C) imply that

$$\Omega = \lim_{T\to\infty} T^{-1}E(S_T S_T') = E(u_0 u_0') + \sum_{k=1}^\infty E(u_0 u_k' + u_k u_0')$$

(Ibragimov and Linnik (1971, Theorem 18.5.3)). Moreover, since this series is absolutely convergent, the spectral density matrix $f_{uu}(\lambda)$ of $\{u_t\}$ exists, is continuous, and $\Omega = 2\pi f_{uu}(0)$. Except where explicitly noted, we shall further require:

(D) $\quad \Omega$ is positive definite.

From the partial sum process $\{S_t\}$ we construct

$$X_T(r) = T^{-1/2}S_{[Tr]} = T^{-1/2}S_{j-1} \quad ((j-1)/T \leqslant r < j/T;\ j = 1, \ldots, T),$$

where $[Tr]$ denotes the integer part of $Tr$. Under the conditions given above a functional central limit theory holds for the random element $X_T(r)$ as $T \uparrow \infty$. In particular, we have:

(1) $\quad X_T(r) \Rightarrow B(r)$

where $B(r)$ is $n$-vector Brownian motion with covariance matrix $\Omega$. In (1) and elsewhere in the paper, the symbol "$\Rightarrow$" signifies weak convergence of the associated probability measures and the limit is taken as the sample size $T \uparrow \infty$. Multivariate invariance principles such as (1) above have recently been given by Eberlain (1986), Phillips and Durlauf (1986), and Phillips (1987c). The reader is referred to these papers for further discussion.

Our main concern will be with multiple time series that are generated by the following model:

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

with

(3) $A = \exp(T^{-1}C)$.

In (3) $C$ is a fixed, real $n \times n$ matrix. Formally we should write $A = A_T$, signifying explicitly the dependence of the coefficient matrix on $T$. Strictly speaking, time series generated by (2) constitute a triangular array of the type $\{\{y_{tT}\}_{t=1}^{T}\}_{T=1}^{\infty}$. However, this formality is not critical to our development and in order not to overburden notation we simply refer to time series generated by (2) as $\{y_t\}_0^{\infty}$. Initial conditions are set at $t = 0$ and $y_0$ may be any random variable (including a constant) whose distribution is fixed and independent of $T$.

We call time series that are generated by (2) and (3) near-integrated. This follows the terminology introduced in Phillips (1987b) for univariate processes. The matrix $C$ in (3) may be interpreted as a noncentrality parameter matrix. It may be used to measure deviations from the following null hypothesis

$$H_0\colon A = I$$

which applies when $C = 0$. In this case $\{y_t\}$ is a vector integrated process of order one (an $I(1)$ process) in the sense that its first differences are stationary (or $I(0)$).

When $C \neq 0$, (3) represents a local alternative to $H_0$. As $T \uparrow \infty$ of course $A \to I_n$. However, the rate of approach to $I_n$ is not so fast that the alternative hypothesis represented by (3) has no impact on the limiting distribution theory that we shall develop. In fact, the rate of approach is controlled so that the effect of the alternative hypothesis (3) on the limiting distribution of statistics based on data generated by (2) is well defined and directly measurable in terms of the noncentrality parameter matrix $C$. Note that an alternate and asymptotically equivalent approach would have been to replace the matrix exponential representation of $A$ in (3) by deviations from $I_n$ of the form: $A = I_n + T^{-1}C$. With this formulation the approach would be analogous to that which is conventionally employed in the statistical analysis of asymptotic power under local alternatives. This alternate approach is used by Chan and Wei (1987) in the scalar case and for innovations that are martingale differences. Their results come within the framework of the theory developed in Phillips (1987b) for univariate processes.

Near-integrated systems such as (2) and (3) accommodate many interesting possibilities. For example, when $C = \mathrm{diag}(c_1, \ldots, c_n)$, $\{y_t\}$ is a multiple time series in which some series may be I(1) processes of the ARIMA type (corresponding to components with $c_i = 0$), some may be stable ARMA processes with near unit roots ($c_j < 0$), and yet others may be mildly explosive ($c_k > 0$). Moreover, if $C$ has nonzero off-diagonal elements the system allows for series which may be near-integrated of different orders. Thus, when $n = 2$ and

$$C = \begin{bmatrix} 0 & c \\ 0 & 0 \end{bmatrix},$$

we have

$$A = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \quad \text{with} \quad a = c/T.$$

In this case $\{y_{2t}\}$ is an I(1) process but $\{y_{1t}\}$ behaves like an I(2) process in finite samples of data since $a \neq 0$ when $c \neq 0$. Note that in this case $y_{1t}$ and $y_{2t}$ may be regarded as being nearly cointegrated because the linear combination which selects $y_{2t}$ reduces the order of integration from (nearly) I(2) to I(1). This is an example of trivial cointegration.

A less trivial example is the following. Let $C = -bb'$ for some nonzero $n$-vector $b$. Define $f = b'b$ and write $b = hf^{1/2}$ where $h = b/(b'b)^{1/2}$. Note that

$$Ah = e^{-f/T}h \tag{4}$$

so that $h$ is an eigenvector of $A$. Note also that

$$h'y_t = e^{-f/T}h'y_{t-1} + h'u_t \tag{5}$$

and since $f > 0$ we deduce that the series $h'y_t$ is nearly stationary. It follows that the time series $\{y_t\}$ is nearly cointegrated in the sense that the linear combination $h'y_t$ is nearly stationary or I(0), as distinct from I(1). Moreover, linear regressions which relate the components of $y_t$ fall in the usual category of spurious regressions (Granger and Newbold (1974), Phillips (1986a); but when (4) and (5) apply they may be interpreted as regressions for nearly cointegrated series. The asymptotic theory we develop therefore applies to cointegrating regressions and delivers asymptotic local power functions for regression based cointegration tests. Finally, in a sequence of models with increasing $f$ we may regard cointegration as the natural limit of a spurious regression as $f \uparrow \infty$.

## 3. SAMPLE MOMENTS OF NEAR-INTEGRATED TIME SERIES

Let $\{y_t\}_0^\infty$ be a near-integrated time series generated by (2) and (3). In developing an asymptotic theory of regression for $y_t$ we make extensive use of the following functional:

$$K_C(r) = \int_0^r e^{(r-s)C}\,dB(s)$$

where $B(s)$ is vector Brownian motion with covariance matrix $\Omega$. $K_C(r)$ is a vector diffusion process and satisfies the stochastic differential equation system:

$$dK_C(r) = CK_C(r)\,dr + dB(r); \qquad K_C(0) = 0. \tag{6}$$

We may also write

$$K_C(r) = B(r) + C\int_0^r e^{(r-s)C}B(s)\,ds,$$

and in this representation the effect of the noncentrality matrix $C$ is more evident. $K_C(r)$ is a Gaussian process and for fixed $r$ the finite dimensional distribution

$$K_C(r) \equiv N(0, Q), \qquad Q = \int_0^r e^{(r-s)C}\Omega e^{(r-s)C'}\,ds \tag{7}$$

is easy to obtain. In this expression and elsewhere in the paper, we use the symbol "$\equiv$" to represent equality in distribution.

Using an approach developed in earlier work (Phillips (1987a, 1987b)) it is easy to study the asymptotic behavior of sample moments of $y_t$. The main results we need are collected together in the following lemma. Here and elsewhere in the paper all asymptotic results apply as $T \uparrow \infty$; and to achieve notational economy we frequently eliminate function arguments and write, for example, $K_C$ in place of $K_C(r)$ and $\int_0^1 K_C$ in place of $\int_0^1 K_C(r)\,dr$.

LEMMA 3.1:

(a) $$T^{-1/2} y_{[Tr]} \Rightarrow K_C;$$

(b) $$T^{-3/2} \sum_1^T y_t \Rightarrow \int_0^1 K_C;$$

(c) $$T^{-2} \sum_1^T y_t y_t' \Rightarrow \int_0^1 K_C K_C';$$

(d) $$T^{-2} \sum_1^T (y_t - \bar{y})(y_t - \bar{y})' \Rightarrow \int_0^1 K_C K_C' - \int_0^1 K_C \int_0^1 K_C';$$

(e) $$T^{-1} \sum_1^T y_{t-1} u_t' \Rightarrow \int_0^1 K_C \, dB' + \Omega_1;$$

*where*

$$\Omega_1 = \sum_{k=1}^{\infty} E(u_0 u_k').$$

This Lemma gives an asymptotic theory for the sample moments of a near-integrated vector process. As in the case of an integrated process, these sample moments (when appropriately standardized) converge weakly to random matrices rather than constants as $T \uparrow \infty$. The limiting distributions of these sample moments are characterized as functionals of the vector diffusion $K_C$. When $C = 0$, $K_C = B$, and the results specialize to those given in earlier work (Phillips and Durlauf (1986)) for I(1) processes. Note that in the case of (a) and (b) we have linear functionals of $K_C$, so that the asymptotic distributions are Gaussian. The first of these is already given in (7). The second is found by a simple calculation to be

(8) $$\int_0^1 K_C \equiv N(0, V)$$

with

(9) $$V = \int_0^1 \int_0^1 \int_0^{r \wedge p} \exp\{(r-s)C\} \Omega \exp\{(p-s)C'\} \, ds \, dp \, dr.$$

In the scalar case (set $n = 1$, $V = v$, $\Omega = \omega^2$, $C = c$) the limiting variance is:

(10) $$v = \omega^2/c^2 + (\omega^2/2c^3)(e^{2c} - 4e^c + 3).$$

Note that for particular cases of (2) in which $A$ is assigned a value in the vicinity of $I_n$ the results of the Lemma may be used to suggest simple asymptotic approximations to the distributions of the sample moments. Thus, if $n=1$ and $A=a$ is close, but not equal, to unity we have $a=e^{c/T}$ so that $c=T\ln a$ and then

$$v=\frac{\omega^2}{(T\ln a)^2}+\frac{\omega^2}{2(T\ln a)^3}[3-4a^T+a^{2T}]$$

is an approximation to the variance of $T^{-3/2}\Sigma_1^T y_t$. In the stationary case ($a<1$) this suggests that $T^{-1/2}\Sigma_1^T y_t$ is approximately $N(0,v_T)$ where

$$v_T=\frac{\omega^2}{(\ln a)^2}+\frac{\omega^2}{2T(\ln a)^3}[3-4a^T+a^{2T}]. \tag{11}$$

The leading term here may be approximated as

$$\begin{aligned} v_t &= \omega^2/(\ln a)^2+O(T^{-1}) \\ &\sim \omega^2/(1-a)^2 \end{aligned} \tag{12}$$

when $a$ is close to unity.

Interestingly, the approximation (12) gives the exact asymptotic variance in the stationary case for all values of $a$ ($|a|<1$). Indeed, we know that in this case

$$T^{-1/2}\sum_1^T y_t \Rightarrow N(0,2\pi f_y(0)) \quad \text{as} \quad T\uparrow\infty$$

(see, for instance, Hall and Heyde (1980), p. 135), where $f_y(\lambda)$ is the spectral density of the stationary process $\{y_t\}_0^\infty$. Here $y_t$ is generated by the stable AR(1) $y_t=ay_{t-1}+u_t$ with stationary errors $u_t$ and fixed autoregressive coefficient $|a|<1$. The spectral density of $y_t$ is given by:

$$f_y(\lambda)=|1-ae^{i\lambda}|^{-2}f_u(\lambda)$$

where $f_u(\lambda)$ is the spectral density of the error process $\{u_t\}$. Moreover,

$$f_y(0)=(1-a)^{-2}f_u(0)=(1-a)^{-2}(\omega^2/2\pi)$$

where

$$\omega^2=E(u_0^2)+2\sum_{k=1}^{\infty}E(u_0u_k).$$

Thus, $2\pi f_y(0)=\omega^2/(1-a)^2$ and so the approximation (12) yields the correct asymptotic variance of $T^{-1/2}\Sigma_1^T y_t$ in the stationary case.

This rather remarkable deduction from the simple approximation (12) extends to the general case of vector processes. Here we find that when (2) is stationary and we set (using the principal value of the logarithm)

$$C=T\ln A\sim T(A-I),$$

the analogue of (11) is:

(13) $$V_T = T^2 V = T^2 C^{-1} \Omega C'^{-1} + O(T^{-1})$$
$$= (I-A)^{-1} \Omega (I-A')^{-1} + O(T^{-1}).$$

Part (b) of Lemma 3.1 and (8), (9), and (13) now suggest the approximation

(14) $$T^{-1/2} \sum_1^T y_t \sim N\left(0, (I-A)^{-1} \Omega (I-A')^{-1}\right) \equiv N\left(0, 2\pi f_{yy}(0)\right)$$

where $f_{yy}(\lambda) = (I - Ae^{i\lambda})^{-1} f_{uu}(\lambda)(I - A'e^{-i\lambda})^{-1}$ is the spectral density matrix of $y_t$ and $f_{uu}(\lambda)$ is the spectral density matrix of $u_t$. Thus, (14) gives for all (stable) $A$ the well known asymptotic result from the theory of stationary processes (see, for example, Hannan (1970, Theorem 11, p. 221)).

## 4 VECTOR AUTOREGRESSIONS WITH NEAR-INTEGRATED PROCESSES

Consider the least squares vector autoregression

(15) $$y_t = \hat{A} y_{t-1} + \hat{u}_t \qquad (t = 1, \ldots, T)$$

where

$$\hat{A} = Y'Y_{-1}(Y_{-1}'Y_{-1})^{-1}, \quad Y' = [y_1, \ldots, y_T], \quad Y_{-1}' = [y_0, \ldots, y_{T-1}].$$

The associated error covariance matrix estimator is:

$$\hat{\Omega}_0 = T^{-1} Y'(I - P_{Y_{-1}})Y,$$

where $P_D = D(D'D)^{-1}D'$ for any matrix $D$ of full column rank. The following theorem provides the asymptotic distribution theory for these least squares regression estimates when the time series is a near-integrated process.

THEOREM 4.1:

(a) $$T(\hat{A} - 1) \Rightarrow C + \left[\int_0^1 dB K_C' + \Omega_1'\right]\left[\int_0^1 K_C K_C'\right]^{-1};$$

(b) $$\hat{A} \underset{p}{\to} I, \qquad \hat{\Omega}_0 \underset{p}{\to} \Omega_0 = E(u_0 u_0');$$

(c) *if condition* $(B)$ *is strengthened to* $(B)'$ $E\|u_0\|^{2\beta} < \infty$ *for some* $\beta > 2$, *then* $\sqrt{T}\left(\hat{\Omega}_0 - \Omega_0\right) \Rightarrow N(0, W)$ *where*

$$W = P_D \left( \sum_{k=0}^{\infty} \Psi_k \right) P_D,$$
$$\Psi_0 = E(u_t u_t' \otimes u_t u_t') - \operatorname{vec}(\Omega_0)\operatorname{vec}(\Omega_0)',$$
$$\Psi_k = \Phi_k + \Phi_k' \qquad (k = 1, 2 \ldots),$$
$$\Phi_k = E(u_t u_{t+k}' \otimes u_t u_{t+k}') - \operatorname{vec}(\Omega_0)\operatorname{vec}(\Omega_0)',$$

*and $D$ is the* $n^2 \times n(n+1)/2$ *duplication matrix.*

Theorem 4.1 extends to near-integrated time series the theory developed in Phillips and Durlauf (1986) for integrated processes. In particular, when $C=0$, part (a) of Theorem 4.1 gives the main distributional result of their Theorem 3.2. When $C \neq 0$ part (a) of Theorem 4.1 shows the effect of near-integration on the asymptotic distribution of the regression coefficients. We see that this entails a shift in the location as well as the shape of the limiting distribution. We also note from part (b) of Theorem 4.1 that simple least squares regression continues to provide consistent estimates of $I_n$ (and hence the asymptotic unit roots of the model) in the presence of serially correlated innovations even when the time series are near-integrated. Part (c) gives the asymptotic distribution of the error covariance matrix estimator $\hat{\Omega}_0$. We observe that this distribution is independent of $C$ and is the same for integrated and near-integrated processes. This is explained by the fact that in both cases $\hat{A}-A \to_p 0$ so that the residuals $\hat{u}_t$ from the regression (16) are asymptotically weakly dependent and consistently estimate the innovation process $u_t$. Conventional normal asymptotics therefore apply in this case, as we would expect for stationary processes.

The above results may be extended to apply to vector autoregressions in the presence of cointegration. In such cases we need to relax condition (D) above and allow $\Omega$ to be singular (see Phillips (1986)). To fix ideas let $J_2$ be an $n \times k$ matrix of orthonormal cointegrating vectors. We shall suppose that $C$ is symmetric (cf. the final example of Section 2) and $CJ_2=0$. Then

$$J_2'y_t = J_2'y_{t-1} + J_2'u_t$$

and since $J_2'y_t$ is stationary we necessarily have $\Omega J_2 = 0$. Let $J_1$ be an $n \times (n-k)$ matrix for which $P=[J_1, J_2]$ is orthogonal. Defining $x_t = P'y_t$ we find that $x_t$ satisfies:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} J_1'AJ_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}$$

or

$$x_t = G_1 x_{1t-1} + G_2 x_{2t-1} + w_t \tag{16}$$

where

$$x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} J_1'y_t \\ J_2'y_t \end{bmatrix}$$

and

$$w_t = \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} = \begin{bmatrix} J_1'u_t \\ J_2'y_t \end{bmatrix}.$$

We deduce that

$$y_t = J_1 x_{1t} + J_2 w_{2t} \tag{17}$$

and $x_{1t}$ is a near-integrated time series of dimension $n-k$. Thus, (17) decomposes $y_t$ into a stationary component of dimension $k$ and near-integrated process of dimension $n-k$. This generalizes the common trend decomposition of Stock and Watson (1986).

THEOREM 4.2: *If $\{w_t\}_0^\infty$ is strictly stationary and ergodic with nonsingular covariance matrix $E(w_0 w_0')$ and satisfies conditions $(A)$, $(B)'$, $(C)$, and $(D)$, then*

$$\hat{A} \underset{p}{\rightarrow} P\begin{bmatrix} I_{n-k} & \vdots & \bar{G}_2 \\ 0 & \vdots & \end{bmatrix} P' = \bar{A}, \quad say, \tag{18}$$

$$T(\hat{A}-\bar{A})J_1 \Rightarrow J_1\bar{C} + P\left\{E\int_0^1 d\bar{B}K_{\bar{C}}' + F\right\}\left\{\int_0^1 K_{\bar{C}}K_{\bar{C}}'\right\}^{-1}, \tag{19}$$

$$\sqrt{T}(\hat{A}-\bar{A})J_2 \Rightarrow N\big(0, (P\otimes I)\bar{Q}(P'\otimes I)\big), \tag{20}$$

$$\sqrt{T}(\hat{A}-\bar{A}) \Rightarrow N\big(0, (P\otimes J_2)\bar{Q}(P'\otimes J_2')\big), \tag{21}$$

*where $\bar{B}(r) = (B_1(r)'B_2(r)')'$ is n-vector Brownian motion with covariance matrix $\bar{\Omega} = 2\pi f_{ww}(0)$, $K_{\bar{C}}(r)$ is an $(n-k)$-vector diffusion defined by*

$$K_{\bar{C}}(r) = \int_0^r e^{(r-s)\bar{C}}\, d\bar{B}_1(s), \qquad \bar{C} = J_1'CJ_1,$$

*and $E$, $F$, and $\bar{Q}$ are constant matrices defined by* (A3)–(A5) *in the Appendix and*

$$\bar{G}_2 = E(w_t w_{2t-1}')\{E(w_{2t}w_{2t}')\}^{-1}.$$

We see from (19) and (20) that some linear combinations of the columns of $\hat{A}$ are $O(\sqrt{T})$-consistent and have an asymptotic normal distribution while others are $O(T)$-consistent and have a nonnormal limit distribution. The results closely parallel those obtained in Phillips and Ouliaris (1986) for the case where $A = I$ in (2). In particular, we observe from (21) that $\hat{A}$ has a singular asymptotic normal distribution in the limit (since $\bar{Q}$ is singular) and this distribution does not depend on the noncentrality matrix $C$. In fact, the limiting distribution of $\sqrt{T}(\hat{A}-\bar{A})$ given by (21) is the same for all $C$ and thus the presence of cointegration in a system such as (2) eliminates the differences between integrated and near-integrated processes. This is because cointegration in the regressors induces $O(\sqrt{T})$-consistency as we have seen and the effects of near-integration are of a smaller order (by construction).

## 5 POWER FUNCTIONS FOR UNIT ROOT TESTS

The theory of the preceding section may be used to derive asymptotic power functions for regression based tests for unit roots. To illustrate what is involved we use the framework developed recently by Park and Phillips (1986). This framework allows for multivariate regressions with deterministic regressors as well as I(1) processes and it also accommodates I(1) processes with drift. Park and Phillips propose a general class of statistics (called $H$-statistics) which in the present context are useful for testing the null hypothesis $H_0$: $A = I$ in (2). More specifically, from the vector autoregression (15) the suggested statistic (given by equation (38) of Park and Phillips (1986)) is:

$$H(\hat{A}) = \operatorname{tr}\{\tilde{\Omega}^{-1}(\hat{A}-I)M_T(\hat{A}-I)'\} - 2T\operatorname{tr}\{\tilde{\Omega}^{-1}(\hat{A}-I)\tilde{\Omega}_1\} + T^2\operatorname{tr}(\tilde{\Omega}^{-1}\tilde{\Omega}_1 M_t^{-1}\tilde{\Omega}_1) \tag{22}$$

where $\tilde{\Omega}$ and $\tilde{\Omega}_1$ are consistent estimates of $\Omega$ and $\Omega_1$ respectively and $M_T = Y'_{-1}Y_{-1}$. Under the null hypothesis $H_0$ we have

$$H(\hat{A}) \Rightarrow \int_0^1\int_0^1 dB(r)'\Omega^{-1}dB(s)B(r)'\left(\int_0^1 BB'\right)^{-1}B(s) \equiv \int_0^1\int_0^1 dW(r)'dW(s)W(r)'\left(\int_0^1 WW'\right)^{-1}W(s) \tag{23}$$

where $W(r)$ is $n$-vector standard Brownian motion. When $n=1$ (23) reduces to

$$\frac{\left(\int_0^1 W\,dW\right)^2}{\int_0^1 W^2} = \left\{\frac{\tfrac{1}{2}\left(W(1)^2-1\right)}{\left(\int_0^1 W^2\right)^{1/2}}\right\}^2$$

which is the square of the limiting distribution of the $t$ ratio statistic in an AR(1) with a single unit root and with iid $(0,\omega^2)$ errors. The latter distribution is tabulated in Fuller (1976). Note also that in this scalar case $H(\hat{A})$ in (22) reduces to the square of the statistic $Z_t$ (the modified $t$-ratio statistic) introduced in Phillips (1987a).

Under the alternative hypothesis given by (3) the limiting distribution of $H(\hat{A})$ may be obtained from the results of Theorem 4.1. We find:

$$\begin{aligned} H(\hat{A}) &\Rightarrow \operatorname{tr}\left\{\Omega^{-1}C\left(\int_0^1 K_C K_C'\right)C'\right\} + 2\operatorname{tr}\left\{\Omega^{-1}C\left(\int_0^1 K_C\,dB'\right)\right\} \\ &\quad + \operatorname{tr}\left\{\Omega^{-1}\left(\int_0^1 dB\,K_C'\right)\left(\int_0^1 K_C K_C'\right)^{-1}\left(\int_0^1 K_C\,dB'\right)\right\} \\ &= \operatorname{tr}\left\{\Omega^{-1}(C+\xi)\left(\int_0^1 K_C K_C'\right)(C+\xi)'\right\} \end{aligned} \tag{24}$$

where

$$\xi = \left(\int_0^1 dB\,K_C'\right)\left(\int_0^1 K_C K_C'\right)^{-1}.$$

Now consider the case where the innovation sequence $\{u_t\}$ is iid $(0,\Omega_0)$. In this case $\Omega_1 = 0$ and $\Omega = \Omega_0$. The conventional Wald statistic for testing $H_0$ is:

$$F = \operatorname{tr}\{\hat{\Omega}_0^{-1}(\hat{A}-I)M_T(\hat{A}-I)'\}.$$

Once again we find that

$$F \Rightarrow \operatorname{tr}\left\{\Omega^{-1}(C+\xi)\left(\int_0^1 K_C K_C'\right)(C+\xi)'\right\}.$$

Thus, there is no loss in asymptotic local power from the use of $H(\hat{A})$ even though this statistic is applicable for a wide range of possible innovation processes. Similar results hold for tests of a single unit root (see Phillips (1987b) and Phillips and Perron (1986)).

Phillips and Durlauf (1986, equation (32)) suggested an alternative test of $H_0$ in the multivariate case based on the statistic:

$$G = T^{3/2}\,\mathrm{tr}\left\{(\hat{A}-I)'(\hat{A}-I)\right\} + T^{-1}y_T'\tilde{\Omega}^{-1}y_T$$

where $\tilde{\Omega}$ is a consistent estimate of $\Omega$. Under the null $G \Rightarrow \chi_n^2$ so that tests based on $G$ have the advantage of relying only on conventional tables of the chi squared distribution. Under the local alternative hypothesis (3) we now find that

$$G \Rightarrow K_C(1)'\Omega^{-1}K_C(1) \tag{25}$$

which is a quadratic form in the normal vector

$$K_C(1) \equiv N\left(0, \int_0^1 e^{(1-s)C}\Omega e^{(1-s)C'}\,ds\right).$$

It is interesting to compare the asymptotic behavior of $H(\hat{A})$ and $G$ under local alternatives. Note first that $K_C(1)$ has zero mean so that the limiting distribution of $G$ is a weighted sum of independent central $\chi_1^2$ variates. The limiting distribution of $H(\hat{A})$, on the other hand, is a random quadratic form in the elements of the matrix $C+\xi$. The distribution of this matrix involves a shift in location under the alternative hypothesis that is directly related to the magnitude of the noncentrality matrix $C$. From these observations we may expect the asymptotic local power of $H(\hat{A})$ to be superior to that of $G$.

The poor power properties of the $G$ test are confirmed by closer examination of the scalar case. Here (25) becomes (setting $C=c$):

$$G \Rightarrow \left\{(e^{2c}-1)/2c\right\}\chi_1^2$$

which for small $c$ behaves like $(1+c)\chi_1^2$. Thus, against stationary local alternatives (with $c<0$) $G$ has asymptotic local power less than the size of the test (the latter being delivered by $\chi_1^2$ under the null). Moreover, $c \downarrow -\infty$ we see that

$$\left\{(e^{2c}-1)/2c\right\}\chi_1^2 \underset{p}{\rightarrow} 0.$$

Thus, asymptotic local power tends to zero as $c \downarrow -\infty$ for the $G$ test.

By contrast, in the scalar case (24) reduces to

$$\left(c\left(\int_0^1 J_c^2\right)^{1/2} + \left(\int_0^1 J_c^2\right)^{-1/2}\int_0^1 J_c\,dW\right)^2 \tag{26}$$

where $J_c(r) = \int_0^r e^{(r-s)c}\,dW(s)$ and $W(s)$ is standard Brownian motion. By Lemma 2 of Phillips (1987b) we have

$$(-2c)\int_0^1 J_c^2 \underset{p}{\rightarrow} 1, \qquad (-2c)^{1/2}\int_0^1 J_c\,dW \Rightarrow N(0,1)$$

as $c \downarrow -\infty$. We deduce that (26) diverges to $+\infty$ as $c \downarrow -\infty$. Thus, asymptotic local power tends to unity as $c \downarrow -\infty$ for the $H$ test.

## 6. EXTENSIONS TO MODELS WITH DRIFT

The theory developed in earlier sections may be extended to allow for models with near unit roots and nonzero drift. In this case we replace (2) by

$$(2)' \qquad y_t = \mu + Ay_{t-1} + u_t.$$

When $A = I$ in (2)′ the asymptotic theory has been fully developed recently by Park and Phillips (1986). Note that in this special case of (2)′ we may write

$$(27) \qquad y_t = \mu t + y_t^0$$

where $y_t^0$ is a (driftless) $I(1)$ process satisfying (2) with $A = I$. Under (27) $y_t$ behaves asymptotically as if it were $\mu t$ and we therefore find asymptotic behavior rather different from that which obtains for $y_t^0$. In particular, we find that:

$$(28) \qquad T^{-3}\sum_1^T y_t y_t' \underset{p}{\to} (1/3)\mu\mu',$$

so that the second moment matrix converges in probability to a constant matrix, in contrast to Lemma 3.1(c). Note also that the standardization in (28) is $T^{-3}$ (rather than $T^{-2}$) and the limit matrix is singular when $n > 1$.

For near-integrated time series two major cases can be distinguished. In the first, we replace (3) by

$$(3)' \qquad A = \exp(T^{-3/2}C) \sim I + T^{-3/2}C$$

where the local alternatives are $O(T^{-3/2})$, a choice inspired by the standardization factor $T^{-3}$ in (28). In place of Lemma 3.1(a–c) we now find:

$$(29) \qquad T^{-1}y_{[Tr]} \underset{p}{\to} r\mu;$$

$$(30) \qquad T^{-2}\sum_1^T y_t \underset{p}{\to} (1/2)\mu;$$

$$(31) \qquad T^{-3}\sum_1^T y_t y_t' \underset{p}{\to} (1/3)\mu\mu'.$$

The asymptotic regression theory is complicated by the singularity of the limiting sample second moment matrix (31). Park and Phillips (1986) show how to deal with this complication and their results apply directly in the present case. In particular, let $\bar{\mu}$ and $\bar{A}$ be the least squares regression coefficients from (2)′. Define $h_1 = \mu/(\mu'\mu)^{1/2}$ and let $H = [h_1, H_2]$ be an orthogonal matrix of dimension $n \times n$. We further define $\underline{B} = H_2'B$ and $\underline{\Omega}_1 = H_2'\Omega_1$ and we use the following functional introduced by Park and Phillips (1986):

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

where $B$ is vector Brownian motion, $M$ is a process with continuous sample paths such that $\int_0^1 MM' > 0$ a.s., and $E$ is a (possibly random) matrix of conformable dimension. As in Theorem 3.6 of Park and Phillips (1986) we find that:

(32) $$T(\bar{A} - A) \Rightarrow f(B, \underline{B}^*, \underline{\Omega}_1)H_2',$$

(33) $$T^{3/2}(\bar{A} - A)h_1 \Rightarrow (\mu'\mu)^{-1/2} f(B, \underline{P}, \underline{\delta}).$$

Here $\underline{B}^*$ and $\underline{P}$ may be interpreted as Hilbert space projections. Specifically, let $m = n - 1$ and consider the Hilbert space $L_2[0,1]^m$ of $m$-vector valued, continuous, square integrable functions on the [0,1] interval with inner product $\int_0^1 g_1' g_2$ for $g_1, g_2 \in L_2[0,1]^m$. Define the functions $1(r) = 1$, $2(r) = r$ for $r \in [0,1]$. Then $\underline{B}^*$ is the projection of $\underline{B}$ onto the orthogonal complement of the subspace spanned by $[1(r)I_m, 2(r)I_m]$. Similarly, $\underline{P}$ is the projection of $2(r)$ on the orthogonal complement of the subspace of $L_2[0,1]$ that is spanned by $[1(r), B'(r)]$. Finally $\underline{\delta}$ in (33) is defined by

$$\underline{\delta} = -\left[\left(\int_0^1 s\underline{B}' - \tfrac{1}{2}\int_0^1 \underline{B}'\right)\left(\int_0^1 \underline{B}\,\underline{B}' - \int_0^1 \underline{B}\int_0^1 \underline{B}'\right)^{-1}\right]\underline{\Omega}_1$$

as in Park and Phillips (1986, Theorems 3.3 and 3.6).

We see from (32) that the limiting distribution of $T(\bar{A} - A)$ is degenerate and its support in $R^{n^2}$ is the range space of $I_n \otimes H_2$. The degenerate linear combination of $\bar{A} - A$ at $O(T)$ scaling involves the vector $(\bar{A} - A)h_1$. From (3)′ and (33) we have:

$$T^{3/2}(\bar{A} - I)h_1 \Rightarrow Ch_1 + (\mu'\mu)^{1/2} f(\underline{B}, \underline{P}, \underline{\delta});$$

whereas from (3)′ and (32) we obtain:

$$T(\bar{A} - I) \Rightarrow f(B, \underline{B}^*, \underline{\Omega}_1)H_2'.$$

Thus, the noncentralities induced by the specification (3)′ influence only those linear combinations of $\bar{A}$, viz. $\bar{A}h_1$, which are $O(T^{-3/2})$ consistent. All other linear combinations of $\bar{A}$ have limiting distributions which are invariant to the noncentrality matrix $C$.

The situation is substantially different when alternatives to $A = I$ take the form given in (3) rather than (3)′. This is the second major case of interest for near-integrated processes. In place of Lemma 3.1(a)–(c) and (29)–(31) we now find:

(34) $$T^{-1} y_{[Tr]} \underset{p}{\to} L_C(r)\mu;$$

(35) $$T^{-2}\sum_1^T y_t \underset{p}{\to} \left(\int_0^1 L_C\right)\mu;$$

(36) $$T^{-3}\sum_1^T y_t y_t' \underset{p}{\to} \int_0^1 L_C(r)\mu\mu' L_C(r)'\,dr = \Sigma(C, \mu), \quad \text{say},$$

where

$$L_C(r) = rI + (1/2!)r^2C + (1/3!)r^3C^2 + \cdots .$$

The limiting sample second moment matrix $\Sigma(C, \mu)$ may have rank equal to any integer from zero to $n$. Let $M$ be the subspace of $R^n$ spanned by the vectors $\{C^k\mu\colon k = 0, 1, 2, \dots\}$ and set $l = \dim m$. Then $0 \leqslant l \leqslant n$ and $\operatorname{rank}\{\Sigma(C, \mu)\} = l$, as in the proof of Theorem 1 of Phillips (1974). Let $H_1$ be an $n \times l$ matrix of orthonormal vectors of $M$ and let $H = [H_1, H_2]$ be an orthogonal matrix. Define the $l$-vector of continuous functions $g(r)$ by the equation

$$L_C(r)\mu = H_1 g(r)$$

and set

$$\underline{K}_C(r) = H_2' K_C(r), \qquad \underline{\Omega}_1 = H_2'\Omega_1.$$

The least squares regression coefficient matrix in (2)′ now has limiting distributions given by:

$$T(\bar{A} - A) \Rightarrow f\left(B, \underline{K}_C^*, \underline{\Omega}_1\right)H_2'; \tag{37}$$

$$T^{3/2}(\bar{A} - A)H_1 \Rightarrow f(B, \underline{P}^+, \underline{\delta}^+). \tag{38}$$

It is again convenient to interpret $K_C^*$ and $P^+$ in these functionals as Hilbert space projections. Thus, $\underline{K}_C^*$ is the projection of $\underline{K}_C$ onto the orthogonal complement in $L_2[0,1]^{n-l}$ of the subspace spanned by the functions $[1(r)I_{n-l}, I_{n-l} \otimes g(r)']$; and $\underline{P}^+$ is the projection of $g(r)$ onto the orthogonal complement in $L_2[0,1]^l$ of the subspace spanned by $[1(r)I_l, I_l \otimes \underline{K}_C']$. $\underline{\delta}^+$ in (38) is defined by

$$\underline{\delta}^+ = -\left[\left(\int_0^1 g\underline{K}_C' - \int_0^1 g \int_0^1 \underline{K}_C'\right)\left(\int_0^1 \underline{K}_C\underline{K}_C' - \int_0^1 \underline{K}_C \int_0^1 \underline{K}_C'\right)^{-1}\right]\underline{\Omega}_1.$$

From (37) we see that the limiting distribution of $T(\bar{A} - A)$ is again degenerate in $R^{n^2}$, with support equal to the range of $I \otimes H_2$. The degenerate elements of $\bar{A} - A$ at $O(T)$ scaling now involve the $l$ vectors of $(\bar{A} - A)H_1$. In contrast to (32), we observe that the limiting distribution of $T(\bar{A} - A)$ does depend on the noncentrality matrix. This is to be expected since the alternatives given by (3) involve $O(T^{-1})$ departures from the null. Indeed

$$T(\bar{A} - I) \Rightarrow C + f(B, \underline{K}_C^*, \underline{\Omega}_1)H_2' \tag{39}$$

and this provides a generalization of Theorem 4.1(a) to models with drift. Note, in particular, that when $\mu = 0$ we have $l = 0$, $H_2 = I_n$, $\underline{\Omega}_1 = \Omega_1$, $\underline{K}_C = K_C$, $\underline{K}_C^* = K_C - \int_0^1 K_C$ and (39) extends Theorem 4.1(a) to the special case of fitted drift with $\mu = 0$.

At the other extreme, when $\mu \neq 0$ and $l = n$, we have $H_1 = I_n$, $\underline{K}_C = 0$, $\underline{\Omega}_1 = 0$, $\underline{\delta}^+ = 0$, and $\underline{P}^+$ is the projection of $g(r)$ on the orthogonal complement of $i(r)I_n$

in $L_2[0,1]^n$. Thus $\underline{P}^+ = g - \int_0^1 g$ and (38) yields

$$(40)\qquad T^{3/2}(\bar{A}-A) \Rightarrow f(B, \underline{P}^+, 0) = \left(\int_0^1 dB\underline{P}^{+\prime}\right)\left(\int_0^1 \underline{P}^+\underline{P}^{+\prime}\right)^{-1}$$

$$\equiv N\left(0, \Omega \otimes \left(\int_0^1 gg' - \int_0^1 g \int_0^1 g'\right)^{-1}\right).$$

Note that in this case

$$\Sigma(C,\mu) = \int_0^1 gg'$$

is positive definite, since $l = n$ and, consequently, $M = R^n$. Thus, from (36), the sample moment matrix $T^{-3}Y_{-1}'Y_{-1}$ has a constant, positive definite probability limit and, as we might have expected in such a case, the limiting distribution is normal. Note also that there are no degeneracies in the limiting distribution (40). This extreme case with $\mu \neq 0$ and $l = n$ therefore represents a return to conventional normal asymptotics. The result is explained by the fact that the behavior of $y_t$ is dominated by deterministic components (viz. $Tg(t/T)$) which induce sufficient asymptotic variation over component variates to ensure that the limit of the sample moment matrix of the regressors in (2)′ is constant and nonsingular.

The general case given in (37) and (38) admits the two extremes we have just discussed as well as intermediate cases in which both normal and nonnormal asymptotics apply. These results therefore help to bridge the apparent gap between the nonnormal asymptotics explored in Phillips (1987a), Phillips and Durlauf (1986), and Park and Phillips (1986) and the normal asymptotics obtained in Kramer (1984) and West (1986), the latter for the special case of a single nonstationary regressor with drift.

## 7 MULTIPLE REGRESSION WITH NEAR-INTEGRATED TIME SERIES

The theory developed in Sections 3 and 6 may be applied to multiple (least squares) regressions of the form

$$(41)\qquad x_t = \hat{\alpha} + \hat{\beta}' z_t + \hat{v}_t,$$

where $x_t$ (a scalar) and $z_t$ (an $m$-vector) are quite general near-integrated processes. For our analysis it will be convenient to set $n = m+1$, to define $y_t' = (x_t, z_t')$, and to assume, at first, that the multiple time series $\{y_t\}_0^\infty$ is generated by (2) and (3) with innovations $\{u_t\}_0^\infty$ that satisfy Conditions (A)–(D). Under this set up, some elements of $y_t$ may be I(1) processes, others may be near-integrated; the innovations $u_t$ that drive (2) may be quite general weakly dependent time series; and $x_t$ and $z_t$ may be both contemporaneously and serially correlated.

The following result provides an asymptotic theory for the least squares regression (41). It is a simple consequence of earlier results in Section 3. In the statement of the Theorem we use $F_\beta$ to represent the customary regression $F$ statistic for testing the significance of $\hat{\beta}$ in (41); $t_{\beta_i}$ denotes the conventional $t$

statistic for assessing the significance of $\beta_i$; $R^2$ is the coefficient of determination in the regression; and $DW$ is the Durbin-Watson statistic.

THEOREM 7.1:

(a) $\hat{\beta} \Rightarrow G_{22}^{-1} g_{21}$;

(b) $T^{-1/2}\hat{\alpha} \Rightarrow b'\eta$;

(c) $R^2 \Rightarrow g_{21}' G_{22}^{-1} g_{21} / g_{11}$;

(d) $T^{-1}F_\beta \Rightarrow (1/m) g_{21}' G_{22}^{-1} g_{21} \big/ \left( g_{11} - g_{21}' G_{22}^{-1} g_{21} \right)$;

(e) $T^{-1/2} t_{\beta_i} \Rightarrow \left\{ \left( g_{11} - g_{21}' G_{22}^{-1} g_{21} \right) \left[ G_{22}^{-1} \right]_{ii} \right\}^{-1/2} \left( G_{22}^{-1} g_{21} \right)_i$;

(f) $T(DW) \Rightarrow \eta' \Omega_0 \eta \big/ \left( g_{11} - g_{21}' G_{22}^{-1} g_{21} \right)$;

*where*

$$G = \overset{\quad 1 \qquad m}{\begin{bmatrix} g_{11} & g_{21}' \\ g_{21} & G_{22} \end{bmatrix}} \begin{matrix} 1 \\ m \end{matrix} = \int_0^1 K_C K_C' - \left( \int_0^1 K_C \right) \left( \int_0^1 K_C' \right); \tag{42}$$

$$b = \int_0^1 K_C;$$

$$\eta' = \left( 1, -g_{21}' G_{22}^{-1} \right).$$

Theorem 7.1 generalizes to near-integrated processes the regression theory derived in Phillips (1986) for spurious regressions with I(1) processes. All of the main qualitative results of the regression theory of the latter paper also apply in the context of near-integrated processes. Thus, unlike the theory of regression for stationary processes, the regression coefficients $\hat{\alpha}$ and $\hat{\beta}$ do not converge to constants as $T \uparrow \infty$; $\hat{\beta}$ has a nondegenerate limiting distribution; and the distribution of $\hat{\alpha}$ diverges as $T \uparrow \infty$. Similarly, $R^2$ has a nondegenerate limiting distribution. On the other hand, the distributions of the test statistics $F$ and $t_{\beta_i}$ both diverge as $T \uparrow \infty$ and $DW \to_p 0$ as $T \uparrow \infty$.

Equation (41) may be regarded as a cointegrating regression of the type recently considered by Engle and Granger (1987). In the work of these authors, the null hypothesis in the regression is that of no cointegration (i.e. no linear combination of $x_t$ and $z_t$ is stationary). Their maintained hypothesis is that all of the variables in the regression (here, $x_t$ and $z_t$) are integrated processes. When $C = 0$, Theorem 7.1 gives the asymptotic theory for the regression coefficients, conventional significance tests, and regression diagnostics under the Granger-Engle null hypothesis in such a cointegrating regression. When $C \neq 0$ the theorem delivers the relevant asymptotic theory for the wider class of near-integrated processes. That is, the asymptotic theory is established for a more general maintained hypothesis under which some variables in the regression may be I(1) processes, others may be nearly explosive, while yet others may be nearly stationary. The effects of these extensions are measured through the noncentrality matrix $C$.

Under the alternative hypothesis that the variables in the regression are cointegrated a different asymptotic theory applies. Phillips and Durlauf (1986) developed the relevant asymptotic theory for regressions such as (41) when $z_t$ is an I(1) process and $(x_t, z_t')$ are cointegrated. This theory is easily extended to the case where $z_t$ is near-integrated. Specifically, suppose $\{x_t\}$ is generated by:

(43) $$x_t = \beta' z_t + v_t$$

where

(44) $$z_t = F z_{t-1} + w_t, \qquad F = \exp\left(\frac{1}{T}R\right).$$

Let $u_t' = (v_t, w_t')$ and assume that $\{u_t\}_0^\infty$ satisfies conditions (A)–(D). Then, in place of Theorem 6.1(a) we find $\beta \underset{p}{\rightarrow} \beta$ as $T \uparrow \infty$ and

(45) $$T(\hat{\beta} - \beta) \Rightarrow \left(\int_0^1 K_R K_R'\right)^{-1}\left(\int_0^1 K_R\, dB + \lambda\right)$$

where

$$K_R = \int_0^r e^{(r-s)R}\, dB_2(s), \qquad \lambda = \sum_{k=0}^{\infty} (w_0 v_k), \quad \text{and}$$

$$B'(r) = [\overset{1}{B_1}(r), \overset{m}{B_2}(r)']$$

is $n$-vector Brownian motion. (45) extends Theorem 4.1(a) of Phillips and Durlauf (1986). The two results are very similar and they differ only by the presence of the diffusion process $K_R$ rather than vector Brownian motion in the limiting distribution.

The asymptotic theory of cointegrating regressions may be further extended to cases which allow for drift as well as near integration. This leads to new and rather different results which make use of the theory developed in Section 6. Suppose that the regressors in (43) are generated by

(44)′ $$z_t = m + F z_{t-1} + w_t, \qquad F = \exp\left(\frac{1}{T}R\right)$$

in place of (44). As in (36) we now find that

$$T^{-3}\sum_1^T z_t z_t' \underset{p}{\rightarrow} \int_0^1 L_R(r) m m' L_R(r)'\, dr = \Sigma(R, m).$$

Let $M_R$ be the subspace of $R^m$ spanned by $\{R^k\colon\ k = 0, 1, 2, \ldots, m\}$ and set $l = \dim(M_R)$. Then, as before, $0 \leqslant l \leqslant m$ and rank$\{\Sigma(R, m)\} = l$. Once again, let $H_1$ be an $m \times l$ matrix of orthonormal vectors of $M_R$ and let $H = [H_1, H_2]$ be orthogonal. We define the $l$-vector $g(r)$ by $L_R(r)m = H_1 g(r)$ and write $\underline{K}_R = H_2' K_R$, $\underline{\lambda} = H_2'\lambda$. The least squares regression coefficient $\hat{\beta}$ from the cointegrating regression (43) now has asymptotic distributions which we characterize as follows:

$$T(\hat{\beta}-\beta) \Rightarrow H_2 f(B_1, \underline{K}_R^*, \underline{\lambda})', \tag{46}$$

$$T^{3/2} H_1'(\hat{\beta}-\beta) \Rightarrow f(B_1, g^+, \underline{\lambda}^+)'. \tag{47}$$

Here $\underline{K}_R^*$ is the projection of $\underline{K}_R$ onto the orthogonal complement in $L_2[0,1]^{m-l}$ of the subspace spanned by $I_{m-l} \otimes g(r)'$; $g^+$ is the projection of $g$ onto the orthogonal complement in $L_2[0,1]^l$ of the space spanned by $I_l \otimes \underline{K}_R'(r)$; and

$$\underline{\lambda}^+ = -\left(\int_0^1 g\underline{K}_R\right)\left(\int_0^1 \underline{K}_R \underline{K}_R'\right)^{-1} \underline{\lambda}.$$

Note that when $m=0$ (and hence $l=0$) $H_2 = I_m$, $\underline{K}_R^* = \underline{K}_R = K_R$, $\underline{\lambda} = \lambda$, and (46) reduces to the earlier result (45). At the other extreme, when $m \neq 0$ and $l = n$ we have:

$$T^{3/2}(\hat{\beta}-\beta) \Rightarrow \left(\int_0^1 gg'\right)^{-1} \int_0^1 g\, dB_1$$
$$\equiv N\left(0, \omega_{11}^2 \left(\int_0^1 gg'\right)^{-1}\right)$$

where

$$\omega_{11}^2 = E(v_0^2) + 2\sum_{k=1}^{\infty} E(v_0 v_k)$$

is the variance of the Brownian motion $B_1$.

## 8 CONCLUSION

This paper develops a general asymptotic theory of regression for multiple time series which may be individually characterized as either integrated or near-integrated processes. The limiting distribution theory that we have derived covers vector autoregressions and cointegrating regressions with near-integrated processes. In both cases the asymptotic theory presents some important general departures from conventional theory based on stationary processes. The new asymptotic theory is helpful in characterizing large sample behavior in such regressions whether there are unit roots or near-unit roots in the underlying data generating mechanisms.

The theory we have developed has been applied to analyze the noncentral distributions of certain multivariate tests for unit roots. The results provide some helpful asymptotic local power comparisons among the tests. In particular, they indicate that the $H$ tests introduced by Park and Phillips (1986) involve no loss in asymptotic power over more conventional Wald tests, in spite of the fact that the new tests allow for a wide class of weakly dependent innovation processes, whereas the Wald tests apply strictly for iid innovations only.

Our regression theory includes cases of vector autoregressions with cointegrated regressors. We have also studied cases where the generating mechanism allows for drift as well as near integration. Both these cases lead, in general, to degeneracies in the asymptotic distribution theory. The case of drift and near-integration is particularly interesting because the extent of the degeneracy is contingent on the noncentrality matrix and the drift coefficient. The results in this case provide a bridge between the nonnormal asymptotic theory developed in Phillips and Durlauf (1986) for integrated regressors and more conventional normal asymptotics for regressions with deterministic regressors.

*Department of Economics, Yale University, New Haven, CT 06520, U.S.A.*

*Manuscript received February, 1986; final revision received August, 1987.*

## APPENDIX

PROOF OF LEMMA 3.1: The proofs of (a)–(d) follow lines developed earlier in Phillips (1987a, 1987b). To illustrate we outline here the arguments leading to part (b). From (2) and (3) we deduce the representation:

$$\text{(A1)} \qquad y_t = \sum_{k=0}^{t-1} \exp\{(k/T)C\} u_{t-k} + \exp\{(t/T)C\} y_0$$
$$= \sum_{j=1}^{t} \exp\{((t-j)/T)C\} u_j + \exp\{(t/T)C\} y_0$$

Thus,

$$\text{(A2)} \qquad T^{-3/2}\sum_1^T y_t = T^{-3/2}\sum_{t=1}^{T}\sum_{j=1}^{t} \exp\{((t-j)/T)C\} u_j + T^{-3/2}\sum_{t=1}^{T}\exp\{(t/T)C\} y_0$$
$$= T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t} \exp\{((t-j)/T)C\}\Sigma^{1/2}\int_{(j-1)/T}^{j/T} dX_T(s) + O_p(T^{-1/2})$$
$$= \sum_{t=1}^{T}\sum_{j=1}^{t} \int_{(t-1)/T}^{t/T} dr \int_{(j-1)/T}^{j/T} \exp\{((t-j)/T)C\}\Sigma^{1/2} dX_T(s)$$
$$+ O_p(T^{-1/2}).$$

Now $(t-1)/T \leqslant r \leqslant t/T$ and $(j-1)/T \leqslant s \leqslant j/t$, so that

$$\exp\{((t-j)/T)C\} = \exp\{(r-s)C\}[1 + O(T^{-1})]$$

and (A2) becomes:

$$\sum_{t=1}^{T}\sum_{j=1}^{t} \int_{(t-1)/T}^{t/T} dr \int_{(j-1)/T}^{j/T} \exp\{(r-s)C\}\Sigma^{1/2} dX_T(s) + O_p(T^{-1/2})$$
$$= \int_0^1 dr \int_0^r \exp\{(r-s)C\}\Sigma^{1/2} dX_T(s) + O_p(T^{-1/2})$$
$$\Rightarrow \int_0^1 K_C(r)\, dr$$

by the continuous mapping theorem (since $\exp\{(r-s)C\}$ is continuous) and (1), proving (b). The proofs of (a), (c), and (d) are entirely analogous Part (e) is proved in Phillips (1986c)

PROOF OF THEOREM 4.1: Define $U' = [u_1, \ldots, u_T]$ and then from (2)

$$\hat{A} = A + U'Y_{-1}(Y_{-1}'Y_{-1})^{-1}$$

so that

$$T(\hat{A} - A) = (T^{-1}U'Y_{-1})(T^{-2}Y_{-1}'Y_{-1})^{-1}$$

Now $A = \exp\{(1/T)C\} = I_n + (1/T)C + O(T^{-2})$ and from Lemma 3 1 and the continuous mapping theorem we deduce that as $T \uparrow \infty$:

$$T(\hat{A} - I) \Rightarrow C + \left[\int_0^1 dB\, K_C' + \Omega_1'\right]\left[\int_0^1 K_C K_C'\right]^{-1}$$

as required for part (a) of the Theorem. The first part of (b) follows directly. To prove the second part of (b) we now show that as $T \uparrow \infty$

$$\hat{\Omega}_0 = T^{-1}U'U - T^{-1}U'Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}'U \underset{p}{\rightarrow} \Omega_0$$

as required, since the second term in the above expression is $O_p(T^{-1})$ and the first term converges to $\Omega_0$ almost surely as $T \uparrow \infty$ by the McLeish strong law for dependent sequences. The proof of part (c) is identical to the proof of Theorem 3.3 of Phillips and Durlauf (1986)

PROOF OF THEOREM 4.2: Let $\hat{G}_1$ and $\hat{G}_2$ be the least squares regression coefficients from (16) Using Lemma 3.1 we find that

$$T(\hat{G}_1 - G_1) \Rightarrow \left\{E\int_0^1 d\bar{B}\, K_{\bar{C}}' + F\right\}\left\{\int_0^1 K_{\bar{C}} K_{\bar{C}}'\right\}^{-1},$$

$$\sqrt{T}(\hat{G}_2 - \bar{G}_2) \Rightarrow N(0, \bar{Q}), \quad \text{where}$$

$$\bar{B}(r) = (\underset{(n-k)}{\bar{B}_1'(r)}, \underset{(k)}{\bar{B}_2'(r)})'$$

is $n$-vector Brownian motion with covariance matrix $\bar{\Omega} = 2\pi f_{ww}(0)$ and

(A3) $$K_{\bar{C}}(r) = \int_0^r e^{(r-s)\bar{C}}\, dB_1(s), \quad \bar{C} = J_1'CJ_1,$$

$$E = I_n - E_* F_*^{-1}[0, I_k],$$

$$E_* = E(w_t w_{2t-1}'),$$

$$F_* = E(w_{2t} w_{2t}'),$$

(A4) $$F = \begin{bmatrix} \Omega_{11}^{(1)\prime} \\ \Omega_{12}^{(1)\prime} \end{bmatrix} - E_* F_*^{-1} \Omega_{12}^{(1)\prime},$$

$$\bar{\Omega}^{(1)} = \sum_{k=1}^{\infty} E(w_0 w_k') = \begin{matrix} \\ (n-k) \\ (k) \end{matrix}\overset{\qquad (n-k) \quad (k)}{\begin{bmatrix} \Omega_{11}^{(1)} & \vdots & \Omega_{12}^{(1)} \\ \Omega_{21}^{(1)} & \vdots & \Omega_{22}^{(1)} \end{bmatrix}},$$

(A5) $$\bar{Q} = HQ_*H',$$

$$H = [I \otimes F_*^{-1}, E_* F_*^{-1} \otimes F_*^{-1}],$$

$$Q_* = M_D(2\pi f_{\xi\xi}(0))M_D,$$

$$M_D = \begin{bmatrix} I_{n^2} & 0 \\ 0 & P_D \end{bmatrix}, \qquad P_D = D(D'D)^{-1}D',$$

$$\xi_t = \begin{bmatrix} w_t \otimes w_{t-1} - \operatorname{vec}(E_*) \\ w_t \otimes w_t - \operatorname{vec}(F_*) \end{bmatrix},$$

and $D$ is the duplication matrix of order $n^2 \times n(n+1)/2$

Now $y_t = Px_t$ so that $\hat{A} = P\hat{G}P'$ and we deduce that

$$\hat{A} \underset{p}{\rightarrow} P[\bar{G}_1, \bar{G}_2]P' = P\left[\begin{array}{c:c} J_1'J_1 & \bar{G}_2 \\ 0 & \end{array}\right]P' = P\bar{G}P' = \bar{A}, \quad \text{say}$$

Noting that

$$\bar{A} = J_1 J_1' + P\bar{G}_2 J_2',$$

we see that

$$T(\hat{A} - \bar{A})J_1 = P\{T(\hat{G} - \bar{G})\}P'J_1$$
$$= P\{T(\hat{G}_1 - \bar{G}_1)\}$$

and (19) follows directly Similarly,

$$\sqrt{T}(\hat{A} - \bar{A})J_2 = P\{\sqrt{T}(\hat{G} - \bar{G})\}P'J_2$$
$$= P\{\sqrt{T}(\hat{G}_2 - \bar{G}_2)\}$$

and we have (20) Finally, writing $\zeta \equiv N(0, (P \otimes I)\bar{Q}(P' \otimes I))$ we obtain

$$\sqrt{T}(\hat{A} - \bar{A})P \Rightarrow [0, \zeta],$$

and thus

$$\sqrt{T}(\hat{A} - \bar{A}) \Rightarrow [0, \zeta]P' \equiv N(0, (P \otimes J_2)\bar{Q}(P' \otimes J_2'))$$

as required for (21)

## REFERENCES


CHAN, N. H., AND C. Z. WEI (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics* (forthcoming).

EBERLAIN, E (1986): "On Strong Invariance Principles Under Dependence Assumptions," *Annals of Probability*, 14, 260–270.

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

EVANS, G B. A., AND N. E SAVIN (1981): "Testing for Unit Roots: 1," *Econometrica*, 49, 753–779

——— (1984): "Testing for Unit Roots: 2," *Econometrica*, 52, 1241–1269.

FULLER (1976): *Introduction to Statistical Time Series*. New York. John Wiley

GRANGER, C. W J, AND P. NEWBOLD (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

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

HANNAN, E J (1970): *Multiple Time Series*. New York· John Wiley.

IBRAGIMOV, I. A, AND Y V LINNIK (1971): *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

KRAMER, W. (1984): "On the Consequences of Trend for Simultaneous Equation Estimation," *Economics Letters*, 14, 23–30

NELSON, C R, AND CHARLES PLOSER (1982). "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162

PARK, J. Y, AND P. C B. PHILLIPS (1986): "Statistical Inference in Regressions with Integrated Processes: Part 1," Cowles Foundation Discussion Paper No. 811.

PHILLIPS, P C. B (1974): "The Estimation of Some Continuous Time Models," *Econometrica*, 42, 803–824.

——— (1986a): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340

——— (1986b): "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 B\, dB'$," Cowles Foundation Discussion Paper No. 796 (forthcoming in *Journal of Multivariate Analysis*).

——— (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–302

——— (1987b): "Towards a Unified Asymptotic Theory of Autoregression," *Biometrika*, 74, 535–547

——— (1987c): "Asymptotic Expansions in Nonstationary Vector Autoregressions," *Econometric Theory*, 3, 45–68.

PHILLIPS, P. C B, AND S. N. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

PHILLIPS, P C B., AND S OULIARIS (1986): "Testing for Cointegration," Cowles Foundation Discussion Paper No. 809

PHILLIPS, P. C B, AND P PERRON (1986): "Testing for a Unit Root in Time Series Regression," Cowles Foundation Discussion Paper No 795 (forthcoming in *Biometrika*)

STOCK, J., AND M. WATSON (1986): "Testing for Common Trends," Harvard University Discussion Paper No. 1222.

WEST, K. D. (1986): "Asymptotic Normality when Regressors have a Unit Root," Princeton Discussion Paper in Economics #110

WITHERS, C S. (1981): "Conditions for Linear Processes to be Strong Mixing," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 57, 477–480