

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS  
AT YALE UNIVERSITY**

**Box 2125, Yale Station  
New Haven, Connecticut 06520**

**COWLES FOUNDATION DISCUSSION PAPER NO. 781**

**Note:** Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

**Regression Theory for Near-Integrated Time Series**

by

**Peter C. B. Phillips**

**January 1986**

# REGRESSION THEORY FOR NEAR-INTEGRATED TIME SERIES

by

P. C. B. Phillips\*

Cowles Foundation for Research in Economics  
Yale University

## O. ABSTRACT

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 on function spaces. This theory generalizes the limit theory that was derived in Phillips and Durlauf (1985) for integrated processes. It is also shown to imply a general central limit theory for standardized sums of stationary processes. The theory is applied to the study of vector autoregressions and cointegrating regressions of the type recently advanced by Granger and Engle (1985). A noncentral limiting distribution theory is derived for the unit root tests that have been proposed by Dickey and Fuller (1979, 1981), by Evans and Savin (1981, 1984) and by the author (1985a). This noncentral distribution theory yields some interesting insights into the asymptotic power properties of the various tests.

January, 1986

---

\*My thanks as always go to Glenna Ames for her skill and effort in typing the manuscript of this paper. The research reported here was supported by the NSF under Grant No. SES 8218792.

## 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.

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 shall describe as near-integrated. A formal definition is given in Section 3 of the paper but the class may be taken to include stationary time series with a strongly autoregressive component and non-stationary 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. Section 3 defines the concept of a near-integrated vector random process. A limit theory for the sample moments of such a process is developed using weak convergence on function spaces. The resulting theory generalizes the limit theory derived recently in Phillips and Durlauf (1985) for integrated processes. A rather remarkable by-product of the theory given in this Section for the asymptotic distribution of the sample mean of a near-integrated process is shown to be a general central limit theorem for the standardized sum of a stationary process. In Section 4 the theory is applied to the study of vector autoregressions. A noncentral limiting distribution theory is derived in Section 5 for the unit root tests that have been proposed by Dickey and Fuller (1979, 1981), by Evans and Savin (1981, 1984) and by the author (1985a). These noncentral distributions help in the analysis of the local asymptotic power properties of the various tests. Section 6 develops a general asymptotic theory for multiple regressions with near-integrated time series. The results of this Section include as special cases: spurious regressions of integrated processes, which have been studied recently by the author (1985b) elsewhere; and cointegrating regressions of the type that have been advanced by Granger and Engle (1985). Some conclusions are given in Section 7. Proofs of results that appear in the text of the paper are provided in the Mathematical Appendix.

## 2. PRELIMINARIES

Let  $\{u_t\}_1^\infty$  be a sequence of random  $n$ -vectors on a probability space  $(\Omega, \mathcal{B}, P)$ . We introduce the vector of partial sums  $S_t = \sum_{j=1}^t u_j$  and set  $S_0 = 0$ . The following assumption will be used frequently in our theoretical development.

## ASSUMPTION 2.1

- (a)  $E(u_t) = 0$  all  $t$  ;
- (b)  $\sup_{i,t} E|u_{it}|^{\beta+\epsilon} < \infty$  for some  $\beta > 2$  and  $\epsilon > 0$  ;
- (c)  $\Sigma = \lim_{T \rightarrow \infty} T^{-1} E(S_T S_T')$  exists and is positive definite
- (d)  $\{u_t\}_1^\infty$  is strong mixing with mixing numbers  $\alpha_m$  that satisfy:
- (1) 
$$\sum_1^\infty \alpha_m^{1-2/\beta} < \infty .$$

The conditions imposed on the innovation sequence  $\{u_t\}_1^\infty$  by Assumption 2.1 are rather weak and allow for many weakly dependent and heterogeneously distributed time series. They include a wide variety of possible data generating mechanisms such as finite order ARMA models under very general conditions on the underlying errors. Note that condition (b) of Assumption 2.1 controls the allowable heterogeneity of the process, whereas (d) controls the extent of permissible temporal dependence in the process in relation to the probability of outlier occurrences. Thus, the summability condition (1) is satisfied when the mixing decay rate is  $\alpha_m = O(m^{-\lambda})$  for some  $\lambda > \beta/(\beta-2)$ . As  $\beta$  approaches 2 and the probability of outliers rises (under the weakening moment condition (b)) the mixing decay rate thereby increases and the effect of outliers is then required under (1) to wear off more quickly.

Note that if  $\{u_t\}$  is weakly stationary then

$$(2) \quad \Sigma = E(u_1 u_1') + \sum_{k=1}^{\infty} E(u_1 u_k' + u_k u_1')$$

and the convergence of this series is implied by the mixing condition (1) (theorem 18.5.3 of Ibragimov and Linnik (1971)).

From the sequence of partial sums  $\{S_t\}_1^T$  we now construct the random

element:

$$(3a) \quad X_T(r) = T^{-1/2} \Sigma^{-1/2} S_{[Tr]} = T^{-1/2} \Sigma^{-1/2} S_{j-1} ; \quad ((j-1)/T \leq r < j/T, \quad j = 1, \dots, T)$$

$$(3b) \quad X_T(1) = T^{-1/2} \Sigma^{-1/2} S_T$$

where  $[a]$  denotes the integer part of  $a$ .  $X_T(r)$  lies in the product metric space  $D^n = D[0,1] \times \dots \times D[0,1]$  where  $D[0,1]$  is the space of all real valued functions on the interval  $[0,1]$  that are right continuous and have finite left limits. We endow  $D^n$  with the metric

$$d_n(f, g) = \max_i \{d_0(f_i, g_i) : i = 1, \dots, n; f_i, g_i \in D[0,1]\}$$

where  $d_0(,)$  is the modified Skorohod metric (see Billingsley (1968), p. 112) under which  $D[0,1]$  is separable and complete.

Under very general conditions we may establish a central limit theory for  $X_T(r)$  on the function space  $D^n$ . We shall, in particular, make use of the following result which is proved in Phillips (1985c, theorem 2.2).

*LEMMA 2.2. If  $\{u_t\}_1^\infty$  is a sequence of random  $n$ -vectors satisfying Assumption 2.1 then  $X_T(r) \Rightarrow W(r)$ , a vector Wiener process, as  $T \uparrow \infty$ .*

The notation " $\Rightarrow$ " in the statement of Lemma 2.2 is used to signify the weak convergence of the probability measure of  $X_T(r)$  to the probability measure (here, multivariate Wiener measure) of the random function  $W(r)$ . The result is a multivariate functional central limit theorem (CLT) i.e. a CLT on the function space  $D^n$ . It may also be described as a multivariate invariance principle following early (univariate) work by Donsker (1951) and Erdős and Kac (1946).

The limit process  $W(r)$  in Lemma 2.2 is popularly known as the vector

Wiener process or as vector Brownian motion. The sample paths of  $W(r)$  lie almost surely (Wiener measure) in  $C^n = C[0,1] \times \dots \times C[0,1]$  ( $n$  copies) where  $C[0,1]$  is the space of all real valued continuous functions on  $[0,1]$ . Moreover, the vector random function  $W(r)$  is Gaussian, with independent increments (so that  $W(s)$  is independent of  $W(r) - W(s)$  for  $0 < s < r \leq 1$ ) and with independent elements (so that  $W_i(r)$  is independent of  $W_j(r)$ ,  $i \neq j$ ).

We shall also have occasion to use the following random element of  $C^n$  :

$$(4a) \quad Z_T(r) = T^{-1/2} \Sigma^{-1/2} S_{[Tr]} + T^{-1/2} (Tr - [Tr]) \Sigma^{-1/2} u_{[Tr]+1}$$

$$(4b) \quad Z_T(1) = T^{-1/2} \Sigma^{-1/2} S_T .$$

Now  $Z_T(r)$  differs from  $X_T(r)$  by a term of  $O_p(T^{-1/2})$ . It follows that  $Z_T(r) \Rightarrow W(r)$  as  $T \uparrow \infty$ , where the convergence to vector Brownian motion may be interpreted in the sense of the uniform topology (see theorem 2.3 of Phillips (1985c)).

### 3. NEAR-INTEGRATED PROCESSES

Let  $\{y_t\}_1^\infty$  be a multiple ( $n \times 1$ ) time series generated by the model:

$$(5) \quad y_t = Ay_{t-1} + u_t ; \quad t = 1, 2, \dots$$

$$(6) \quad A = \exp(T^{-1}C) .$$

To complete the specification of (5) either of the commonly proposed initial conditions may be employed:

(i)  $y_0 = c$ , a constant; or

(ii)  $y_0 =$  random with a certain specified distribution.



In addition, the innovation sequence  $\{u_t\}_1^\infty$  in (5) will be required to satisfy the weak dependence and moment conditions of Assumption 2.1.

In (6)  $C$  is an  $n \times n$  matrix which we may use to measure deviations from the null hypothesis:

$$H_0 : A = I_n .$$

Of course,  $H_0$  applies when  $C = 0$ . The time series  $\{y_t\}_1^\infty$  is then said to be an integrated process. This may be formalized as follows:

*DEFINITION 3.1. A time series  $\{y_t\}_1^\infty$  that is generated by (5) and (6) with  $C = 0$  and in which the innovation sequence  $\{u_t\}_1^\infty$  satisfies Assumption 2.1 is called an integrated (vector) process.*

When  $C \neq 0$ , (6) represents a local alternative to  $H_0$ . As  $T \uparrow \infty$  of course  $A \rightarrow I_n$ . However, the rate of approach to  $I_n$  is not so fast that the alternative hypothesis represented by (6) 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 (6) on the limiting distribution of statistics based on data generated by (5) is well defined and directly measurable in terms of the (noncentrality) parameter matrix  $C$ .

Note that an alternative approach would have been to replace the matrix exponential representation of  $A$  in (6) by deviations from  $I_n$  of the form:

$$(7) \quad A = I_n + T^{-1}C .$$

With this formulation (suggested in Phillips (1985d)) the approach would be

analogous to that which is conventionally employed in the statistical analysis of asymptotic power under local alternatives. In (7), as in (6), deviations of  $O(T^{-1})$  from  $I_n$  are employed to ensure compatibility with the rate at which the statistics under study converge as  $T \uparrow \infty$ .

In our analysis here, alternatives to  $H_0$  of the form given in (6) rather than (7) will be used. This choice is made because under (6) the limiting distribution theory is more elegantly obtained and is more readily related to the null than it is under (7). Working with alternatives of the form (6) we now formalize the concept of a near-integrated process.

*DEFINITION 3.2.* A time series  $\{y_t\}_1^\infty$  that is generated by (5) and (6) with  $C \neq 0$  and in which the innovation sequence  $\{u_t\}_1^\infty$  satisfies Assumption 2.1 is called a near-integrated (vector) process.

Our first result concerns the asymptotic behavior of sample moments of the process  $\{y_t\}_1^\infty$ .

*LEMMA 3.3.* If  $\{y_t\}_1^\infty$  is a near-integrated process generated by (5) and (6) then as  $T \uparrow \infty$ :

- (a)  $T^{-3/2} \Sigma_1^T y_t \Rightarrow \Sigma^{1/2} \int_0^1 W(r) dr + C \int_0^1 B(r) dr$  ;
- (b)  $T^{-2} \Sigma_1^T y_t y_t' \Rightarrow \Sigma^{1/2} \int_0^1 W(r) W(r)' dr \Sigma^{1/2} + C \int_0^1 B(r) W(r)' \Sigma^{1/2} + \Sigma^{1/2} \int_0^1 W(r) B(r)' dr C' + C \int_0^1 B(r) B(r)' dr C'$  ;
- (c)  $T^{-1} \Sigma_1^T y_{t-1} u_t' \Rightarrow \Sigma^{1/2} \int_0^1 W(r) dW(r)' \Sigma^{1/2} + (1/2) (\Sigma - \Sigma_u) + C \int_0^1 B(r) dW(r)' \Sigma^{1/2}$  ;

where

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

$$\Sigma_u = \lim_{T \rightarrow \infty} T^{-1} \Sigma_1^T E(u_t u_t') ,$$

$$B(r) = \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} W(s) ds ,$$

and  $W(r)$  is a vector Wiener process on  $C^n$ . Results (a)-(c) hold under either of the initial conditions (i) or (ii).

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  $T \uparrow \infty$ . In the Lemma the limiting distributions of these sample moments are characterized as functionals of the vector Wiener process  $W(r)$ . Other representations than those given here are also possible and some of these are obtained in the proof of the Lemma given in the Appendix. Thus, in the case of (a) we find that:

$$(8) \quad \Sigma^{1/2} \int_0^1 \int_0^1 W(r) dr + C \int_0^1 B(r) dr = \int_0^1 \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dW(s) dr \equiv N(0, V)$$

where

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

In the scalar case ( $n = 1$ ) this limiting covariance matrix is easily evaluated to be ( $V = v$ ,  $\Sigma = \sigma^2$ ,  $C = c$ ) :

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

Note that for particular cases of (5) in which  $A$  is assigned a value in the vicinity of  $I_n$  the results of the Lemma may be used to deduce 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 may set  $c = T \ln a$  so that  $a = e^{c/T}$  and then

$$(11) \quad v = \frac{\sigma^2}{(T \ln a)^2} + \frac{\sigma^2}{2(T \ln a)^3} [3 - 4a^T + a^{2T}]$$

is an approximation to the variance of  $T^{-3/2} \sum_1^T y_t$ . This implies that  $T^{-1/2} \sum_1^T y_t$  is approximately  $N(0, v_T)$  where

$$(12) \quad v_T = \frac{\sigma^2}{(\ln a)^2} + \frac{\sigma^2}{2T(\ln a)^3} [3 - 4a^T + a^{2T}] .$$

In the stationary case ( $a < 1$ ) the leading term here may be approximated as:

$$(13) \quad \begin{aligned} v_T &= \sigma^2 / (\ln a)^2 + O(T^{-1}) \\ &\sim \sigma^2 / (1-a)^2 \end{aligned}$$

when  $a$  is close to unity.

Interestingly, this approximation (13) gives the exact asymptotic variance in the stationary case for all values of  $a$  ( $|a| < 1$ ). Indeed, we know that in this case

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

(see, for instance, Hall and Heyde (1980), p. 135) where  $f_y(\lambda)$  is the spectral density of the stationary process  $\{y_t\}_1^\infty$ . Here  $y_t$  is generated by the stable AR(1)  $y_t = ay_{t-1} + u_t$  with stationary errors  $u_t$ . The spectral density of  $y_t$  is therefore given by:

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

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

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

where

$$\sigma^2 = E(u_1^2) + 2\sum_{k=1}^{\infty} E(u_1 u_k)$$

as given earlier by (2). Thus,  $2\pi f_y(0) = \sigma^2/(1-a)^2$  and this proves that (13) yields the correct asymptotic variance of  $T^{-1/2}\sum_1^T y_t$  in the stationary case.

This rather remarkable deduction from the simple approximation (13) extends easily to the general case of vector processes. Here we find that when (5) is stationary and we set (using the principal value of the logarithm)

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

the analogue of (12) is:

$$\begin{aligned} V_T &= T^2 V = T^2 C^{-1} \Sigma C^{-1} + O(T^{-1}) \\ &= (I-A)^{-1} \Sigma (I-A')^{-1} + O(T^{-1}) . \end{aligned}$$

Thus, part (a) of Lemma 3.3 and (8) and (9) now imply that as  $T \uparrow \infty$  :

$$\begin{aligned} (14) \quad T^{-1/2} \sum_1^T y_t &\Rightarrow N\left(0, (I-A)^{-1} \Sigma (I-A')^{-1}\right) \\ &\equiv N(0, 2\pi f_{yy}(0)) \end{aligned}$$

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  $u_t$ . (14)

now follows since

$$2\pi f_{uu}(0) = E(u_1 u_1') + \sum_{k=1}^{\infty} \{E(u_1 u_k') + E(u_k u_1')\} = \Sigma .$$

Once again, (14) gives precisely for all (stable)  $A$  the well known result from the theory of stationary processes (see, for example, Hannan (1970, theorem 11, p. 221)).

Additionally, by setting  $C = 0$  in Lemma 3.3 we deduce the limit theory that applies in the case of integrated processes. Thus, in part (a) of the Lemma, we find in this case that as  $T \uparrow \infty$  :

$$(15) \quad T^{-3/2} \Sigma_1^T y_t \Rightarrow \Sigma \int_0^1 W(r) dr \equiv N(0, (1/3) \Sigma)$$

which is precisely the result obtained in Phillips and Durlauf (1985, lemma 3.1) by direct methods.

From this analysis we conclude that the limiting distribution theory given by Lemma 3.3 for near-integrated processes may be specialized to yield not only the appropriate limit theory for the sample moments of integrated processes but also a general central limit theory for standardized sums of stationary processes.

#### 4. VECTOR AUTOREGRESSIONS WITH NEAR-INTEGRATED PROCESSES

We shall study the least squares vector autoregression

$$(16) \quad y_t = \hat{A} y_{t-1} + \hat{u}_t ; \quad t = 1, \dots, T$$

where

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

and the associated error covariance matrix estimator is:

$$\hat{\Sigma} = T^{-1} Y' (I - P) Y , \quad P = Y_{-1} (Y_{-1}' Y_{-1})^{-1} Y_{-1}' .$$

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.** *If  $\{y_t\}_1^\infty$  is a near-integrated process generated by (5) and (6) then as  $T \uparrow \infty$  :*

- (a)  $T(\hat{A}-I) \Rightarrow C + \left\{ \Sigma^{1/2} \int_0^1 W(r) dW(r)' \Sigma^{1/2} + (1/2) (\Sigma - \Sigma_u) + C \int_0^1 B(r) dW(r)' \Sigma^{1/2} \right\}'$   
 $\cdot \left\{ \Sigma^{1/2} \int_0^1 W(r) W(r)' dr \Sigma^{1/2} + C \int_0^1 B(r) W(r)' dr \Sigma^{1/2} + \Sigma^{1/2} \int_0^1 W(r) dB(r)' dr C' \right.$   
 $\left. + C \int_0^1 B(r) B(r)' dr C' \right\}^{-1}$  ;
- (b)  $\hat{A} \xrightarrow{p} I$  ;
- (c)  $\hat{\Sigma} \xrightarrow{p} \Sigma_u$  ;

where  $W(r)$  is a vector Wiener process on  $C^n$  and  $B(r)$  is defined in Lemma 3.3.

**THEOREM 4.2.** *Let  $v_t = (u_{it}^2 - E(u_{it}^2))_{n \times 1}$ . If  $\{v_t\}_1^\infty$  is a weakly stationary process and satisfies Assumption 2.1 and if  $\{y_t\}$  is a near-integrated process generated by (5) and (6) then as  $T \uparrow \infty$*

$$T^{1/2} \text{vec}(\Sigma^* - \Sigma_u) \Rightarrow N(0, Q)$$

where

$$Q = P_D \sum_{k=0}^{\infty} \{ \Psi_k - (\text{vec } \Sigma_u) (\text{vec } \Sigma_u)' \} P_D$$

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

$$P_D = D(D'D)^{-1} D'$$

and  $D$  is the  $n^2 \times n(n+1)/2$  duplication matrix of Magnus and Neudecker (1980).

Theorems 4.1 and 4.2 extend to near-integrated processes the theory developed in Phillips and Durlauf (1985) 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 the near-integration of the process 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 parts (b) and (c) 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 (5) and (6)) under serial correlation even when the time series are near-integrated.

From Theorem 4.2 we see that the asymptotic distribution of the error covariance matrix estimator  $\hat{\Sigma}$  is independent of  $C$  and is therefore the same for integrated and near-integrated processes. This is explained by the fact that in both cases  $\hat{A} \xrightarrow{p} I_n$  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 apply in this case, therefore, as we would expect for stationary processes and for processes that satisfy Assumption 2.1.

## 5. POWER FUNCTIONS FOR TESTS OF THE RANDOM WALK HYPOTHESIS

The theory of the preceding section may be used to develop power functions for regression based tests of the random walk hypothesis. In order to relate our results to earlier work by Dickey and Fuller (1979, 1981), by Savin and Evans (1981, 1984) and by the present author (1985) we shall confine our attention here to the important scalar case. Thus, we set  $n = 1$ ,  $A = a$ ,  $\Sigma = \sigma^2$ ,  $\Sigma_u = \sigma_u^2$ ,  $C = c$  and  $\hat{\Sigma} = s_u^2$ . The model given



by (5) and (6) for the generation of the time series is now:

$$(17) \quad y_t = ay_{t-1} + u_t ; \quad t = 1, 2, \dots$$

$$(18) \quad a = e^{c/T} .$$

We shall assume, as before, that the innovation sequence  $\{u_t\}_1^\infty$  in (17) satisfies Assumption 2.1.

In addition to the least squares regression coefficient  $\hat{a}$  we shall also consider the regression  $t$  statistic defined by

$$(19) \quad t_a = (\sum_1^T y_{t-1}^2)^{1/2} (\hat{a}-1)/s_u$$

and the following new test statistics proposed by the author (1985):

$$(20) \quad Z_a = T(\hat{a}-1) - (1/2)(s_{T\ell}^2 - s_u^2)/(T^{-2}\sum_1^T y_{t-1}^2)$$

$$(21) \quad Z_t = (\sum_1^T y_{t-1}^2)^{1/2} (\hat{a}-1)/s_{T\ell} - (1/2)(s_{T\ell}^2 - s_u^2)/s_{T\ell} (T^{-2}\sum_1^T y_{t-1}^2)^{1/2}$$

where

$$s_{T\ell}^2 = T^{-1}\sum_1^T \hat{u}_t^2 + 2T^{-1}\sum_{s=1}^{\ell}\sum_{t=s+1}^T \hat{u}_t \hat{u}_{t-s}$$

and  $\ell$  is the lag truncation number employed in the estimation of  $\sigma^2$  by  $s_{T\ell}^2$ . We have used regression residuals  $\hat{u}_t$  in the construction of  $s_{T\ell}^2$  and  $s_u^2$ , although this is not strictly necessary. First differences in the time series  $y_t$  may also be used (as in Phillips (1985a)) since these reproduce precisely the innovations  $u_t$  under the null hypothesis that  $a = 1$ .

*THEOREM 5.1.* Let  $\{y_t\}_1^\infty$  be a near-integrated (scalar) process generated by (17) and (18). If

(a)  $\sup_t E|u_t|^{4p} < \infty$  for some  $p > 1$  ;

(b) the mixing numbers  $\alpha_m$  of the sequence  $\{u_t\}_1^\infty$  are such that

$$\alpha_m = O(m^{-2p/(p-1)-\varepsilon})$$

for  $p > 1$  as in (a) and some  $\varepsilon > 0$  ;

(c) the lag truncation number  $l \uparrow \infty$  or  $T \uparrow \infty$  such that  $l = o(T^{1/4})$  ;

then as  $T \uparrow \infty$  :

$$(22) \quad Z_a \Rightarrow c + \int_0^1 \{W(r) + cF(r)\} dW(r) / \int_0^1 \{W(r) + cF(r)\}^2 dr$$

$$(23) \quad Z_t \Rightarrow c \left[ \int_0^1 \{W(r) + cF(r)\}^2 dt \right]^{1/2} + \int_0^1 \{W(r) + cF(r)\} dW(r) / \left[ \int_0^1 \{W(r) + cF(r)\}^2 dr \right]^{1/2}$$

where  $W(r)$  is a Wiener process on  $C[0,1]$  and  $F(r) = \int_0^r e^{(r-s)c} W(s) ds$  .

This Theorem gives the noncentral ( $c \neq 0$ ) limiting distributions of the two test statistics  $Z_a$  and  $Z_t$  under the sequence of local alternatives (as  $T \uparrow \infty$ )

$$(24) \quad H_1 : a = e^{c/T}$$

to the unit root null hypothesis

$$(25) \quad H_0 : a = 1 .$$

As discussed in Phillips (1985a) the test statistics  $Z_a$  and  $Z_t$  are constructed to allow for quite general weakly dependent and heterogeneously distributed innovations  $\{u_t\}_1^\infty$  in the null model  $y_t = y_{t-1} + u_t$  . Moreover, the limiting distributions of  $Z_a$  and  $Z_t$  were shown: (i) to be

independent of nuisance parameters (specifically, the correlation sequence of the innovations  $\{u_t\}_1^\infty$ ); and (ii) to be identical to the limiting distributions of  $T(\hat{a}-1)$  and  $t_a$ , respectively, where the latter were obtained under the null hypothesis (25) and under the assumption that the innovation sequence  $\{u_t\}_1^\infty$  is  $iid(0, \sigma^2)$ . Thus, asymptotic critical values for the test statistics  $Z_a$  and  $Z_t$  may be readily obtained from the published tables of the latter distributions in Fuller (1976).

Our next theorem demonstrates an important equivalence between the asymptotic power properties of  $Z_a$  and  $Z_t$  (under quite general weakly dependent innovations) and those of the tests based on  $T(\hat{a}-1)$  and  $t_a$  (under iid errors).

*THEOREM 5.2. If the time series  $\{y_t\}_1^\infty$  is generated by (17) and (18) and if the innovation sequence  $\{u_t\}_1^\infty$  is  $iid(0, \sigma^2)$  then as  $T \rightarrow \infty$ :*

$$(26) \quad T(\hat{a}-1) \Rightarrow c + \int_0^1 \{W(r) + cF(r)\} dW(r) / \int_0^1 \{W(r) + cF(r)\}^2 dr ;$$

$$(27) \quad t_a \Rightarrow c \left[ \int_0^1 \{W(r) + cF(r)\}^2 dt \right]^{1/2} + \int_0^1 \{W(r) + cF(r)\} dW(r) / \left[ \int_0^1 \{W(r) + cF(r)\}^2 dr \right]^{1/2}$$

where  $W(r)$  is a Wiener process on  $C[0,1]$  and  $F(r) = \int_0^r e^{(r-s)c} W(s) ds$ .

The noncentral ( $c \neq 0$ ) limiting distribution of  $T(\hat{a}-1)$  represented by (26) is identical to that of  $Z_a$  given earlier by (22), the latter holding under much more general conditions. Similarly, the noncentral limiting distribution of the regression  $t$  statistic (27) is identical to that of  $Z_t$  as given by (23). We deduce from these results that the new tests based on  $Z_a$  and  $Z_t$  have the same asymptotic power properties under a wide range of possible innovation processes  $\{u_t\}_1^\infty$  in (17) as the regression test based on  $T(\hat{a}-1)$  and  $t_a$ , respectively,

have for the much narrower class of  $\text{iid}(0, \sigma^2)$  innovations. Thus, for example, under iid innovations  $\{u_t\}_1^\infty$ ,  $Z_a$  and  $T(\hat{a}-1)$  have the same asymptotic critical value under the null (25) and the same local power under the alternative hypothesis (24). However,  $Z_a$  retains this asymptotic critical value and this local asymptotic power for a wide range of weakly dependent and heteroskedastic innovations. By contrast, the nominal critical value of  $T(\hat{a}-1)$  becomes biased when the innovations are no longer  $\text{iid}(0, \sigma^2)$ ; and the correct asymptotic critical value of this test for a stated size is now generally dependent on nuisance parameters. Thus, it becomes necessary to introduce some test statistic such as  $Z_a$  even to ensure the correct asymptotic size in this case. Theorems 5.1 and 5.2 then tell us that there is no loss of asymptotic power in the use of the appropriately corrected statistic  $Z_a$  over that which obtained for  $T(\hat{a}-1)$  under iid errors. A similar argument applies to the statistic based on  $Z_t$ .

## 6. MULTIPLE REGRESSION WITH NEAR INTEGRATED TIME SERIES

The theory developed earlier in this paper may now be applied to multiple (least squares) regressions of the form:

$$(28) \quad x_t = \hat{\alpha} + \hat{\beta}' z_t + \hat{u}_t ; \quad t = 1, \dots, T$$

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 that the multiple time series  $\{y_t\}_1^\infty$  is generated by (5) and (6) with innovations  $\{u_t\}_1^\infty$  that satisfy Assumption 2.1. Under this set up, some elements of  $y_t$  may be integrated processes, others may be near-integrated; the innovations  $u_t$  that drive (5)

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 (28). 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 (28);  $t_{\beta_i}$  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 6.1.* *If (28) is estimated by least squares and if  $\{y_t\}_1^\infty$  is a near-integrated process generated by (5) and (6) with  $y_t' = (z_t, z_t')$ , then as  $T \uparrow \infty$ :*

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

where

$$(29) \quad G = \begin{bmatrix} 1 & m \\ g_{11} & g_{21}' \\ g_{21} & G_{22} \end{bmatrix}^1_m = \int_0^1 J(r)J(r)'dr - \int_0^1 J(r)dr \int_0^1 J(r)'dr ;$$

$$b = \int_0^1 J(r)dr ;$$

$$J(r) = \Sigma^{1/2}W(r) + CB(r) = \Sigma^{1/2}W(r) + C \int_0^r \exp\{(r-s)C\} \Sigma^{1/2}W(s) ds ;$$

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

$$\Sigma = \lim_{T \rightarrow \infty} E\{T^{-1}(\Sigma_1^T u_t)(\Sigma_1^T u_t)'\};$$

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

and  $W(r)$  is a vector Wiener process on  $C^n$ .

Theorem 6.1 generalizes to near-integrated processes the regression theory derived in Phillips (1985b) for integrated processes. We see from Theorem 6.1 that 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  $DW \xrightarrow{p} 0$  as  $T \uparrow \infty$ .

Equation (28) may be regarded as a cointegrating regression of the type recently considered by Granger and Engle (1985). 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 6.1 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 integrated

processes, others may be nearly explosive, while yet others may be nearly stationary. The effects of these extensions are, of course, measured through the noncentrality matrix  $C$ .

Under the alternative hypothesis that the variables in the regression are cointegrated a different asymptotic theory applies. The reader is referred to Phillips and Durlauf (1985) for a detailed development of the relevant asymptotic theory for regressions such as (28) in this case.

## 7. 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 important general departures from conventional theory based on stationary processes. The new asymptotic theory, it is hoped, will be helpful in describing 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 also been applied to analyze the non-central distributions of certain tests of the random walk hypothesis under local alternatives. The results provide some helpful asymptotic power comparisons among the tests. In particular, they indicate that the new tests introduced by the author in an earlier paper (1985) involve no loss in asymptotic power over existing tests such as those of Dickey-Fuller (1979, 1981) in spite of the fact that the new tests permit a wide class of weakly dependent and heterogeneously distributed innovations.

## MATHEMATICAL APPENDIX

Proof of Lemma 3.3. From (5) and (6) we deduce the representation:

$$\begin{aligned} y_t &= \sum_{k=0}^{t-1} \exp\{(k/T)C\} u_{t-k} + \exp\{(t/T)C\} y_0 \\ (A1) \quad &= \sum_{j=1}^t \exp\{((t-j)/T)C\} u_j + \exp\{(t/T)C\} y_0 . \end{aligned}$$

Thus,

$$\begin{aligned} T^{-3/2} \sum_{l=1}^T y_t &= T^{-3/2} \sum_{i=1}^T \sum_{j=1}^i \exp\{((i-j)/T)C\} u_j + T^{-3/2} \sum_{i=1}^T \exp\{(i/T)C\} y_0 \\ &= T^{-1} \sum_{i=1}^T \sum_{j=1}^i \exp\{((i-j)/T)C\} \Sigma^{1/2} \int_{(j-1)/T}^{j/T} dX_T(s) + O_p(T^{-1/2}) \\ (A2) \quad &= \sum_{i=1}^T \sum_{j=1}^i \int_{(i-1)/T}^{i/T} dr \int_{(j-1)/T}^{j/T} \exp\{((i-j)/T)C\} \Sigma^{1/2} dX_T(s) + O_p(T^{-1/2}) . \end{aligned}$$

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

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

and (A2) becomes:

$$\begin{aligned} &\sum_{i=1}^T \sum_{j=1}^i \int_{(i-1)/T}^{i/T} dr \int_{(j-1)/T}^{j/T} \exp\{(r-s)C\} \Sigma^{1/2} dX_T(s) + O_p(T^{-1/2}) \\ &= \int_0^1 dr \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dX_T(s) + O_p(T^{-1/2}) \\ &= \Sigma^{1/2} \int_0^1 X_T(r) dr + C \int_0^1 \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} X_T(s) ds dr + O_p(T^{-1/2}) \\ &\Rightarrow \Sigma^{1/2} \int_0^1 W(r) dr + C \int_0^1 \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} W(s) ds dr \\ &= \Sigma^{1/2} \int_0^1 W(r) dr + C \int_0^1 B(r) dr \end{aligned}$$



as  $T \uparrow \infty$ , in view of Lemma 2.2 and the continuous mapping theorem. This proves part (a) of the Lemma.

To prove (b) we write

$$\begin{aligned} T^{-2} \sum_{i=1}^T y_t y_t' &= T^{-2} \sum_{i=1}^T \left[ \sum_{j=1}^i \exp\{((i-j)/T)C\} u_j \right] \left[ \sum_{k=1}^i u_k' \exp\{((i-k)/T)C\} \right] + o_p(T^{-1/2}) \\ &= \sum_{i=1}^T \int_{(i-1)/T}^{i/T} dr \left[ \sum_{j=1}^i \int_{(j-1)/T}^{j/T} \exp\{(r-s)C\} \Sigma^{1/2} dX_T(s) \right] \\ &\quad \cdot \left[ \sum_{k=1}^i \int_{(k-1)/T}^{k/T} dX_T(t) \Sigma^{1/2} \exp\{(r-t)C'\} \right] + o_p(T^{-1/2}) \end{aligned}$$

$$(A3) = \int_0^1 \int_0^r \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dX_T(s) dX_T(t) \Sigma^{1/2} \exp\{(r-t)C'\} dr + o_p(T^{-1/2})$$

$$(A4) \Rightarrow \int_0^1 \int_0^r \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dW(s) dW(t) \Sigma^{1/2} \exp\{(r-t)C'\} dr$$

as  $T \uparrow \infty$ . The continuous mapping theorem may be employed here since the integrand is continuous, nonstochastic and of bounded variation. For example,

$$\int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dX_T(s) = \Sigma^{1/2} X_T(r) + C \int_0^r \exp\{(r-s)C\} X_T(s) ds$$

which is a continuous functional on  $D[0,1]^n$ . Thus, the first term of

(A4) may be written as:

$$\begin{aligned} &\Sigma^{1/2} \int_0^1 X_T(r) X_T(r)' dr \Sigma^{1/2} + C \int_0^1 \int_0^r \exp\{(r-s)C\} X_T(s) ds X_T(r)' dr \Sigma^{1/2} \\ &+ \Sigma^{1/2} \int_0^1 X_T(r) \int_0^r X_T(t)' \exp\{(r-t)C'\} dt C' \\ &+ C \int_0^1 \left[ \int_0^r \exp\{(r-s)C\} X_T(s) ds \right] \left[ \int_0^r X_T(t)' \exp\{(r-t)C'\} dt \right] dr C' \end{aligned}$$

which is also a continuous functional on  $D[0,1]^n$ . Since  $X_T(t) \Rightarrow W(t)$  as  $T \uparrow \infty$  we deduce that the above functional converges weakly to:

$$\begin{aligned}
& \int_0^1 \Sigma^{1/2} W(r) W(r)' dr \Sigma^{1/2} + C \int_0^1 \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} W(s) ds W(r)' dr \Sigma^{1/2} \\
& + \int_0^1 \Sigma^{1/2} \int_0^r W(r) W(t)' \Sigma^{1/2} \exp\{(r-t)C'\} dt dr C' \\
& + C \int_0^1 \left[ \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} W(s) ds \right] \left[ \int_0^r W(t)' \Sigma^{1/2} \exp\{(r-t)C'\} dt \right] dr C' \\
(A5) \quad & = \int_0^1 \Sigma^{1/2} W(r) W(r)' dr \Sigma^{1/2} + C \int_0^1 B(r) W(r)' dr \Sigma^{1/2} \\
& + \int_0^1 \Sigma^{1/2} W(r) B(r)' dr C' + C \int_0^1 B(r) B(r)' dr C'
\end{aligned}$$

proving (b). Furthermore we note that

$$\begin{aligned}
\int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dW(s) &= \Sigma^{1/2} W(r) + C \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} W(s) ds \\
(A6) \quad &= \Sigma^{1/2} W(r) + CB(r)
\end{aligned}$$

by partial integration, which is valid here because the integrand is non-stochastic, continuous and of bounded variation. Using the representation (A6) in (A4) we readily deduce the equivalence of (A4) and (A5).

Next, we consider

$$\begin{aligned}
& T^{-1} \Sigma_1^T y_{t-1} u_t' = \exp\{(-1/T)C\} [T^{-1} \Sigma_1^T y_t u_t' - T^{-1} \Sigma_1^T u_t u_t'] \\
& = T^{-1} \Sigma_1^T y_t u_t' - T^{-1} \Sigma_1^T u_t u_t' + O_p(T^{-1}) \\
& = T^{-1} \Sigma_{i=1}^T [\Sigma_{j=1}^i \exp\{((i-j)/T)C\} u_j + \exp\{(i/T)C\} y_0] u_i' - T^{-1} \Sigma_1^T u_t u_t' + O_p(T^{-1}) \\
(A6) \quad & = T^{-1} \Sigma_{i=1}^T \Sigma_{j=1}^i \exp\{((i-j)/T)C\} u_j u_i' - T^{-1} \Sigma_1^T u_t u_t' + O_p(T^{-1/2}) .
\end{aligned}$$

It is convenient to use the random element  $Z_T(t)$  defined by (4) in what follows. We note that  $dZ_T(t) = T^{1/2} \Sigma^{-1/2} u_j dt$  for  $(j-1)/T \leq t < j/T$  and then (A6) may be written in the form:

$$\begin{aligned}
& \Sigma_{i=1}^T \int_{(i-1)/T}^{i/T} \left[ \Sigma_{j=1}^i \exp\{((i-j)/T)C\} \int_{(j-1)/T}^{j/T} \Sigma^{1/2} dZ_T(s) \right] dZ_T(r) \Sigma^{1/2} - T^{-1} \Sigma_1^T u_t u_t' + O_p(T^{-1/2}) \\
& = \Sigma_{i=1}^T \int_{(i-1)/T}^{i/T} \left[ \Sigma_{j=1}^i \int_{(j-1)/T}^{j/T} \exp\{(r-s)C\} \Sigma^{1/2} dZ_T(s) \right] dZ_T(r) \Sigma^{1/2} - T^{-1} \Sigma_1^T u_t u_t' + O_p(T^{-1/2}) \\
& = \Sigma_{i=1}^T \int_{(i-1)/T}^{i/T} \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dZ_T(s) dZ_T(r) \Sigma^{1/2} \\
& \quad + \Sigma_{i=1}^T \int_{(i-1)/T}^{i/T} \int_r^{i/T} \exp\{(r-s)C\} \Sigma^{1/2} dZ_T(s) dZ_T(r) \Sigma^{1/2} - T^{-1} \Sigma_1^T u_t u_t' + O_p(T^{-1/2}) \\
(A7) \quad & = \int_0^1 \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dZ_T(s) dZ_T(r) \Sigma^{1/2} \\
& \quad + \Sigma_{i=1}^T \int_{(i-1)/T}^{i/T} \int_r^{i/T} \Sigma^{1/2} dZ_T(s) dZ_T(r) \Sigma^{1/2} - T^{-1} \Sigma_1^T u_t u_t' + O_p(T^{-1/2}) .
\end{aligned}$$

The second term of (A7) may be evaluated as:

$$\begin{aligned}
& T \sum_{i=1}^T \int_{(i-1)/T}^{i/T} (i/T - r) dr u_i u_i' \\
&= T \sum_{i=1}^T u_i u_i' \left[ ir/T - r^2/2 \right]_{(i-1)/T}^{i/T} \\
\text{(A8)} \quad &= (1/2T) \sum_{i=1}^T u_i u_i' .
\end{aligned}$$

We also note that:

$$\text{(A9)} \quad \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} dZ_T(s) = \Sigma^{1/2} Z_T(r) + C \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} Z_T(s) ds .$$

Substituting (A8) and (A9) in (A7) we obtain

$$\begin{aligned}
\text{(A10)} \quad & \Sigma^{1/2} \int_0^1 Z_T(r) dZ_T(r)' \Sigma^{1/2} - (1/2T) \sum_{i=1}^T u_i u_i' \\
&+ C \int_0^1 \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} Z_T(s) ds dZ_T(r)' \Sigma^{1/2} + O_p(T^{-1/2}) .
\end{aligned}$$

Now define

$$B_T(r) = \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} Z_T(s) ds = \exp(rC) \int_0^r \exp(-sC) \Sigma^{1/2} Z_T(s) ds$$

and then

$$\begin{aligned}
\int_0^1 B_T(r) dZ_T(r)' &= [B_T(r) Z_T(r)']_0^1 - \int_0^1 dB_T(r) Z_T(r)' \\
&= B_T(1) Z_T(1)' - C \int_0^1 B_T(r) Z_T(r)' dr - \Sigma^{1/2} \int_0^1 Z_T(r) Z_T(r)' dr .
\end{aligned}$$

Thus,

$$\begin{aligned}
T^{-1} \Sigma_1^T y_{t-1} u_t' &= \Sigma^{1/2} \int_0^1 Z_T(r) dZ_T(r)' \Sigma^{1/2} - (1/2T) \Sigma_1^T u_t u_t' + CB_T(1) Z_T(1)' \Sigma^{1/2} \\
&\quad - C^2 \int_0^1 B_T(r) Z_T(r)' dr \Sigma^{1/2} - C \Sigma^{1/2} \int_0^1 Z_T(r) Z_T(r)' dr \Sigma^{1/2} + o_p(T^{-1/2}) .
\end{aligned}$$

But  $Z_T(t) \Rightarrow W(t)$  as  $T \uparrow \infty$  and

$$B_T(r) \Rightarrow \int_0^r \exp\{(r-s)C\} \Sigma^{1/2} W(s) ds = B(r)$$

by the continuous mapping theorem. Moreover, by Lemma 3 and equation (A14) of Phillips and Durlauf (1985) we know that as  $T \uparrow \infty$

$$\int_0^1 Z_T(r) dZ_T(r)' \Rightarrow \int_0^1 W(r) dW(r)' + (1/2) I_n .$$

Combining these results, we deduce that as  $T \uparrow \infty$  :

$$\begin{aligned}
T^{-1} \Sigma_1^T y_{t-1} u_t' &\Rightarrow \Sigma^{1/2} \int_0^1 W(r) dW(r)' \Sigma^{1/2} + (1/2) (\Sigma - \Sigma_u) \\
&\quad + CB(1) W(1)' \Sigma^{1/2} - C^2 \int_0^1 B(r) W(r)' dr \Sigma^{1/2} - C \Sigma^{1/2} \int_0^1 W(r) W(r)' dr \Sigma^{1/2} \\
&= \Sigma^{1/2} \int_0^1 W(r) dW(r)' \Sigma^{1/2} + (1/2) (\Sigma - \Sigma_u) + C \int_0^1 B(r) dW(r)' \Sigma^{1/2} .
\end{aligned}$$

This proves part (c).

Proof of Theorem 4.1. Define  $U' = [u_1, \dots, u_T]$  and then from (5)

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

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 3.2 and the continuous mapping theorem we deduce that as  $T \uparrow \infty$  :

$$\begin{aligned} T(\hat{A}-I) \Rightarrow & C + \left\{ \Sigma^{1/2} \int_0^1 W(r)W(r)' dr \Sigma^{1/2} \right\} + (1/2)(\Sigma - \Sigma_u) + C \int_0^1 B(r) dW(r)' \Sigma^{1/2} \\ & \cdot \left\{ \Sigma^{1/2} \int_0^1 W(r)W(r)' dr \Sigma^{1/2} + C \int_0^1 B(r)W(r)' dr \Sigma^{1/2} \right. \\ & \left. + \Sigma^{1/2} \int_0^1 W(r) dB(r)' C' + C \int_0^1 B(r)B(r)' dr C' \right\}^{-1} \end{aligned}$$

proving part (a) of the Theorem. (b) then follows directly. To prove (c) we note that as  $T \uparrow \infty$

$$\Sigma^* = T^{-1}U'U - T^{-1}U'Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}'U \xrightarrow{p} \Sigma_u$$

as required, since the second term in the above expression is  $O_p(T^{-1})$  and the first term converges to  $\Sigma_u$  almost surely as  $T \uparrow \infty$  .

Proof of Theorem 4.2. The proof is identical to the proof of Theorem 3.3 of Phillips and Durlauf (1985).

Proof of Theorem 5.1. From Theorem 4.1 we have as  $T \uparrow \infty$  :

$$T(\hat{a}-1) \Rightarrow c + \left\{ \sigma^2 \int_0^1 W(r) dW(r) + (1/2)(\sigma^2 - \sigma_u^2) + c\sigma \int_0^1 B(r) dW(r) \right\} \left\{ \int_0^1 (\sigma W(r) + cB(r))^2 dr \right\}$$

Here

$$B(r) = \int_0^r e^{(r-s)c} \sigma W(s) ds = \sigma F(r) .$$

Moreover, as shown in Theorem 5.2 of Phillips (1985a) under the stated

conditions (a), (b) and (c)  $s_{T\ell}^2 \xrightarrow{p} \sigma^2$  as  $T \uparrow \infty$ . Thus, from (20) and Lemma 3.2 we deduce that as  $T \uparrow \infty$ :

$$\begin{aligned} Z_a &\Rightarrow c + \left\{ \int_0^1 (\sigma W(r) + cB(r))^2 dr \right\}^{-1} \left\{ \sigma^2 \int_0^1 W(r) dW(r) + (1/2)(\sigma^2 - \sigma_u^2) + c\sigma \int_0^1 B(r) dW(r) \right\} \\ &\quad - (1/2)(\sigma^2 - \sigma_u^2) \left\{ \int_0^1 (\sigma W(r) + cB(r))^2 dr \right\}^{-1} \\ &= c + \left\{ \int_0^1 (W(r) + cF(r))^2 dr \right\}^{-1} \left\{ \int_0^1 W(r) dW(r) + c \int_0^1 F(r) dW(r) \right\} \end{aligned}$$

as required for (22). The proof of (23) is entirely analogous and is therefore omitted.

Proof of Theorem 5.2. This follows directly from Theorem 4.1(a), Lemma 3.2 and the fact that  $\sigma^2 = \sigma_u^2$  in this case since  $\{u_t\}_1^\infty$  is iid(0,  $\sigma^2$ ).

Proof of Theorem 6.1. To prove part (a) we first note that:

$$\hat{\beta} = \left\{ T^{-2} \Sigma_1^T (z_t - \bar{z})(z_t - \bar{z})' \right\}^{-1} \left\{ T^{-2} \Sigma_1^T (z_t - \bar{z})(x_t - \bar{x}) \right\}.$$

Now by Lemma 3.2 we deduce that

$$\begin{aligned} T^{-2} \Sigma_1^T (y_t - \bar{y})(y_t - \bar{y})' &= T^{-2} \Sigma_1^T y_t y_t' - (T^{-3/2} \Sigma_1^T y_t) (T^{-3/2} \Sigma_1^T y_t)' \\ &\Rightarrow \int_0^1 J(r) J(r)' dr - \left( \int_0^1 J(r) dr \right) \left( \int_0^1 J(r)' dr \right) \equiv G. \end{aligned}$$

Partitioning  $G$  as in (29) we obtain:

$$\hat{\beta} \Rightarrow G_{22}^{-1} g_{21}; \text{ as } T \uparrow \infty$$

as required. To prove part (b) we have

$$\hat{\alpha} = \bar{x} - \bar{z}'\hat{\beta}$$

so that as  $T \uparrow \infty$

$$T^{-1/2}\hat{\alpha} \Rightarrow b_1 - b_2'G_{22}^{-1}g_{21} = b_1\eta$$

as required. Next

$$\begin{aligned} R^2 &= \hat{\beta}' [T^{-2}\Sigma_1^T(z_t - \bar{z})(z_t - \bar{z})'] \hat{\beta} / T^{-2}\Sigma_1^T(x_t - \bar{x})^2 \\ &\Rightarrow g_{21}'G_{22}^{-1}g_{21}/g_{11} ; \text{ as } T \uparrow \infty \end{aligned}$$

proving (c). Next

$$\begin{aligned} T^{-1}F_{\beta} &= \frac{T-m-1}{Tm} \frac{R^2}{1-R^2} \\ &\Rightarrow \frac{1}{m} \frac{g_{21}'G_{22}^{-1}g_{21}}{g_{11} - g_{21}'G_{22}^{-1}g_{21}} ; \text{ as } T \uparrow \infty \end{aligned}$$

proving (d). To prove (e) we note that

$$\begin{aligned} t_{\beta_i} &= \hat{\beta}_i / S_{\beta_i} \quad \text{where} \quad s_{\beta_i}^2 = s^2 \left[ \{\Sigma_1^T(z_t - \bar{z})(z_t - \bar{z})'\}_{ii} \right] \\ &\quad \text{and} \quad s^2 = T^{-1}\Sigma_1^T\{y_t - \bar{y} - \hat{\beta}'(z_t - \bar{z})\}^2 . \end{aligned}$$

As  $T \uparrow \infty$  we have:

$$\begin{aligned} \hat{\beta}_i &\Rightarrow (G_{22}^{-1}g_{21})_i \\ T^{-1}s^2 &\Rightarrow g_{11} - g_{21}'G_{22}^{-1}g_{21} . \end{aligned}$$

Now



$$t_{\beta_i} = \beta_i / \left\{ T^{1/2} (T^{-1/2} s) T^{-1} \left( \left[ \left\{ T^{-2} \Sigma_1^T (z_t - \bar{z}) (z_t - \bar{z})' \right\}^{-1} \right]_{ii} \right)^{1/2} \right\}$$

so that as  $T \uparrow \infty$

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

proving (e). Finally,

$$\begin{aligned} TDW &= T^{-1} \Sigma_2^T \{x_t - x_{t-1} - \hat{\beta}'(z_t - z_{t-1})\}^2 / \left[ T^{-2} \Sigma_1^T \{x_t - \bar{x} - \hat{\beta}'(z_t - \bar{z})\}^2 \right] \\ &= \hat{b}' [T^{-1} \Sigma_2^T (y_t - y_{t-1}) (y_t - y_{t-1})'] \hat{b} / \hat{b}' [T^{-2} \Sigma_1^T (y_t - \bar{y}) (y_t - \bar{y})'] \hat{b} \end{aligned}$$

where  $\hat{b}' = (1, \hat{\beta}')$ . Now

$$y_t - y_{t-1} = T^{-1} C y_{t-1} + u_t + o_p(T^{-3/2})$$

so that

$$T^{-1} \Sigma_2^T (y_t - y_{t-1}) (y_t - y_{t-1})' = T^{-1} \Sigma_2^T u_t u_t' + o_p(1).$$

But, under the stated conditions

$$T^{-1} \Sigma_2^T u_t u_t' \xrightarrow{\text{a.s.}} \Sigma_u = \lim_{T \rightarrow \infty} T^{-1} \Sigma_1^T E(u_t u_t')$$

by the strong law of McLeish (1975). Thus, as  $T \uparrow \infty$

$$TDW \Rightarrow \eta' \Sigma_u \eta / \eta' G \eta = \eta' \Sigma_u \eta / (g_{11} - g_{21}' G_{22}^{-1} g_{21})$$

proving (f).

## REFERENCES

- Billingsley, P. (1968). *Convergence of Probability Measures*. New York:
- Dickey, D. A. and W. A. Fuller (1979). "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, 427-431.
- \_\_\_\_\_ (1981). "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, 49, 1057-1072.
- Donsker, M. D. (1951). "An invariance principle for certain probability limit theorems," *Memoirs of the American Mathematical Society*, 6, 1-12.
- Erdős, P. and M. Kac (1946). "On certain limit theorems in the theory of probability," *Bulletin of the American Mathematical Society*, 52, 292-302.
- 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, Wayne A. (1976). *Introduction to Statistical Time Series*. New York: John Wiley.
- Granger, C. W. J. and R. F. Engle (1985). "Dynamic model specification with equilibrium constraints: Cointegration and error correction," UCSD Discussion Paper No. 85-18.
- 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.
- McLeish, D. L. (1975). "A maximal inequality and dependent strong laws," *Annals of Probability*, 3, 829-839.
- Magnus, J. R. and H. Neudecker (1980). "The elimination matrix: Some lemmas and applications," *SIAM Journal of Algebraic and Discrete Methods*, 1, 422-449.
- Nelson, Charles R. and Charles I. Plosser (1982). "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10, 139-162.
- Phillips, P. C. B. (1985a). "Time series regression with unit roots," Cowles Foundation Discussion Paper No. 740, Yale University.

\_\_\_\_\_ (1985b). "Understanding spurious regressions in econometrics,"  
Cowles Foundation Discussion Paper No. 757, Yale University.

\_\_\_\_\_ (1985c). "Asymptotic expansions in nonstationary vector auto-  
regressions," Cowles Foundation Discussion Paper No. 765, Yale University.

\_\_\_\_\_ (1985d). "Inference from nonstationary economic time series,"  
Proposal to the National Science Foundation, July, 1985.

\_\_\_\_\_ and S. N. Durlauf (1985). "Multiple time series regression with  
integrated processes," Cowles Foundation Discussion Paper No. 768.