COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 740-R

TIME SERIES REGRESSION WITH UNIT ROOTS

P. C. B. PHILLIPS

April, 1985

Revised: February 1986

TIME SERIES REGRESSION WITH A UNIT ROOT

by

P. ·C. B. Phillips[*]

*Cowles Foundation for Research in Economics
Yale University*

Author's address: Peter C. B. Phillips
                    Box 2125, Yale Station
                    New Haven, Connecticut 06520

    Telephone: 203 436-1893

First Version: March 1985

    Revised: February 1986

0. HEADNOTE

This paper studies the random walk in a general time series setting that allows for weakly dependent and heterogeneously distributed innovations. It is shown that simple least squares regression consistently estimates a unit root under very general conditions in spite of the presence of autocorrelated errors. The limiting distribution of the standardized estimator and the associated regression t-statistic are found using functional central limit theory. New tests of the random walk hypothesis are developed which permit a wide class of dependent and heterogeneous innovation sequences. A new limiting distribution theory is constructed based on the concept of continuous data recording. This theory, together with an asymptotic expansion that is developed in the paper for the unit root case, explain many of the interesting experimental results recently reported in Evans and Savin (1981, 1984).

Running Head: Time Series with a Unit Root

1. INTRODUCTION

Autoregressive time series with a unit root have been the subject of much recent attention in the econometrics literature. In part, this is because the unit root hypothesis is of considerable interest in applications, not only with data from financial and commodity markets where it has a long history but also with aggregate time series. The study by Hall (1978) has been particularly influential with regard to the latter, advancing theoretical support for the random walk hypothesis for consumption expenditure and providing further empirical evidence. Moreover, the research program on vector autoregressive (VAR) modeling of aggregate time series (see Doan et al. (1984) and the references therein) has actually responded to this work by incorporating the random walk hypothesis as a Bayesian prior in the VAR specification. This approach has helped to attenuate the dimensionality problem of VAR modeling and seems to lead to decided improvements in forecasting performance (Litterman (1984)).

At the theoretical level there has also been much recent research. This has concentrated on the distribution theory that is necessary to develop tests of the random walk hypothesis under the null and the analysis of the power of various tests under interesting alternatives. Investigations by Dickey (1976), Dickey and Fuller (1979, 1981), Fuller (1976) and Evans and Savin (1981, 1984) have been at the forefront of this research. Related work on regression residuals has been done by Sargan and Bhargava (1983) and by Bhargava (1986). Recently, attention has also been given to more general ARIMA models by Solo (1984) and by Said and Dickey (1985).

All of the research cited in the previous paragraph has been confined to the case where the sequence of innovations driving the model are

1

independent with common variance. Frequently, it is assumed that the innovations are iid $(0, \sigma^2)$ or, further, that they are iid $N(0, \sigma^2)$. Independence and homoskedasticity are rather strong assumptions to make about the errors in most empirical econometric work; and there are good reasons from economic theory (as shown in Hall (1978)) for believing them to be false in the context of aggregate time series that may be characterized as a random walk. For both empirical and theoretical considerations, therefore, it is important to develop tests for unit roots that do not depend on these conditions.

One aim of the present paper is to develop such tests. In doing so, we provide an asymptotic theory for the least squares regression estimator and the associated regression t-statistic which allows for quite general weakly dependent and heterogeneously distributed innovations. The conditions we impose are very weak and are similar to those used recently by White and Domowitz (1984) in the general nonlinear regression context. However, the limiting distribution theory that we employ here is quite different from that of White and Domowitz (1984). It belongs to a general class of functional limit theory on metric spaces, rather than the central limit theory on Euclidean spaces that is more conventionally used in econometrics (as in the excellent recent treatment by White (1984). Our approach unifies and extends the presently known limiting distribution theory for the random walk and more general ARIMA models with a single unit root. It would seem to allow for most of the data we can expect to encounter in time series regression with aggregate economic series. A particularly interesting feature of the new test statistics that we propose in this paper is that their limiting distributions are identical to those found in earlier work under the assumption of iid errors. Thus, we discover

that much of the work done by the authors cited in the earlier paragraph (particularly Fuller (1976), Dickey and Fuller (1979) and Evans and Savin (1981)) under the assumption of iid errors remains relevant for a very much larger class of models.

Another aim of the paper is to present a new limiting distribuiton theory that is based on the concept of continuous data recording. This theory, together with the asymptotic expansion that is developed in Section 7 of the paper for the unit root case, help to explain many of the interesting experimental results reported in the recent papers by Evans and Savin (1981, 1984) in this journal.

2. FUNCTIONAL LIMIT THEORY FOR DEPENDENT HETEROGENEOUSLY DISTRIBUTED DATA

Let $\{y_t\}_{t=1}^{\infty}$ be a stochastic process generated in discrete time according to:

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

(2) $\qquad \alpha = 1 \; .$

Under (2) we have the representation $y_t = S_t + y_0$ in terms of the partial sum $S_t = \Sigma_1^t u_j$ of the innovation sequence $\{u_j\}$ in (1) and the initial condition y_0. We may define $S_0 = 0$. The three alternatives commonly proposed for y_0 are (c.f. White (1958)):

(3a) $\qquad y_0 = c$, a constant, with probability one;

(3b) $\qquad y_0$ has a certain specified distribution;

(3c) $\qquad y_0 = y_T$, where T = the sample size.

(3c) is a circularity condition, due to Hotelling, that is used mainly as a mathematical device to simplify distribution theory (c.f. Anderson (1942)). (3b) is a random initial condition that is frequently used to achieve stationarity in stable models ($|\alpha| < 1$). In this paper, we shall employ (3b). This permits the greatest flexibility in the specification of (1). It allows for non stationary series (with $|\alpha| \geq 1$) and it includes (3a) as a special case (and, in particular, the commonly used condition $y_0 = 0$).

Our concern in this section will be with the limiting distribution of standardized sums such as:

$$(4a) \qquad X_T(r) = \frac{1}{\sqrt{T}\sigma} S_{[Tr]} = \frac{1}{\sqrt{T}\sigma} S_{j-1} \;; \quad (j-1)/T \leq r < j/T \quad (j = 1, \ldots, T)$$

$$(4b) \qquad X_T(1) = \frac{1}{\sqrt{T}\sigma} S_T$$

where [ ] denotes the integer part of its argument and σ is a certain constant defined later (see Assumption 2.1 below). Observe that the sample paths $X_T(r) \in D = D[0,1]$, the space of all real valued functions on [0,1] that are right continuous at each point of [0,1] and have finite left limits. That is, jump discontinuities (or discontinuities of the first kind) are allowable in D. It will be sufficient for our purpose if we endow D with the uniform metric defined by $\|f-g\| = \sup_r |f(r) - g(r)|$ for any $f, g \in D$.

. $X_T(r)$ is a random element in the function space D. Under certain conditions, $X_T(r)$ can be shown to converge weakly to a limit process which is popularly known either as standard Brownian motion or the Wiener process. This result is often referred to as a *functional central limit theorem* (CLT) (i.e. a CLT on a function space) or as an *invariance principle,*

following the early work of Donsker (1951) and Erdos and Kac (1946). The limit process which we denote by $W(r)$, has sample paths which lie in $C = C[0,1]$, the space of all real valued continuous functions on $[0,1]$. Moreover, $W(r)$ is a Gaussian process (for fixed r $W(r)$ is $N(0,r)$) and has independent increments ($W(s)$ is independent of $W(r) - W(s)$ for all $0 < s < r \leq 1$). We shall denote the weak convergence of the process $X_T(r)$ to $W(r)$ by the notation $X_T(r) \Rightarrow W(r)$; and, when the meaning is clear from the context, we shall sometimes suppress the argument r and simply write $X_T \Rightarrow W$. Here and elsewhere in the paper the symbol "$\Rightarrow$" is used to signify the weak convergence of the associated probability measures as $T \uparrow \infty$. Note that many finite dimensional CLT's follow directly from this result (e.g. the case in which $r = 1$ yields the Lindeberg-Lévy theorem when the u_j are iid $(0, \sigma^2)$). The reader is referred to Billingsley (1968) for a detailed introduction to the subject and to Pollard (1984) for an excellent recent treatment.

The conditions under which $X_T \Rightarrow W$ are very general indeed and extend to a wide class of nonstationary, weakly dependent and heterogeneously distributed innovation sequences $\{u_t\}_1^\infty$. Billingsley (1968, Ch. 4) proves a number of such results for strictly stationary series satisfying weak dependence conditions. His results have recently been extended by many authors in the probability literature (see Hall and Heyde (1980, Ch. 5) for a good discussion of this literature and some related results for martingales and near martingales). Amongst the most general results that have been established are those of McLeish (1975a, 1977) and Herrndorf (1983, 1984a, 1984b). We shall employ a result of Herrndorf (1984b) in our own development because it applies most easily to the weakly dependent and heterogeneously distributed innovations that we wish to allow for in the context of time series such

as (1).

To begin we must be precise about the sequence $\{u_t\}_1^\infty$ of allowable innovations in (1). In what follows we shall assume that $\{u_t\}_1^\infty$ is a sequence of random variables that satisfy:

ASSUMPTION 2.1

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

    (b) $\sup_t E|u_t|^\beta < \infty$ *for some* $\beta > 2$;

    (c) $\sigma^2 = \lim_{T\to\infty} E(T^{-1}S_T^2)$ *exists and* $\sigma^2 > 0$;

    (d) $\{u_t\}_1^\infty$ *is strong mixing with mixing coefficients* α_m *that satisfy:*

(5) $$\Sigma_1^\infty \alpha_m^{1-2/\beta} < \infty .$$

These conditions allow for both temporal dependence and heteroskedasticity in the process $\{u_t\}_1^\infty$. For the definition of strong mixing and the mixing coefficients α_m that appear in (d) the reader is referred, for example, to White (1984). Condition (d) controls the extent of the temporal dependence in the process $\{u_t\}_1^\infty$, so that, although there may be substantial dependence amongst recent events, events which are separated by long intervals of time are almost independent. In particular, the summability requirement (5) on the mixing coefficients is satisfied when the mixing decay rate is $\alpha_m = O(m^{-\lambda})$ for some $\lambda > \beta/(\beta-2)$. The summability condition (5) also controls the mixing decay rate in relation to the probability of outliers as determined by the moment existence condition (b). Thus, as β approaches 2 and the probability of outliers rises (under the weakening moment condition (b)) the mixing decay rate increases and the effect of outliers is required under (5) to wear off more quickly. This tradeoff between moment and mixing conditions was first developed by McLeish (1975b) in the context of strong

laws for dependent sequences. Condition (b) also controls the allowable heterogeneity in the process by ruling out unlimited growth in the β'th absolute moments of u_t .

Condition (c) is a convergence condition on the average variance of the partial sum S_T . It is a common requirement in much central limit theory although it is not strictly a necessary condition (see, for example, Herrndorf (1983)). However, if $\{u_t\}$ is weakly stationary then

$$(6) \qquad \sigma^2 = E(u_1^2) + 2\Sigma_{k=1}^{\infty} E(u_1 u_k)$$

and the convergence of the series is implied by the mixing condition (5) (see Ibragimov and Linnik (1971),Theorem 18.5.3). Even in this case, however, we still require $\sigma^2 > 0$ to exclude degenerate results. Once again, this is a conventional requirement.

Assumption 2.1 allows for a wide variety of possible generating mechanisms for the sequence of innovations $\{u_t\}_1^{\infty}$. These include all Gaussian and many other stationary finite order ARMA models under very general conditions on the underlying errors (see Withers (1981)).

We shall make extensive use of the following two results in our theoretical development. The first is a functional central limit theorem that is due to Herrndorf; and the second is the continuous mapping theorem, which is given a very thorough treatment in Billingsley (1968, Section 5).

LEMMA 2.2. If $\{u_t\}_1^{\infty}$ satisfies Assumption 2.1 then as $T \uparrow \infty$ $X_T \Rightarrow W$, a standard Wiener process on C .

LEMMA 2.3. If $X_T \Rightarrow W$ *as* $T \uparrow \infty$ *and* h *is any continuous functional on* D *(continuous, that is, except for at most a set of points* $D_h \subset D$ *for which* $P(W \in D_h) = 0$ *) then* $h(X_T) \Rightarrow h(W)$ *as* $T \uparrow \infty$

3. LARGE SAMPLE $(T \uparrow \infty)$ ASYMPTOTICS

We denote the ordinary least squares (OLS) estimator of α in (1) by $\hat{\alpha} = \Sigma_1^T y_t y_{t-1} / \Sigma_1^T y_{t-1}^2$. Appropriately centered and standardized we have

$$(7) \qquad T(\hat{\alpha}-1) = \{T^{-1}\Sigma_1^T y_{t-1}(y_t - y_{t-1})\}/\{T^{-2}\Sigma_1^T y_{t-1}^2\}$$

and we shall consider the limiting behavior of this statistic as $T \uparrow \infty$. We shall also consider the conventional regression t-statistic:

$$(8) \qquad t_\alpha = \left(\Sigma_1^T y_{t-1}^2\right)^{1/2} (\hat{\alpha}-1)/s$$

where

$$(9) \qquad s^2 = T^{-1}\Sigma_1^T(y_t - \hat{\alpha}y_{t-1})^2 .$$

Both (7) and (8) have been suggested as test statistics for detecting the presence of a unit root in (1). The distributions of these statistics under both the null hypothesis $\alpha = 1$ and certain alternatives $\alpha \neq 1$ have been studied recently by Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984) and Nankervis and Savin (1985). The work of these authors concentrates altogether on the special case in which the innovation sequence $\{u_t\}_1^\infty$ is iid $(0, \sigma^2)$. In related work, Solo (1984) has studied the asymptotic distribution of the Lagrange multiple (LM) statistic in a general ARIMA setting. His results are also established under the assumption that iid innovations drive the model.

Our approach relies on the theory of weak convergence on D. It leads to rather simple characterizations of the limiting distributions of (7) and (8) in terms of functionals of a Wiener process. The main advantage of the approach is that the results hold for a very wide class of error processes in the model (1).

THEOREM 3.1. *If* $\{u_t\}_1^\infty$ *satisfies Assumption 2.1 and if* $\sup_t E|u_t|^{\beta+\eta} < \infty$ *for some* $\eta > 0$ *(where* $\beta > 2$ *is the same as that in Assumption 2.1), then as* $T \uparrow \infty$:

(a) $T^{-2}\Sigma_1^T y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr$;

(b) $T^{-1}\Sigma_1^T y_{t-1}(y_t - y_{t-1}) \Rightarrow (\sigma^2/2)(W(1)^2 - \sigma_u^2/\sigma^2)$;

(c) $T(\hat{\alpha}-1) \Rightarrow (1/2)(W(1)^2 - \sigma_u^2/\sigma^2)/\int_0^1 W(r)^2 dr$;

(d) $\hat{\alpha} \underset{p}{\to} 1$;

(e) $t_\alpha \Rightarrow (\sigma/2\sigma_u)(W(1)^2 - \sigma_u^2/\sigma^2)/\{\int_0^1 W(r)^2 dr\}^{\frac{1}{2}}$;

where

$$\sigma_u^2 = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(u_t^2) ,$$

$$\sigma^2 = \lim_{T\to\infty} E(T^{-1}S_T^2)$$

and $W(r)$ *is a standard Wiener process on* C.

When the innovation sequence $\{u_t\}_1^\infty$ is iid$(0, \sigma^2)$ we have $\sigma_u^2 = \sigma^2$, leading to the following simplification of part (c) of Theorem 3.1:

(10) $\qquad T(\hat{\alpha}-1) \Rightarrow (1/2(W(1)^2 - 1)/\int_0^1 W(r)^2 dr$.

Result (10) was first given by White (1958, p. 1196), although his expres-

sion is incorrect as stated since his standardization of $\hat{\alpha}$ is $g(T)(\hat{\alpha}-1)$ with $g(T) = T/\sqrt{2}$. Unfortunately, this rather minor error recurs at several points in the paper by Rao (1978, pp. 187-188). Lai and Wei (1982) and Lai and Seigmund (1984) also give (10) as stated above.

Theorem 3.1 extends (10) to the very general case of weakly dependent and heterogeneously distributed data. Interestingly, our result shows that the limiting distribution of $T(\hat{\alpha}-1)$ has the same general form for a very wide class of innovation processes $\{u_t\}_1^\infty$.

The differences between (c) of Theorem 3.1 and (10) may be illustrated with a simple example. Suppose that the generating process for $\{u_t\}$ is the moving average

$$u_t = \epsilon_t + \theta\epsilon_{t-1} , \quad t = 1, 2, \ldots$$

with ϵ_t iid$(0, \sigma_\epsilon^2)$. Then

$$\sigma_u^2 = \text{plim}_{T\to\infty} \, T^{-1}\Sigma_1^T u_t^2 = (1 + \theta^2)\sigma_\epsilon^2 ,$$

$$\sigma^2 = \lim_{T\to\infty} \, T^{-1}E(S_T^2) = (1+\theta)^2\sigma_\epsilon^2$$

and we have

$$T^{-1}\Sigma_1^T y_{t-1}u_t \Rightarrow \frac{\sigma_\epsilon^2}{2}\left[(1+\theta)^2 W(1)^2 - (1 + \theta^2)\right] ,$$

which can also be verified by direct calculation. In this case

$$T(\hat{\alpha}-1) \Rightarrow (1/2)[W(1)^2 - (1 + \theta^2)/(1+\theta)^2]/\int_0^1 W(r)^2 dr$$

generalizing (10) and, of course, reducing to it when $\theta = 0$.

Part (d) of Theorem 3.2 shows that, unlike the stable AR(1) with $|\alpha| < 1$, OLS retains the property of consistency when there is a unit root even in the presence of substantial serial correlation. This extremely simple result seems not to have been derived at this level of generality before, although closely related results for ARIMA models have been obtained recently by Tiao and Tsay (1983) and by Said and Dickey (1984). The robustness of the consistency of $\hat{\alpha}$ in this case is rather extraordinary, allowing for a wide variety of error processes that permit serious misspecifications in the usual random walk formulation of (1) with white noise errors. Intuitively, when the model (1) has a unit root the strength of the signal (as measured by the sample variation of the regressor y_{t-1}) dominates the noise by a factor of $O(T)$, so that the effects of any regressor-error correlation are annihilated in the regression as $T \uparrow \infty$.

Part (e) of Theorem 3.1 gives the limiting distribution of t_α. This distribution, like that of the coefficient estimator, depends on the variance ratio σ_u^2/σ^2. We note that in the Lagrange multiplier approach (c.f. Evans and Savin (1981) and Solo (1984)) we would employ the variance estimator $s'^2 = T^{-1}\Sigma_1^T(y_t - y_{t-1})^2$ in the test statistic, using the null hypothesis (2). Writing the Lagrange multiplier statistic as $LM = t'^2$ where $t' = (\Sigma_1^T y_{t-1}^2)^{1/2}(\hat{\alpha}-1)/s'$ we deduce from part (e) of Theorem 3.1 that

$$(11) \qquad LM \Rightarrow (\sigma/2\sigma_u)^2 [W(1)^2 - \sigma_u^2/\sigma^2]^2 / \int_0^1 W(r)^2 dr .$$

Solo (1984) derived the special case of (11) in which $\sigma^2 = \sigma_u^2$.

Theorem 3.1 provides an interesting example of a functional of a partial sum that does not necessarily converge weakly to the same functional of Brownian motion. To show this, it is most convenient to replace $X_T(r)$

as defined by (4) by its close relative, the random element

$$Y_T(r) = \frac{1}{\sqrt{T}\sigma} S_{[Tr]} + \frac{Tr - [Tr]}{\sqrt{T}\sigma} u_{[Tr]+1} \; ; \quad (j-1)/T \leq r < j/T \quad (j = 1, \ldots, T)$$

$$Y_T(1) = \frac{1}{\sqrt{T}\sigma} S_T$$

which lies in $C[0,1]$. In fact, $Y_T(r) \Rightarrow W(r)$ under the same conditions as those prescribed earlier for $X_T(r)$ in Lemma 2.2. However, the sample path of $Y_T(r)$ is continuous and of bounded variation on $[0,1]$ so that we may define and evaluate by partial integration the following Riemann Stieltjes integral:

$$(12) \qquad \int_0^1 Y_T(r) \, dY(r) = \frac{1}{2}\left[Y_T^2(r)\right]_0^1 = \frac{1}{2}Y_T(1)^2 \; .$$

The corresponding integral for the limit process $W(r)$ must be defined as a stochastic integral, for which the rule of partial integration used in (12) does not apply. Instead, we have the well known result (see, for example, Hida (1980), p. 158):

$$(13) \qquad \int_0^1 W \, dW = (1/2)(W(1)^2 - 1)$$

which may be obtained directly from the Ito formula. On the other hand, we deduce from (12) and Lemma 2.3 that

$$(14) \qquad \int_0^1 Y_T \, dY_T \Rightarrow (1/2)W(1)^2 \; .$$

The problem arises because all elements of $C[0,1]$ except for a set of Wiener measure zero are of unbounded variation (Billingsley (1968), p. 63).

In particular, the sample paths of $W(t)$ are almost surely of unbounded variation and thus the integral $\int_0^1 W dW$ does not exist in the same sense as the integral $\int_0^1 Y_T dY_T$. It follows that the latter integral does not define a continuous mapping on $C[0,1]$ and we cannot appeal to the continuous mapping theorem to deduce that $\int Y_T dY_T \Rightarrow \int W dW$ when $Y_T \Rightarrow W(t)$. In fact, as (13) and (14) demonstrate the result is not correct.

We may, however, proceed as in the proof of Theorem 3.1 in the Appendix. Alternatively, since $dY_T(t) = \sqrt{T} u_j dt / \sigma$ we find by direct integration that:

$$\int_{(j-1)/T}^{j/T} Y_T dY_T = \frac{1}{T\sigma^2} S_{j-1} u_j + \frac{u_j^2}{2T\sigma^2}$$

and summing over $j = 1, \ldots, T$ we deduce that:

$$(15) \qquad T^{-1} \Sigma_1^T S_{j-1} u_j = \sigma^2 \int_0^1 Y_T dY_T - \Sigma_1^T u_j^2 / 2T \Rightarrow (\sigma^2/2) W(1)^2 - \sigma_u^2 / 2$$

as given by part (b) of Theorem 3.1. Note also, in view of (13), that the sum (15) converges to $\sigma^2 \int_0^1 W dW$ if and only if $\sigma_u^2 = \sigma^2$.

4. ESTIMATION OF (σ_u^2, σ^2)

The limiting distributions given in Theorem 3.1 depend on unknown parameters σ_u^2 and σ^2 . These distributions are therefore not directly useable for statistical testing. However, both these parameters may be consistently estimated and the estimates may be used to construct modified statistics whose limiting distributions are independent of (σ_u^2, σ^2) . As we shall see, these new statistics (given below by (21) and (22)) provide very general

tests for the presence of a unit root in (1).

As shown in the proof of Theorem 3.1 $T^{-1}\Sigma_1^T u_t^2 \rightarrow \sigma_u^2$ a.s. as $T \uparrow \infty$. This provides us with the simple estimator

(16) $\qquad s_u^2 = T^{-1}\Sigma_1^T (y_t - y_{t-1})^2 = T^{-1}\Sigma_1^T u_t^2$,

which is consistent for σ_u^2 under the null hypothesis (2). Since $\hat{\alpha} \underset{p}{\rightarrow} 1$ by Theorem 3.1 we may also use $T^{-1}\Sigma_1^T (y_t - \hat{\alpha} y_{t-1})^2$ as a consistent estimator of σ_u^2.

Consistent estimation of $\sigma^2 = \lim_{T \to \infty} E(T^{-1}S_T^2)$ is more difficult. The problem is essentially equivalent to the consistent estimation of an asymptotic covariance matrix in the presence of weakly dependent and heterogeneously distributed observations.[1] The latter problem has recently been examined by White and Domowitz (1984). A detailed treatment is also available in Chapter VI of White (1984).

We start by defining

$$\sigma_T^2 = \text{var}(T^{-1/2}S_T)$$

$$= T^{-1}\Sigma_1^T E(u_t^2) + 2T^{-1}\sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^{T} E(u_t u_{t-\tau})$$

and by introducing the approximant

$$\sigma_{T\ell}^2 = T^{-1}\Sigma_1^T E(u_t^2) + 2T^{-1}\sum_{\tau=1}^{\ell} \sum_{t=\tau+1}^{T} E(u_t u_{t-\tau}) .$$

---

[1]This is most easily seen by noting that $S_T/\sqrt{T} \Rightarrow N(0, \sigma^2)$, according to the invariance principle Lemma 2.2.

We shall call ℓ the lag truncation number. For large T and large $\ell < T$, $\sigma_{T\ell}^2$ may be expected to be very close to σ_T^2 if the total contribution in σ_T^2 of covariances such as $E(u_t u_{t-\tau})$ with long lags $\tau > \ell$ is small. This will be true if $\{u_t\}_1^\infty$ satisfies Assumption 2.1. Formally, we have:

LEMMA 4.1. If the sequence $\{u_t\}_1^\infty$ satisfies Assumption 2.1 and if $\ell \uparrow \infty$ as $T \uparrow \infty$ then $\sigma_T^2 - \sigma_{T\ell}^2 \to 0$ as $T \uparrow \infty$.

This Lemma suggests that under suitable conditions on the rate at which $\ell \uparrow \infty$ as $T \uparrow \infty$ we may proceed to estimate σ^2 from finite samples of data by sequentially estimating $\sigma_{T\ell}^2$. The problem is explored by White (1984, Ch. 6). We define

$$(17) \qquad s_{T\ell}^2 = T^{-1} \Sigma_1^T u_t^2 + 2T^{-1} \Sigma_{\tau=1}^\ell \Sigma_{t=\tau+1}^T u_t u_{t-\tau} .$$

The following result establishes that $s_{T\ell}^2$ is a consistent estimator of σ^2.

THEOREM 4.2. If

(a) $\{u_t\}_1^\infty$ *satisfies Assumption 2.1(a), (c) and (d) and part (b) of Assumption 2.1 is replaced by the stronger moment condition:*

$$\sup_t E|u_t|^{2\beta} < \infty , \quad \text{for some } \beta > 2 ;$$

(b) $\ell \uparrow \infty$ *as* $T \uparrow \infty$ *such that* $\ell = o(T^{1/4})$; *then* $s_{T\ell}^2 \underset{p}{\to} \sigma^2$ *as* $T \uparrow \infty$.

According to this result, if we allow the number of estimated autocovariances to increase as $T \uparrow \infty$ but control the rate of increase so that $\ell = o(T^{1/4})$ then $s_{T\ell}^2$ yields a consistent estimator of σ^2. White and Domowitz (1984) provide some guidelines for the selection of ℓ. Inevit-

ably the choice of ℓ will be an empirical matter. In our own case, a preliminary investigation of the sample autocorrelations of $u_t = y_t - y_{t-1}$ will help in selecting an appropriate choice of ℓ . Since the sample auto-correlations of first differenced economic time series usually decay quickly it is likely that in moderate sample sizes quite a small value of ℓ will be chosen.

Rather than using the first differences $u_t = y_t - y_{t-1}$ in the construction of $s^2_{T\ell}$, we could have used the residuals $\hat{u}_t = y_t - \hat{\alpha} y_{t-1}$ from the least squares regression. Since $\hat{\alpha} \underset{p}{\to} 1$ as $T \uparrow \infty$ this estimator is also consistent for σ^2 under the null hypothesis (2). Moreover, this estimator is consistent for σ^2 under explosive alternatives to (2) (i.e. when $\alpha > 1$) and may, therefore, be preferred to $s^2_{T\ell}$ when such cases seem likely.

We remark that $s^2_{T\ell}$ is not constrained to be non negative as it is presently defined in (17). When there are large negative sample serial co-variances, $s^2_{T\ell}$ can take on negative values. In a related context, Newey and West (1985) have recently suggested a modification to variance estimators such as $s^2_{T\ell}$ which ensures that they are non negative. In the present case, the modification yields:

(18) $\qquad \tilde{s}^2_{T\ell} = T^{-1} \Sigma^T_1 u^2_t + 2T^{-1} \Sigma^\ell_{\tau=1} w_{\tau\ell} \Sigma^T_{t=\tau+1} u_t u_{t-\tau}$

where

(19) $\qquad w_{\tau\ell} = 1 - \tau/(\ell+1)$.

It is simple to motivate the weighted variance estimator (18). When $\{u_t\}^\infty_1$ is weakly stationary, $\sigma^2 = 2\pi f_u(0)$ where $f_u(\lambda)$ is the spectral density

of u_t. In this case, $(1/2\pi)\tilde{s}_{T\ell}^2$ is the value at the origin $\lambda = 0$ of the Bartlett estimate

$$(20) \quad \hat{f}_u(\lambda) = (1/2\pi)\Sigma_{\tau=-\ell-1}^{\ell+1}[1 - |\tau|/(\ell+1)]C(\tau)e^{-i\lambda\tau} \;, \quad C(\tau) = T^{-1}\Sigma_{t=|\tau|+1}^{T}u_t u_{t-|\tau|}$$

of $f_u(\lambda)$ (see, for example, Priestley (1981), pp. 439-440). Since the Bartlett estimate (20) is non negative everywhere, we deduce that $\tilde{s}_{T\ell}^2 \geq 0$ also. Of course, weights other than (19) are possible and may be inspired by other choices of lag window in the density estimate (20).

5. NEW TESTS FOR A UNIT ROOT

The consistent estimates s_u^2 and $s_{T\ell}^2$ may be used to develop new tests for unit roots that apply under very general conditions. We define the statistics:

$$(21) \quad Z_\alpha = T(\hat{\alpha}-1) - (1/2)(s_{T\ell}^2 - s_u^2)/(T^{-2}\Sigma_1^T y_{t-1}^2)$$

and

$$(22) \quad Z_t = (\Sigma_1^T y_{t-1}^2)^{1/2}(\hat{\alpha}-1)/s_{T\ell} - (1/2)(s_{T\ell}^2 - s_u^2)\left[s_{T\ell}(T^{-2}\Sigma_1^T y_{t-1}^2)^{1/2}\right]^{-1} .$$

Z_α is a transformation of the standardized estimator $T(\hat{\alpha}-1)$ and Z_t is a transformation of the regression t-statistic (8).

The limiting distributions of Z_α and Z_t are given by:

THEOREM 5.1. *If the conditions of* Theorem 4.2 *are satisfied then as* $T \uparrow \infty$:

$$(a) \quad Z_\alpha \Rightarrow \frac{(W(1)^2 - 1)/2}{\int_0^1 W(t)^2 dt}$$

and

$$(b) \quad Z_t \Rightarrow \frac{(W(1)^2 - 1)/2}{\left\{\int_0^1 W(t)^2 dt\right\}^{1/2}}$$

under the null hypothesis that $\alpha = 1$ *in* (1).

Theorem 5.1 demonstrates that the limiting distributions of the two statistics Z_α and Z_t are invariant within a very wide class of weakly dependent and possibly heterogeneously distributed innovations $\{u_t\}_1^\infty$.

Moreover, the limiting distribution of Z_α is identical to that of $T(\hat{\alpha}-1)$ when $\sigma_u^2 = \sigma^2$ (see (10) above). The latter distribution has recently been computed by Evans and Savin (1981) using numerical methods. These authors present tabulations and graphical plots of the limiting pdf and their article also contains a detailed tabulation of the limiting cdf, which is suitable for testing purposes. Since Evans and Savin work with the normalization $g(T)(\hat{\alpha}-1)$, in which $g(T) = T/\sqrt{2}$, the modified statistic

$$(23) \qquad Z_\alpha' = (1/\sqrt{2})Z_\alpha$$

may be used to ensure compatibility with their published tables. Fuller (1976, p. 371) provides a tabulation of the limiting distribution (10) for the standardization $T(\hat{\alpha}-1)$, so that his table may be used directly in significance testing with our statistic Z_α.

The limiting distribution of Z_t given in Theorem 5.1 is identical to that of the regression t-statistic when $\sigma^2 = \sigma_u^2$ (see Theorem 3.1). This is, in fact, the limiting distribution of the t-statistic when the innovation sequence $\{u_t\}_1^\infty$ is iid $(0, \sigma^2)$. The latter distribution

has been calculated using Monte Carlo methods by Dickey (1976) and tabulations of percentage points of the distribution are reported in Fuller (1976, table 8.5.2, p. 373).

Theorem 5.1 shows that much of the work of these authors on the distribution of the OLS estimator $\hat{\alpha}$ and the regression t-statistic under iid innovations remains relevant for a very much larger class of models. In fact, our results show that their tabulations appear to be relevant in almost any time series with a unit root. To test the unit root hypothesis (2) one simply computes (21), (23) or (22) and compares these to the relevant critical values given by Evans and Savin (1981) and Fuller (1976).

6. CONTINUOUS RECORD ASYMPTOTICS

In certain econometric applications a near-continuous record of data is available for empirical work. Prominent examples occur in various financial, commodity and stock markets as well as in certain recent energy usage experiments. Undoubtedly, trends in this direction will accelerate in the next decade with ongoing computerizations of banking and credit facilities and electronic monitoring of sales activity. Moreover, financial and foreign exchange markets, in particular, now offer empirical researchers the opportunity of working with data recorded at many different frequencies (weekly, daily, hourly or even minute by minute in some cases). For these reasons, it is of intrinsic interest to study the behavior of econometric estimators and test statistics as the time interval (h) between sampled observations is allowed to vary and, possibly, to tend to zero. When $h \downarrow 0$, we obtain in the limit a continuous record of observations over a finite time span, comparable to a seismographic recording. We shall call

asymptotics of this type continuous record asymptotics.[1]

As we shall show below, there is a very interesting relationship, at least in certain cases, between the behavior of the statistics we have been considering when the sample size $T \uparrow \infty$ and when the sampling interval $h \downarrow 0$. This relationship, together with the results of Section 7, help to explain many of the recent Monte Carlo results reported by Evans and Savin (1981, 1984). Moreover, continuous record $(h \downarrow 0)$ asymptotics have an additional advantage in that they bring into prominence the role of initial observation conditions. Such conditions can be of considerable importance in the statistical behavior of certain econometric estimators and tests in finite samples of data. Yet their effects normally disappear entirely in conventional large sample $(T \uparrow \infty, h$ fixed) asymptotic theory. We shall illustrate the effects of such initial conditions in the context of autoregressions such as (1) in the presence of a unit root.

We start by considering a triangular array of random variables $\{\{y_{nt}\}_{t=1}^{T_n}\}_{n=1}^{\infty}$ defined as follows. Given n, the sequence $\{y_{nt}\}_{1}^{T_n}$ is generated by the random walk

(24) $y_{nt} = y_{nt-1} + u_{nt}$; $t = 1, \ldots, T_n$; $y_{n0} = y(0)$

where the innovations $\{u_{nt}\}_{t=1}^{T_n}$ are iid $(0, \sigma^2 h_n)$ are independent of $y(0)$ and $T_n h_n = N$ is a fixed positive constant for all n. Moreover, as $n \uparrow \infty$ we shall require $T_n \uparrow \infty$ and $h_n \downarrow 0$ so that $T_n h_n = N$ remains

---

[1]Note that earlier work on continuous time econometric modeling (see, for instance, Bergstrom (1984) and the articles published in Bergstrom (1976)) used small sampling interval $(h \downarrow 0)$ methods in a different context: viz., to compare various estimation procedures by considering how their conventional $(T \uparrow \infty)$ asymptotic properties differed as $h \downarrow 0$.

constant and $T_n \in Z^+$. The sequence $\{\{u_{nt}\}_1^{T_n}\}_1^\infty$ will be called a triangular array of iid$(0, \sigma^2 h_n)$ variates and $\{\{y_{nt}\}_1^{T_n}\}_1^\infty$ will be called a triangular array of random walks.

Each row of the triangular array $\{\{y_{nt}\}_1^{T_n}\}_1^\infty$ may be interpreted as a sequence of random variables generated by a random walk in discrete time with a sampling interval $= h_n$. The array itself represents a sequence of random walks with sampling intervals that decrease $(h_n\downarrow)$ as we get deeper $(n\uparrow)$ into the array. The convergence of the sampling interval $h_n \downarrow 0$ as $n \uparrow \infty$ and the requirement that $h_n T_n = N$ be fixed are the only connections that link the random variables in different rows of the array. The interval $[0,N]$ may be regarded as a fixed time span over which we observe the random walk at discrete points in time determined by the sampling interval h_n . The triangular array $\{\{y_{nt}\}_1^{T_n}\}_1^\infty$ then provides a formal framework within which h_n may vary and by means of which we may investigate limiting behavior as $h_n \downarrow 0$.

Let $S_{ni} = \Sigma_{j=1}^i u_{nj}$ $(1 \leq i \leq T_n)$ with $S_{n0} = 0$ as usual. We form the random function

$$Y_n(r) = \sigma^{-1} S_{ni-1} , \quad (i-1)/T_n \leq r < i/T_n \quad (i = 1, \ldots, T_n)$$

$$Y_n(1) = \sigma^{-1} S_{nT_n}$$

and observe that $Y_n \in D$. As $n \uparrow \infty$ $Y_n(r)$ converges weakly to a constant multiple of a standard Wiener process. Specifically, we have:

LEMMA 6.1. If

(a) $\{\{u_{nt}\}_1^{T_n}\}_1^\infty$ *is a triangular array of* iid$(0, \sigma^2 h_n)$ *variates;*

(b) $T_n \in Z^+$, $T_n \uparrow \infty$ *and* $h_n \downarrow 0$ *as* $n \uparrow \infty$ *in such a way that the product* $T_n h_n = N > 0$ *remains constant; then* $Y_n(r) \Rightarrow N^{1/2}W(r)$ *as* $n \uparrow \infty$ *where* $W(r)$ *is a standard Wiener process.*

Let $\hat{\alpha}_n = \Sigma_1^{T_n} y_{nt} y_{nt-1} / \Sigma_1^{T_n} y_{nt-1}^2$ be the coefficient and

$t_{\alpha_n} = (\Sigma_1^{T_n} y_{nt-1}^2)^{1/2} (\hat{\alpha}_n - 1)/s_n$ be the associated t-ratio in a least squares

regression on (24). Here, $s_n = \{T_n^{-1} \Sigma_1^{T_n} (y_{nt} - \hat{\alpha}_n y_{nt-1})^2\}^{1/2}$ is the standard

error of the regression. The limiting behavior of these statistics as

$h_n \downarrow 0$ is given in our next result.

THEOREM 6.2. If $\{\{y_{nt}\}_1^{T_n}\}_1^\infty$ *is a triangular array of random walks for which the innovation sequence* $\{\{u_{nt}\}_1^{T_n}\}_1^\infty$ *satisfies the conditions of Lemma 6.1 then as* $n \uparrow \infty$:

(a) $h_n \Sigma_1^{T_n} y_{nt-1}^2 \Rightarrow N^2 \sigma^2 \{\int_0^1 W(r)^2 dr + 2(y(0)/\sigma N^{1/2}) \int_0^1 W(r) dr + y(0)^2/\sigma^2 N\}$;

(b) $\Sigma_1^{T_n} y_{nt-1}(y_{nt} - y_{nt-1}) \Rightarrow (N\sigma^2/2)\{W(1)^2 - 1 + 2(y(0)/\sigma N^{1/2})W(1)\}$;

(c) $h_n^{-1}(\hat{\alpha}_n - 1) \Rightarrow (1/N)[\int_0^1 W(r)^2 dr + 2(y(0)/\sigma N^{1/2}) \int_0^1 W(r) dr + y(0)^2/\sigma^2 N]^{-1}$

$\cdot [(1/2)(W(1)^2 - 1) + (y(0)/\sigma N^{1/2})W(1)]$;

(d) $t_{\alpha_n} \Rightarrow [\int_0^1 W(r)^2 dr + 2(y(0)/\sigma N^{1/2}) \int_0^1 W(r) dr + y(0)^2/\sigma^2 N]^{-1/2}$

$\cdot [(1/2)(W(1)^2 - 1) + (y(0)/\sigma N^{1/2})W(1)]$;

where $W(r)$ *is a standard Wiener process.*

Theorem 6.2 shows that for small h_n the distribution of $\hat{\alpha}_n$ and

that of t_{α_n} may be approximated by suitable functionals of Brownian motion.

These functionals involve the initial condition $y(0)$, which may be either constant or random. If $y(0)$ is random then it is independent of the Wiener process $W(r)$ that appears in the functionals given in parts (c) and (d) (recall that $y(0)$ is independent of innovation sequence $\{\{u_{nt}\}_1^{T_n}\}_1^\infty$). For large N the distributions may be well approximated by:

$$(25) \qquad h_n^{-1}(\hat{\alpha}_n - 1) \sim N^{-1}[\int_0^1 W(r)^2 dr]^{-1}[(1/2)(W(1)^2 - 1)] \ ;$$

$$(26) \qquad t_{\alpha_n} \sim [\int_0^1 W(r)^2 dr]^{-\frac{1}{2}}[(1/2)(W(1)^2 - 1)] \ .$$

(25) and (26) correspond, as we would expect, to the conventional large sample $(T \uparrow \infty)$ asymptotics and are special cases of our earlier results in Theorem 3.1 with $\sigma^2 = \sigma_u^2$.

When the time span N is not large, (25) and (26) may not be good approximations. Theorem 6.2 suggests that the initial value $y(0)$ or, more specifically, the ratio $c = y(0)/\sigma$ plays an important role in the determining of the adequacy of these approximations. Thus, when c is large the effect on the limiting distributions given in parts (c) and (d) of Theorem 6.2 is substantial. In fact as $c \uparrow \infty$ it is easy to deduce from these expressions that:

$$(27) \qquad h_n^{-1}(\hat{\alpha}_n - 1) \sim (1/cN^{\frac{1}{2}})W(1) \equiv N(0, 1/c^2 N) \ ;$$

$$(28) \qquad t_{\alpha_n} \sim W(1) \equiv N(0,1) \ .$$

Theorem 6.2 helps to explain several of the phenomena discovered in the experimental investigation of Evans and Savin (1981). These authors

found: (i) that the finite sample distribution of $\hat{\alpha}$ was very well approx-
imated by its asymptotic distribution (using conventional large sample
$(T \uparrow \infty)$ asymptotics with h fixed) even for quite small samples when the
initial value $y(0) = 0$; and (ii) that changes in $c = y(0)/\sigma$ precipi-
tate substantial changes in the distribution of $\hat{\alpha}$; specifically, the
distribution of $T(\hat{\alpha}-1)$ noticeably concentrates as c increases.

Observation (ii) is well explained by Theorem 6.2, which shows that
c is an important parameter in the limiting distribution of $h_n^{-1}(\hat{\alpha}_n - 1)$
as $h_n \downarrow 0$ over finite data spans $[0,N]$. This is to be contrasted with
the usual $(T \uparrow \infty)$ asymptotic theory, which obscures the dependence of
the distribution of $\hat{\alpha}$ on $y(0)/\sigma$. The second phenomenon noted by Evans-
Savin, that the distribution of $T(\hat{\alpha}-1)$ concentrates as c increases, is
directly corroborated by (27). Thus, using $N = T_n h_n$, we deduce from
(27) that:

$$T_n(\hat{\alpha}_n - 1) \sim (N^{1/2}/c)W(1) \equiv N(0, N/c^2) \underset{p}{\rightarrow} 0 , \quad \text{as} \quad c \uparrow \infty .$$

The fact that the asymptotic $(T \uparrow \infty)$ distribution of $\hat{\alpha}$ is a very
good approximation in finite samples when $y(0) = 0$ and the innovations
are iid (Evans-Savin observation (i) above) is also well explained by our
analytical results. In particular, Theorem 6.2 shows that the asymptotic
distribution applies not only as $T \uparrow \infty$ in the conventional sense with
h fixed (our Theorem 3.1(c) with $\sigma^2 = \sigma_u^2$) but also as $h_n \downarrow 0$ with a
fixed data span N (our Theorem 6.2(c) with $y(0) = 0$). Thus, the
limiting distribution theory operates in two different directions with
identical results when $y(0) = 0$ and the innovations are iid. Moreover,
this limiting distribution is actually the finite sample distribution of
the least squares continuous record estimator when the (continuous) stochastic

process is Gaussian.

To show this we observe that the natural limit of (24) as $n \uparrow \infty$ may be regarded as the stochastic differential equation

$$(29) \qquad dy(t) = \theta y(t)dt + \zeta(dt) \ , \quad (0 \le t \le N)$$

with $\theta = 0$. In (29) $y(t)$ is now a random function of continuous time over $[0,N]$ and $\zeta(dt)$ is a σ-additive random measure which is defined on all subsets of the real line with finite Lebesgue measure such that:

$$(30) \qquad E(\zeta(dr)) = 0 \ , \quad E(\zeta(dr)^2) = \sigma^2 dr$$

(see, for example, Rozanov (1967), p. 8). The stochastic integral $\int_0^t \zeta(dr)$ has zero mean, variance $= \sigma^2 t$ and uncorrelated increments. We note that (24) may be regarded as a random walk in discrete time-that is satisfied by equispaced observations over intervals of length h_n generated by (29) with $\theta = 0$. Then we may write, for example, $u_{nt} = \int_{th_n - h_n}^{th_n} \zeta(dr)$ and $\{u_{nt}\}_1^{T_n}$ is an orthogonal sequence with common variance $\sigma^2 h_n$. If we go further in the specification of (29) and require that the increments of $\int_0^t \zeta(dr)$ are independent then the continuous stochastic process $y(t)$ defined by (29) is Gaussian and $\sigma^{-1} \int_0^t \zeta(dr)$ $(0 \le t \le N)$ is a Wiener process on $C[0,N]$. This is proved by Billingsley (1968, Theorem 19.1, p. 154) and is also a consequence of our Lemma 6.1. In particular, from Lemma 6.1 we have $Y_n(r) \Rightarrow N^{1/2} W(r)$ $(0 \le r \le 1)$ as $n \uparrow \infty$. Now set $t = Nr$ $(0 \le t \le N)$, $y_n(t) = \sigma Y_n(t/N)$ and define $V(t) = N^{1/2} W(t/N)$. Clearly, $V(t)$ is a Wiener process on $C[0,N]$. Moreover, as $n \uparrow \infty$ $y_n(t) \Rightarrow \sigma V(t) \equiv y(t)$ $(0 \le t \le N)$, where "$\equiv$" signifies equality in distribution. Thus, the triangular array $\{\{y_{nt}\}_1^{T_n}\}_1^\infty$ $(T_n h_n = N$, $y_{n0} = 0)$ converges weakly as

$n \uparrow \infty$ to the continuous Gaussian process $y(t)$ $(0 \leq t \leq N)$ generated by the stochastic differential equation (29) with $\theta = 0$, $y(0) = 0$.

If we now consider the problem of estimating the parameter θ in (29) from the continuous record $\{y(t); 0 \leq t \leq N\}$ least squares suggests the criterion

$$(31) \qquad \min_{\theta} \int_0^N (\dot{y} - \theta y)^2 dt .$$

Here we write $\dot{y} = dy(t)/dt$ in a purely formal way since this is all that is needed for our present purpose. (31) leads to the estimator:

$$(32) \qquad \hat{\theta} = \int_0^N y\dot{y}\,dt / \int_0^N y^2 dt = \int_0^N y\,dy / \int_0^N y^2 dt .$$

Provided the integral $\int_0^N y\,dy$ in the numerator of (32) is interpreted as a stochastic integral it is not necessary to be more specific about the interpretation of $\dot{y}$ as a generalized stochastic process.

The estimator $\hat{\theta}$ was originally suggested by Bartlett (1946). Its properties were subsequently studied by Grenander (1950) and more recently by Brown and Hewitt (1975) and by Feigin (1979). As our next result shows the distribution of $\hat{\theta}$ can be simply expressed in terms of the initial value $y(0)$ and a standard Wiener process.

THEOREM 6.3. If $y(t)$ $(0 \leq t \leq N)$ *is generated by* (29) *where* $\zeta(dr)$ *is Gaussian and independent of* $y(0)$, *then* $\hat{\theta}$ *has the same distribution in finite samples (i.e. finite* N *) as the functional*

$$(33) \quad (1/N) \left[\int_0^1 W(r)^2 dr + 2(y(0)/\sigma N^{1/2}) \int_0^1 W(r)\,dr + y(0)^2/\sigma^2 N \right]^{-1}$$

$$\cdot \left[(1/2)(W(1)^2 - 1) + (y(0)/\sigma N^{1/2}) W(1) \right]$$

of the standard Wiener process $W(r)$.

This verifies the result we stated earlier: viz., the limiting distribution of $h_n^{-1}(\hat{\alpha}_n - 1)$ as $n \uparrow \infty$ (Theorem 6.2(c)) is the same as the finite sample (fixed N) distribution of the continuous record estimator $\hat{\theta}$.

Since the triangular array $\{\{y_{nt}\}_1^{T_n}\}_1^\infty$ converges weakly to the Gaussian process $y(t)$ as $n \uparrow \infty$ and since both estimators $\hat{\alpha}_n$ and $\hat{\theta}$ are obtained by least squares, this is a result that may well have been anticipated. Interestingly, while the limiting distribution of $h_n^{-1}(\hat{\alpha}_n - 1)$ and the distribution of $\hat{\theta}$ are equivalent, the distribution of $\hat{\alpha}_n$ is degenerate in the limit as $n \uparrow \infty$. In fact, $\hat{\alpha}_n \underset{p}{\to} 1$ as $n \uparrow \infty$ and $h_n \downarrow 0$. $\hat{\theta}$, on the other hand, is a consistent estimator of $\theta = 0$ only as $N \uparrow \infty$. The consistency of $\hat{\alpha}_n$ is explained by the fact that, although there is not an infinite span of data (N is fixed), there is, in the limit as $n \uparrow \infty$, an infinite amount of independent incremental data on the random walk (24) (and, hence, on the coefficient $\alpha = 1$) because the triangular array $\{\{u_{nt}\}_1^{T_n}\}_1^\infty$ has iid innovations in every row. This is sufficient to ensure that $\hat{\alpha}_n \underset{p}{\to} 1$ as $n \uparrow \infty$.

7. AN ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF $T(\hat{\alpha}-1)$

A general refinement (to higher order) of the functional central limit theorem discussed in Section 2 does not yet appear to be available in the probability literature. However, it is relatively easy to develop asymptotic expansions in special cases such as the limit Theorem 3.1. In order to proceed one needs to endow the random sequence $\{u_t\}_1^\infty$ with somewhat stronger properties than those of Assumption 2.1. To facilitate our analysis here, we shall consider the special case in which the u_t are iid $N(0, \sigma^2)$. In this case it is easy to see that, for fixed $r \in [0,1]$, $X_T(r)$ is

$N(0, [Tr]/T)$. In fact, when $r = 0$ we have $X_T(0) = W(0) = 0$ and when $0 < r \leq 1$ we have:

$$X_T(r) = W(r)([Tr]/Tr)^{1/2}$$

$$= W(r)\{1 - (Tr - [Tr])/Tr\}^{1/2}$$

$$(34) \qquad \equiv W(r)\left\{1 - \frac{1}{2}\frac{Tr - [Tr]}{Tr}\right\} + 0_p(T^{-2})$$

where, as before, "$\equiv$" signifies equality in distribution. (34) provides a simple asymptotic expansion for the finite dimensional (in fact, one dimensional) distribution of $X_T(r)$ with r fixed. Note that, since $W(r)/r \xrightarrow[a.s.]{} 0$ as $r \downarrow 0$ (see, for example, Hida (1980), p. 57), the expansion (34) remains well defined in the neighborhood of $t = 0$. Higher order finite dimensional distributions of $X_T(r)$ may be treated in a similar way. The error on the approximation $X_T(r) \sim W(r)$ is seen to be of $0_p(T^{-1})$ in (34). This suggests that certain functionals of $X_T(r)$ may be expected to differ from the same functionals of $W(r)$ by quantities of the same order. In particular

$$(35) \qquad \int_0^1 X_T(r)^2 dr \equiv \int_0^1 W(r)^2 dr + 0_p(T^{-1})$$

and this expansion may be verified directly by developing an expansion for the characteristic function of $\int_0^1 X_T(r)^2 dr$. Since the algebra is lengthy we shall not report it here.

Our main concern is to develop an expansion for the distribution of $T(\hat{a}-1)$. We therefore consider next the numerator of this statistic, viz.

$$(36) \qquad T^{-1}\Sigma_1^T y_{t-1}(y_t - y_{t-1}) = (\sigma^2/2)X_T(1)^2 - (2T)^{-1}\Sigma_1^T u_i^2 + y_0\bar{u}$$

using formula (A3) from the Appendix. We shall confine our attention to the case where the initial value $y_0 = 0$. Since $X_T(1)^2 \equiv W(1)^2 + O_p(T^{-1})$ (36) becomes (in distribution):

$$(\sigma^2/2)(W(1)^2 - 1) - (1/2T)\Sigma_1^T(u_i^2 - \sigma^2) + O_p(T^{-1})$$

$$\equiv (\sigma^2/2)(W(1)^2 - 1) - (\sigma^2/\sqrt{2T})\xi + O_p(T^{-1})$$

where $\xi \equiv N(0,1)$ and is independent of the Wiener process $W(r)$. The distribution of ξ follows directly from the Lindeberg-Levy theorem. Note that ξ is dependent on a quadratic function of the u_i , whereas $W(r)$ depends on partial sums which are linear in the u_i . Hence, ξ and $W(r)$ are uncorrelated and, being normal, are therefore independent. We deduce the following result.

THEOREM 7.1. *If* y_t *is generated from the random walk* (1) *with* $\alpha = 1$ *and initial value* $y_0 = 0$ *and if the* u_i *are iid* $N(0, \sigma^2)$ *then*

$$(37) \qquad T(\hat{\alpha}-1) \equiv \frac{(1/2)(W(1)^2 - 1) - (1/\sqrt{2T})\xi}{\int_0^1 W(r)^2 dr} + O_p(T^{-1})$$

where $W(r)$ *is a standard Wiener process and* ξ *is* $N(0,1)$ *and independent of* $W(r)$.

(37) provides the first term in the asymptotic expansion of the distribution of $T(\hat{\alpha}-1)$ about its limiting distribution. We observe that the term of $1/\sqrt{T}$ in this expansion contributes no adjustment to the mean of the limiting distribution. This is to be contrasted with the expansion of the distribution of $\sqrt{T}(\hat{\alpha}-\alpha)$ when $|\alpha| < 1$ that was obtained in earlier work by the author (1977). In the latter case the mean adjustment of the

$O(1/\sqrt{T})$ term in the expansion was substantial for α less than but close to unity.

The expansion (37) suggests that the location of the limiting distribution should be an accurate approximation in moderate samples. This is confirmed in the results of the sampling experiment in Evans and Savin (1981). It will be of interest to discover the extent to which (37) improves upon the asymptotic distribution of $\hat{\alpha}$ at various finite sample sizes. The numerical computations that are necessary to explore this question will be performed at a later date.

8. CONCLUSION

The model (1) and (2) that we have considered above is much more general than it may appear. It applies, for example, to virtually any ARMA model with a unit root and even ARMAX systems with a unit root and with stable exogenous processes that admit a Wold decomposition. In the former case, we may write

$$(38) \qquad a(L)(1-L)y_t = b(L)e_t$$

for given finite order lag polynomials $a(L)$ and $b(L)$ in the lag operator L. Then, upon inversion, (38) becomes

$$(39) \qquad y_t = y_{t-1} + u_t , \quad u_t = a(L)^{-1}b(L)e_t$$

and u_t will satisfy the weak dependence and heteroskedasticity conditions of Assumption 2.1 under very general conditions on the innovations and lag polynomials of (38). In the latter case, we may write

$$(40) \qquad a(L)(1-L)y_t = b(L)x_t + c(L)e_t$$

with

$$d(L)x_t = f(L)v_t$$

and then upon inversion we have

$$y_t = y_{t-1} + u_t \ , \quad u_t = a(L)^{-1}c(L)e_t + a(L)^{-1}b(L)d(L)^{-1}f(L)v_t$$

which is once again of the form (1) with u_t satisfying the required assumptions under general conditions on e_t , v_t and the lag polynomials.

Our results show that in quite complicated time series models such as (38) and (40) it is not necessary to estimate the model or even to identify the model in order to consistently estimate or test for a unit root in the time series. One needs only to construct the first order serial correlation coefficient and associated test statistics (21) or (22) and use the appropriate limiting distributions given in Theorems 3.1 and 5.1 for statistical testing. This approach applies under conditions that are of even wider applicability than the models (38) and (40). In a certain sense, this general idea is already implicit in the Box-Jenkins modeling approach. However, none of the traditional theory in this research (given, for example, by Box and Jenkins (1976) or Granger and Newbold (1977)) allows for estimation or testing procedures that have anything approaching the range of applicability of the approach developed here.

The research reported in this paper is currently being extended in various ways. First, the methods that we have developed make it very easy to perform similar analyses on models like (1) with a drift and a time trend. The new tests for the presence of a unit root given here may also

be extended to such models. Second, multivariate generalizations of time series models such as (1) may also be studied by our methods. This generalization opens the way to a detailed asymptotic analysis of nonstationary vector autoregressions, spurious regressions of the type considered by Granger and Newbold (1974, 1977) and cointegrating regressions of the type advanced recently by Granger and Engle (1985). Third, the asymptotic local power properties of the tests developed herein and those of other authors such as Dickey and Fuller (1979, 1981) may be studied by procedures which are entirely analogous to those devised here but which allow for local departures from unit root formulations. Finally, the methods outlined in Section 7 for the refinement of the first order asymptotic theory may be extended to apply in quite general time series models with a unit root. All of these extensions are currently under investigation by the author.

MATHEMATICAL APPENDIX

Proof of Lemma 2.2. See Herrndorf (1984b), Corollary 1, p. 142.

Proof of Lemma 2.3. See Billingsley (1968), Corollary 1, p. 31.

Proof of Theorem 3.1

To prove (a) and (b) we write each statistic as a functional of $X_T(r)$ on $D[0,1]$. Thus, in the case of (a), we have

$$T^{-2} \Sigma_1^T y_{t-1}^2 = T^{-2} \Sigma_{i=1}^T \left(\Sigma_{j=1}^{i-1} u_j + y_0 \right)^2$$

$$= T^{-2} \Sigma_1^T (S_{i-1}^2 + 2 y_0 S_{i-1} + y_0^2)$$

$$= \sigma^2 \Sigma_1^T \int_{(i-1)/T}^{i/T} (1/T\sigma^2) S_{[Tr]}^2 \, dr + 2 y_0 \sigma T^{-1/2} \Sigma_1^T \int_{(i-1)/T}^{i/T} (1/\sqrt{T}\sigma) S_{[Tr]} \, dr + y_0^2/T$$

$$= \sigma^2 \int_0^1 X_T^2(r) \, dr + 2 y_0 \sigma T^{-1/2} \int_0^1 X_T(r) \, dr + y_0^2/T$$

$$\text{(A1)} \quad \Rightarrow \sigma^2 \int_0^1 W(r)^2 \, dr \,, \quad \text{as} \quad T \uparrow \infty \,;$$

by Lemmas 2.2-2.3. Note that (A1) holds whether y_0 is a constant (see (3a)) or is random (see (3b)).

In the above derivation

$$\text{(A2)} \qquad \sigma^2 = \lim_{T \to \infty} E(T^{-1} S_T^2)$$

as in part (a) of Assumption 2.1.

To prove (b) we have:

(A2) $\quad T^{-1}\Sigma_1^T y_{t-1}(y_t - y_{t-1}) = \Sigma_{i=1}^T (1/\sqrt{T})(S_{i-1} + y_0)(u_i/\sqrt{T})$

$$= T^{-1}\Sigma_1^T S_{i-1} u_i + y_0 \bar{u}$$

$$= (2T)^{-1}\Sigma_1^T (S_i^2 - S_{i-1}^2 - u_i^2) + y_0 \bar{u}$$

$$= (\sigma^2/2)\Sigma_1^T [X_T(r)^2]_{(i-1)/T}^{i/T} - (2T)^{-1}\Sigma_1^T u_i^2 + y_0 \bar{u}$$

(A3) $\qquad = (\sigma^2/2)X_T(1)^2 - (2T)^{-1}\Sigma_1^T u_i^2 + y_0 \bar{u} \; .$

Under the conditions of the theorem (in particular, the requirement that $\sup_t E|u_t|^{\beta+\eta} < \infty$ for some $\beta > 2$ and $\eta > 0$) we deduce that:

(A4) $\qquad T^{-1}\Sigma_1^T u_i^2 \xrightarrow[a.s.]{} \sigma_u^2 = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(u_i^2)$

and

(A5) $\qquad \bar{u} \xrightarrow[a.s.]{} 0$

by the strong law of McLeish (1975b, Theorem 2.10 with condition (2.12)). Now $X_T(1) \Rightarrow W(1)$ by Lemma 2.2. It follows from (A3)-(A5) and the continuous mapping theorem (Lemma 2.3) that as $T \uparrow \infty$:

(A6) $\qquad T^{-1}\Sigma_1^T y_{t-1} u_t \Rightarrow (\sigma^2/2)W(1)^2 - (\sigma_u^2/2) = (\sigma^2/2)\{W(1)^2 - \sigma_u^2/\sigma^2\}$

proving (b) of the theorem.

In view of (A1) and (A6), result (c) of the theorem is also a direct consequence of the continuous mapping theorem. Part (d) of the theorem follows immediately from result (c).

To prove part (e) we first write

$$s^2 = T^{-1}\Sigma(y_t - \hat{\alpha}y_{t-1})^2 = T^{-1}\Sigma_1^T u_t^2 - 2(\hat{\alpha}-1)T^{-1}\Sigma_1^T y_{t-1}u_t + (\hat{\alpha}-1)^2 T^{-1}\Sigma_1^T y_{t-1}^2 .$$

Then, in view of (A4) and parts (a), (b) and (d) of the theorem we deduce that $s^2 \xrightarrow{p} \sigma_u^2$ as $T \uparrow \infty$. Thus, it is of no consequence in the limiting distribution whether we use s^2 or $s'^2 = T^{-1}\Sigma(y_t - y_{t-1})^2$ (as in the LM approach) in the construction of the t_α statistic (8). Part (e) of the theorem now follows directly from the continuous mapping theorem using parts (a) and (c).

Proof of Lemma 4.1. The proof follows the same line as the proof of Lemma 6.17, pp. 149-152 of White (1984). We need only note that, since $\{u_t\}_1^\infty$ is strong mixing and $\sup_t E|u_t|^\beta < \infty$, the following inequality:

$$|E(u_t u_{t-\tau})| \leq c\alpha_\tau^{1-2/\beta}$$

holds for some constant c (see, for example, Ibragimov and Linnik (1971), Theorem 17.2.2, p. 307). Thus

$$|2T^{-1}\Sigma_{\tau=\ell+1}^{T-1}\Sigma_{t=\tau+1}^{T} E(u_t u_{t-\tau})| \leq 2cT^{-1}\Sigma_{\tau=\ell+1}^{T-1}(T-\tau)\alpha_\tau^{1-2/\beta}$$

and the right side of this inequality tends to zero as $T \uparrow \infty$ since $\ell \uparrow \infty$ and $\Sigma_1^\infty \alpha_m^{1-2/\beta} < \infty$. We deduce that $\sigma_T^2 - \sigma_{T\ell}^2 \to 0$ as $T \uparrow \infty$.

Proof of Theorem 4.2

The proof follows the same lines as the proof of Theorem 6.20, pp. 155-159 of White (1984), although we have no need to treat estimated residuals here. We first note that the assumptions of the theorem ensure that the

conditions of Lemma 4.1 hold. Thus, $\sigma_T^2 - \sigma_{T\ell}^2 \to 0$ as $T \uparrow \infty$. Next we have:

$$s_{T\ell}^2 - \sigma_{T\ell}^2 = T^{-1}\Sigma_1^T\{u_t^2 - E(u_t^2)\} + 2T^{-1}\Sigma_{s=1}^\ell\Sigma_{t=s+1}^T\{u_t u_{t-s} - E(u_t u_{t-s})\}$$

and

$$T^{-1}\Sigma_1^T\{u_t^2 - E(u_t^2)\} \xrightarrow[a.s.]{} 0 \ , \quad T \uparrow \infty$$

as in the proof of Theorem 3.1. Writing $Z_{ts} = u_t u_{t-s} - E(u_t u_{t-s})$, it remains to show that $T^{-1}\Sigma_{s=1}^\ell\Sigma_{t=s+1}^T Z_{ts} \xrightarrow[p]{} 0$ as $T \uparrow \infty$. But this part of the proof follows as in the proof given by White (1984, pp. 155-157). It is necessary, however, to correct the error that occurs on page 156 of White (1984) in the use of his Lemma 6.19. When one allows for the fact that s may increase with T the conclusion of Lemma 6.19 should be amended to:

$$E(\Sigma_{t=s+1}^T Z_{ts})^2 \le s(T-s)B < sTB$$

for a suitable constant B. It follows that (see p. 156 of White (1984) for details):

$$P\left[|T^{-1}\sum_{s=1}^\ell \sum_{t=s+1}^T Z_{ts}| \ge \epsilon\right]$$

$$\le \Sigma_{s=1}^\ell sTB\ell^2/\epsilon^2 T^2$$

$$= B\ell^3(\ell+1)/2\epsilon^2 T$$

which tends to zero if $\ell = o(T^{1/4})$. We deduce that $s_{T\ell}^2 \xrightarrow[p]{} \sigma^2$ as required.

Proof of Theorem 5.1

By Theorems 3.1 and 4.2 we have as $T \uparrow \infty$:

$$T(\hat{a}-1) \Rightarrow \frac{(1/2)(W(1)^2 - \sigma_u^2/\sigma^2)}{\int_0^1 W(r)^2 dr} \; ,$$

$$T^{-2} \Sigma_1^T y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr \; ,$$

$$s_{T\ell}^2 \xrightarrow[p]{} \sigma^2 \quad \text{and} \quad s_u^2 \xrightarrow[p]{} \sigma_u^2$$

Part (a) now follows directly by the continuous mapping theorem. In the same way we deduce that:

$$Z_t \Rightarrow \frac{(W(1)^2 - \sigma_u^2/\sigma^2)/2}{\{\int_0^1 W(r)^2 dr\}^{1/2}} - \frac{(\sigma^2 - \sigma_u^2)/2}{\sigma^2 \{\int_0^1 W(r)^2 dr\}^{1/2}}$$

and part (b) follows as required.

Proof of Lemma 6.1. Define $\xi_{nt} = \sigma^{-1} h_n^{-\frac{1}{2}} u_{nt}$. Then

$$Y_n(r) = h_n^{\frac{1}{2}} \Sigma_1^{i-1} \xi_{nt} \; ; \quad (i-1)T_n \leq r < i/T_n \; , \quad (i = 1, \ldots, T_n)$$

$$= N^{\frac{1}{2}} T_n^{-\frac{1}{2}} \Sigma_1^{i-1} \xi_{nt} \; .$$

But $\{\{\xi_{nt}\}_1^{T_n}\}_1^\infty$ is a triangular array of iid(0,1) variates so that

$$W_n(r) = T_n^{-\frac{1}{2}} \Sigma_1^{i-1} \xi_{nt} \; , \quad (i-1)/T_n \leq r < i/T_n$$

$$\Rightarrow W(r)$$

as $n \uparrow \infty$ (see, for example, McLeish (1977), Corollary 2.11). It follows

that as $n \uparrow \infty$ $Y_n(r) \Rightarrow N^{1/2}W(r)$ as required.

Proof of Theorem 6.2

Using (24), Lemma 6.1 and the continuous mapping theorem we obtain:

$$h_n \Sigma_1^{T_n} y_{nt-1}^2 = (N/T_n) \Sigma_1^{T_n} \left\{ \Sigma_1^{t-1} u_{ni} + y(0) \right\}^2$$

$$= (N/T_n) \Sigma_1^{T_n} \{ S_{nt-1}^2 + 2y(0) S_{nt-1} \} + Ny(0)^2$$

$$= N\sigma^2 \int_0^1 Y_n(r)^2 dr + 2N\sigma y(0) \int_0^1 Y_n(r) dr + Ny(0)^2$$

$$\Rightarrow N^2 \sigma^2 \int_0^1 W(r) dr + 2N^{3/2} \sigma y(0) \int_0^1 W(r) dr + Ny(0)^2$$

as $n \uparrow \infty$. This proves part (a).

To prove part (b) we write:

$$\Sigma_1^{T_n} y_{nt-1}(y_{nt} - y_{nt-1}) = \Sigma_1^{T_n} \{ \Sigma_1^{t-1} u_{ni} + y(0) \} u_{nt}$$

$$= h_n \sigma^2 \Sigma_1^{T_n} (\Sigma_1^{t-1} \xi_{ni}) \xi_{nt} + \sigma y(0) h_n^{1/2} \Sigma_1^{T_n} \xi_{nt}$$

$$= N\sigma^2 \Sigma_1^{T_n} (T_n^{-1/2} \Sigma_1^{t-1} \xi_{ni})(T_n^{-1/2} \xi_{nt}) + \sigma y(0) N^{1/2} T^{-1/2} \Sigma_1^{T_n} \xi_{nt}$$

$$= (N\sigma^2/2) \Sigma_1^{T_n} \{ [W_n(r)^2]_{(t-1)/T}^{t/T} - T_n^{-1} \xi_{nt}^2 \} + \sigma y(0) N^{1/2} W_n(1)$$

$$= (N\sigma^2/2) W_n(1)^2 - (N\sigma^2/2) T_n^{-1} \Sigma_1^{T_n} \xi_{nt}^2$$

$$\Rightarrow (N\sigma^2/2) W(1)^2 - (N\sigma^2/2) + \sigma y(0) N^{1/2} W(1)$$

by the strong law of large numbers, Lemma 6.2 and the continuous mapping theorem.

Part (c) follows directly from the expression

$$\hat{\alpha}_n - 1 = \Sigma_1^{T_n} y_{nt-1}(y_{nt} - y_{nt-1})/\Sigma_1^{T_n} y_{nt-1}^2$$

and parts (a) and (b) above.

To prove part (d) we note first, after a simple calculation, that

$$(1/h_n)s_n^2 \rightarrow \sigma^2 T_n^{-1}\Sigma_1^{T_n}\xi_{nt}^2 + 2(h_n/N)(1 - \hat{\alpha}_n)\Sigma_1^{T_n} y_{nt-1}u_{nt} + (h_n/N)(\hat{\alpha}_n - 1)^2\Sigma_1^{T_n} y_{nt-1}^2$$

$$\xrightarrow{p} \sigma^2$$

as $n \uparrow \infty$. Then

$$t_{\alpha_n} = (h_n\Sigma_1^{T_n} y_{nt-1}^2)^{1/2}(1/h_n)(\hat{\alpha}_n - 1)/(h_n^{-1/2}s_n)$$

$$\rightarrow [(1/2)(W(1)^2 - 1) + (y(0)/\sigma N^{1/2})W(1)]$$

$$\div [\int_0^1 W(r)^2 dr + 2(y(0)/\sigma N^{1/2})\int_0^1 W(r)dr + y(0)^2/\sigma^2 N]^{1/2}$$

as required.

Proof of Theorem 6.3

From the solution of (29) we have (under the null hypothesis $\theta = 0$)
$y(t) = \int_0^t \zeta(dr) + y(0)$. Thus,

$$y(t)/\sigma = V(t) + y(0)/\sigma$$

Since $\int_0^t \zeta(dr)$ is Gaussian by assumption $V(t)$ is here a Wiener process
on $C[0,N]$. Transform $t \rightarrow Nu = t$ with $u \in [0,1]$. Note that
$V(t) = V(Nu) \equiv N(0, Nu)$ so that we may write $V(Nu) \equiv N^{1/2}W(u)$ where
$W(u)$ is a Wiener process on $C[0,1]$. Now

$$\int_0^N y^2 dt = \sigma^2 \int_0^N V(t)^2 dt + 2\sigma y(0) \int_0^N V(t) dt + y(0)^2 N$$

$$\cong \sigma^2 N^2 \int_0^1 W(u) du + 2\sigma y(0) N^{3/2} \int_0^1 W(u) du + y(0)^2 N$$

and

$$\int_0^N y dy = \sigma^2 \int_0^N V dV + \sigma y(0) \int_0^N dV$$

$$= \sigma^2 N \int_0^1 W dW + \sigma y(0) N^{1/2} W(1)$$

$$= (\sigma^2 N/2) \{W(1)^2 - 1\} + \sigma y(0) N^{1/2} W(1)$$

and the required result follows.

REFERENCES

Anderson, R. L., (1942), "Distribution of the serial correlation coefficient," *Annals of Mathematical Statistics*, 13, 1-13.

Bartlett, M. S. (1946), "On the theoretical specification and sampling properties of autocorrelated time series," *Journal of the Royal Statistical Society*, 7, 27-41.

Bergstrom, A. R. (1976), *Statistical Inference in Continuous Time Economic Models*. North Holland: Amsterdam.

_____ (1984), "Continuous time stochastic models and issues of aggregation over time," Ch. 20, pp. 1145-1212 in M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics*, Vol. 2. North Holland: Amsterdam.

Bhargava, A. (1986), "On the theory of testing for unit roots in observed time series," mimeographed, University of Pennsylvania.

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

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

Brown, B. M. and J. I. Hewitt (1975), "Asymptotic likelihood theory for diffusion processes," *Journal of Applied Probability*, 12, 228-238.

Dickey, D. A. (1976), "Estimation and hypothesis testing for nonstationary time series," Ph.D. Thesis, Iowa State University, Ames, Iowa.

Dickey, D. A. and W. A. Fuller (1976), "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.

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

Donsker, M. D. (1951), "An invariance principle for certain probability limit theorems," *Memoirs of the American Mathematical Society*, 6, 1-12.

Erdös, 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.

Feigin, P. D. (1979), "Some comments about a curious singularity," *Journal of Applied Probability*, 16, 440-444.

Fuller, Wayne A. (1976), *Introduction to Statistical Time Series*. John Wiley & Sons: New York.

Granger, C. W. J. and P. Newbold (1974), "Spurious regressions in econometrics," *Journal of Econometrics*, 2, 111-120.

_____ (1977), *Forecasting Economic Time Series*. Academic Press: New York.

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.

Grenander, U. (1950), "Stochastic processes and statistical inference," *Arkiv För Matematik*, 1, 195-275.

Hall, P. and C. C. Heyde (1980), *Martingale Limit Theory and its Application*. Academic Press: New York.

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

Herrndorf, N. (1983), "The invariance principle for φ-mixing sequences," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 63, 97-108.

_____ (1984a), "A functional central limit theorem for ρ-mixing sequences," *Journal of Multivariate Analysis*, 15, 141-146.

_____ (1984b), "A functional central limit theorem for weakly dependent sequences of random variables," *Annals of Probability*, 12, 141-153.

Hida, T. (1980), *Brownian Motion*. Springer-Verlag: New York.

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

Lai, T. L. and C. Z. Wei (1982), "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *Annals of Statistics*, 10, 154-166.

_____ and D. Seigmund (1983), "Fixed accuracy estimation of an autoregressive parameter," *Annals of Statistics*, 11, 478-485.

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

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

_____ (1975b), "A maximal inequality and dependent strong laws," *Annals of Probability,* 3, 829-839.

_____ (1977), "On the invariance principle for nonstationary mixingales," *Annals of Probability,* 5, 616-621.

Nankervis, J. C. and N. E. Savin (1985), "Testing the autoregressive parameter with the t statistic," *Journal of Econometrics,* 27, 143-162.

Newey, W. K. and K. D. West (1985), "A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix," Princeton Discussion Paper No. 92.

Phillips, P. C. B. (1977), "Approximations to some finite sample distributions associated with a first-order stochastic difference equation," *Econometrica,* 45, 463-485.

Pollard, D. (1984), *Convergence of Stochastic Processes.* Springer-Verlag: New York.

Priestley, M. B. (1981), *Spectral Analysis and Time Series,* Vol. 1. Academic Press: New York.

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

Rozanov, Y. A. (1967), *Stationary Random Processes.* Holden Day: San Francisco.

Said, S. E. and D. A. Dickey (1985), "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika,* 71, 599-608.

Sargan, J. D. and A. Bhargava (1983), "Testing residuals from least squares regression for being generated by a Gaussian random walk," *Econometrica,* 51, 153-174.

Solo, V. (1984), "The order of differencing in ARIMA models," *Journal of the American Statistical Association,* 79, 916-921.

Tiao, G. C. and R. S. Tsay (1983), "Consistency properties of least squares estimates of autoregressive parameters in ARMA models," *Annals of Statistics,* 11, 856-871.

White, H. (1984), *Asymptotic Theory for Econometricians.* Academic Press: New York.

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

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

Withers, C. S. (1981), "Conditions for linear processes to be strong mixing," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete,* 57, 477-480.