COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 740

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References in publications to Discussion Papers (other than mere acknowledgement by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

TIME SERIES REGRESSION WITH A UNIT ROOT

P. C. B. PHILLIPS

April, 1985

TIME SERIES REGRESSION WITH A UNIT ROOT

by

P. C. B. Phillips*

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

This paper studies the random walk in a general time series setting that allows for weakly dependent and heterogeneously distributed innovations of the type recently considered in [39] and [40]. It is shown that simple least squares regression consistently estimates the unit root 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 [17] and [18].

*My thanks go to Don Andrews and David Pollard for some helpful discussions, and, as always, to Glena 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

Autoregressive time series with a unit root have been the subject of much recent attention in the econometric 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 [24] 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 [14] 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 [27].

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. A series of recent investigations by Dickey, Fuller and their associates in [11], [12], [13], [19], [20] and by Savin, Evans and Nankervis in [17], [18], [31], [32] have been at the forefront of this research. Related work on regression residuals has been done by Sargan and Bhargava [6, 37] and on ARIMA models by Solo [38].

All of the research cited above has been confined to the case where the sequence of innovations driving the model are independent with common variance. Frequently, it is assumed that the innovations are i.i.d. $(0, \sigma^2)$ or further that they are n.i.d. $(0, \sigma^2)$. Independence and homoskedasticity are rather strong assumptions to make in most empirical econometric work; and there are good reasons from economic theory (as shown in [24]) 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 of the random walk hypothesis 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 of the unit root and its 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 [39] and White and Domowitz [40] in the general nonlinear regression context. However, the limiting distribution theory that we employ here is quite different from that of [39] and [40]. 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 [39]). Our approach unifies and extends the presently known limiting distribution theory for the random walk and seems 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 i.i.d. errors. Thus, we discover that much of the work done by the authors cited in the earlier paragraph (particularly Fuller [19], Dickey and Fuller [12] and Evans and Savin [17]) under the assumption of i.i.d. errors remains relevant for a very much larger class of models.

Another aim of the paper is to present a new limiting distribution theory that is based on the concept of continuous data recording. This theory, together with the asymptotic expansion that is developed in Section 8 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 [17, 18] in this journal.

2. INVARIANCE PRINCIPLES FOR DEPENDENT HETEROGENEOUSLY DISTRIBUTED VARIABLES

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

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

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

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 (see [41]):

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

(3b) y_0 has a certain specified distribution;

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

(3c) is a circularity condition due to Hotelling that is used as a mathematical device to simplify distribution theory (see [1] or [2]), while (3b) is an initial condition that is frequently used to achieve stationarity We will employ (3a) which permits the greatest flexibility in the specification of (1), allowing in particular for non stationary series where $|\alpha| \geq 1$; but our main results also apply under (3b).

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

$$X_T(t) = \frac{1}{\sqrt{T}\sigma} S_{[Tt]} = \frac{1}{\sqrt{T}\sigma} S_{j-1} \; ; \quad (j-1)/T \leq t < j/T \quad (j = 1, \ldots, T) \tag{4}$$

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

where [ ] denotes the integer part of its argument and σ is a certain constant defined later (see Theorem 3.1). Observe that the sample paths $X_T(t) \in D = D[0,1]$, 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 $\|f-g\| = \sup_t |f(t) - g(t)|$ for any $f, g \in D$.

$X_T(t)$ is a random element in the function space $D[0,1]$. Under certain conditions, $X_T(t)$ can be shown to converge weakly to a limit process which is popularly known either as 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 [15] and Erdos and Kac [16]. The limit process which we denote by $W(t)$, has sample paths which lie in $C[0,1]$, the space of all real valued continuous functions on $[0,1]$. Moreover, $W(t)$ is a Gaussian process (for fixed t $W(t)$ is $N(0,t)$) and has independent increments ($W(s)$ is independent of $W(t) - W(s)$ for all $0 < s < t \leq 1$). We will denote the weak convergence of the process $X_T(t)$ to $W(t)$ by the notation $X_T(t) \Rightarrow W(t)$. Note that many finite dimensional CLT's follow directly from this result (e.g. the case in which $t = 1$ yields the Lindeberg-Lévy theorem when the u_j are i.i.d. $(0, \sigma^2)$). The reader is referred to Billingsley [7] for a detailed introduction to the subject and to Pollard [34] for an excellent recent treatment.

The conditions under which $X_T(t) \Rightarrow W(t)$ are very general indeed and extend to a wide class of nonstationary, weakly dependent and heterogeneously distributed innovation sequences $\{u_t\}$. Billingsley [7, Ch. 4(21)] 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 [23, 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 [29, 30]. His results are, in fact, the ones that we will use in our own development because they apply most easily to the weakly dependent and heterogeneously distributed data that we wish to allow for in the context of a 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 will assume that $\{u_t\}_1^\infty$ is a sequence of random variables defined on a probability space $(\Omega, \mathcal{B}, P)$ and satisfying the zero mean condition:

(5) $$E(u_t) = 0 , \quad \text{for all } t .$$

We want to allow for both temporal dependence and heterogeneity in the process $\{u_t\}_1^\infty$. As discussed by White [39] and White and Domowitz [40], a convenient and general way of doing this is through the use of mixing conditions on the process. These conditions allow for serial correlation and heteroskedasticity but control the extent of the temporal dependence so that although there may be substantial dependence amongst recent events, events which are separated by long intervals of time are almost independent. Formally, we introduce the following two measures of dependence between the σ-algebras $\mathcal{F}$ and $\mathcal{G}$:

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

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

Now let $\mathcal{F}_a^b$ denote the σ-field generated by $\{u_a, u_{a+1}, \ldots, u_b\}$ and $\mathcal{R}_a^b$ be the σ-field generated by $\{u_a + u_{a+1} + \ldots + u_b = S_b - S_{a-1}\}$ for all $a \leq b$. We define:

$$\varphi_m = \sup_n \sup_{j\geq n+m} \varphi(\mathcal{F}_1^n, \mathcal{R}_{n+m}^j) \tag{6}$$

$$\alpha_m = \sup_n \sup_{j\geq n+m} \alpha(\mathcal{F}_1^n, \mathcal{R}_{n+m}^j) . \tag{7}$$

The measures (6) and (7) were introduced by McLeish [29]. Both quantities measure the extent of dependence between events which are separated by at least m time periods. We will assume that

$$\varphi_m \downarrow 0 , \quad \alpha_m \downarrow 0 \quad \text{as} \quad m \uparrow \infty . \tag{8}$$

These conditions are somewhat weaker than the related conditions (e.g. see

[39]) which define φ-mixing and α-mixing processes, respectively; and they (i.e. (8)) are, in fact, implied by the latter [29]. Following [29] we will also be more specific about the rates at which φ_m and α_m approach zero as $m \uparrow \infty$. Adopting a somewhat stronger definition than that of McLeish [29, p. 167] we will say that φ_m (or α_m) is of size-p if φ_m (or α_m) $= O(m^{-\rho-\varepsilon})$ for some $\varepsilon > 0$ as $m \uparrow \infty$.

The following invariance principle is due to McLeish [29, pp. 168-169]:

LEMMA 2.1. If $\{u_t\}_1^\infty$ *is a random sequence satisfying* (5) *and*

(a) $E(T^{-1}S_T^2) \to \sigma^2 > 0$ *as* $T \uparrow \infty$

(b) $\{u_t^2\}$ *is uniformly integrable (e.g. see* [10, *p. 96*]*)*

(c) $\sup_t (E|u_t|^\beta) < \infty$ *for some* $2 \leq \beta \leq \infty$

(d) $T^{-1}E(S_{k+T} - S_k)^2 \to \sigma^2$ *as* $\min(k,T) \uparrow \infty$,

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

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

In the above (b), (c) and (d) are moment conditions which control the degree of heterogeneity in the process and prevent the variances (and possibly higher moments) from fluctuating too wildly; (a) (and also (d)) control the normalization of the functional at a rate which ensures a non degenerate distribution on $C[0,1]$; and (e) controls the mixing decay rate in relation to the probability of outliers as determined by the moment conditions. Thus, as there is a higher probability of outliers $(\beta \to 2)$ the mixing decay rate of α_m increases so that the effect of the outliers dies off more quickly.

For stationary processes we have a similar result which holds under somewhat simpler conditions:

LEMMA 2.2. If $\{u_t\}_1^\infty$ *is a weakly stationary random sequence satisfying* (5) *and*

(a) $E|u_1|^\beta < \infty$ *for some* $2 \leq \beta < \infty$;

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

then

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

and if $\sigma^2 > 0$

$$X_T(t) \Rightarrow W(t) \quad \text{as} \quad T \uparrow \infty \ .$$

These results cover a very wide class of processes that are commonly used in econometric modeling. Lemma 2, for instance, includes stationary linear processes of the form $u_t = \Sigma_{j=-\infty}^\infty a_j \varepsilon_{t-j}$ where $\Sigma_{-\infty}^\infty a_j^2 < \infty$ and the sequence $\{\varepsilon_i\}$ is i.i.d. $(0, \sigma^2)$ under mild additional conditions on the sequence of constants a_j (which ensure that the mixing numbers φ_n or α_n satisfy (c)). In particular the Lemma applies to many stationary finite order ARMA processes, which have exponential mixing decay rates under very general conditions. Lemma 1 extends this class further by allowing for processes with a moderate degree of heterogeneity as well as temporal dependence.

The following result which is given a very thorough treatment in [7] will be used extensively in the following sections.

LEMMA 2.3. If h *is any continuous functional on* C[0,1] *(continuous that is, except for at most a set of points* D_h *for which* $P(W \in D_h) = 0$ *) then* $X_T(t) \Rightarrow W$ *implies that* $h(X_T(t)) \Rightarrow h(W)$.

3. CONSISTENT ESTIMATION OF A UNIT ROOT

Let

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

denote the ordinary least squares (OLS) estimator of α in (1). We write

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

and consider the limiting behavior of the standardized statistic (8).

THEOREM 3.1. If $\{u_t\}_1^\infty$ *satisfies the conditions of either* Lemma 2.1 *or* 2.2 *and if* $\sup_t E|u_t|^{\beta+\eta} < \infty$ *for some* $2 \leq \beta < \infty$ *and any* $\eta > 0$, *then*

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

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

where

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

and

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

(c) $$T(\hat{\alpha}-1) \Rightarrow \frac{(1/2)(W(1)^2 - \sigma_u^2/\sigma^2)}{\int_0^1 W(t)^2 dt} .$$

Moreover, (a), (b) *and* (c) *continue to hold whether the initial conditions are given by* (3a) *or* (3b).

Note that when the innovation sequence $\{u_t\}$ is i.i.d. $(0, \sigma^2)$ we have $\sigma^2 = \sigma_u^2$ and

(9) $$T(\hat{\alpha}-1) \Rightarrow \frac{(1/2)(W(1)^2 - 1)}{\int_0^1 W(t)^2 dt} .$$

This result was first given by White [41] although his result (p. 1196 of [41]) is incorrect as stated since his standardization of $\hat{\alpha}$ is $g(T)(\hat{\alpha}-1)$ with $g(T) = T/\sqrt{2}$ (see (3.2) of [41]). Unfortunately, the same error occurs repeatedly in the recent paper by Rao [35] (see, in particular, pp. 187-188 of [35]).

Theorem 3.1 extends (9) 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 (9) may be illustrated with a simple example. Suppose that the generating process for $\{u_t\}$ is the moving average

(10) $$u_t = \varepsilon_t + \theta\varepsilon_{t-1} , \quad t = 1, 2, \ldots$$

with ε_t i.i.d. $(0, \sigma_\varepsilon^2)$. Then

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

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

and we have

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

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

$$(11) \qquad T(\hat{\alpha}-1) \Rightarrow \frac{(1/2)[W(1)^2 - (1+\theta^2)/(1+\theta)^2]}{\int_0^1 W(t)^2 dt}$$

generalizing (9) and, of course, reducing to it when $\theta = 0$.

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 $X_T(t)$ as defined in (4) by its close relative, the random element

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

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

which lies in $C[0,1]$. In fact, $Y_T(t) \Rightarrow W(t)$ under the same conditions as those prescribed earlier for $X_T(t)$ in Lemmas 2.1-2.2. However, the sample path of $Y_T(t)$ is continuous and of bounded variation on [0,1] so that we may define and evaluate by partial integration the following Riemann Stieltjes integral:

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

The corresponding integral for the limit process $W(t)$ must be defined as a stochastic integral,[1] for which the rule of partial integration used in (12) does not apply. Instead, we have the well known result (e.g., see [26, p. 158]):

---

[1]E.g., see [36]. For a good recent introduction to the subject in the econometrics literature, see [5].

$$\text{(13)} \qquad \int_0^1 WdW = (1/2)(W(1)^2 - 1)$$

whereas from (12) and Lemma 2.3

$$\text{(14)} \qquad \int_0^1 Y_T dY_T \Rightarrow (1/2)W(1)^2 \ .$$

The problem arises because all elements of C[0,1] except for a set of Wiener measure zero are of unbounded variation [7, p. 63]. In particular, the sample paths of $W(t)$ are almost surely of unbounded variation and thus the integral $\int_0^1 WdW$ 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 $\int Y_T dY_T \Rightarrow \int WdW$ 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. Since $dY_T(t) = \sqrt{T}u_j dt/\sigma$ we find by 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:

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

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

THEOREM 3.2. If $\{u_t\}_1^\infty$ *satisfies the conditions of* Theorem 3.1 *then* $\hat{\alpha} \xrightarrow[p]{} 1$ as $T \to \infty$.

Theorem 3.2 shows that, unlike the stable AR(1) with $|\alpha| < 1$ (in place of our (2) above) OLS retains the property of consistency even in the presence of substantial serial correlation. This extremely simple yet general result seems never to have appeared before in the literature. The robustness of the conclusion of the theorem is most extraordinary, allowing for a wide variety of error processes which certainly permit serious misspecifications in the usual random walk formulation of (1) with white noise errors.

4. THE REGRESSION t-STATISTIC

Define

$$t = (\hat{\alpha}-\alpha)/s(\Sigma_1^T y_{t-1}^2)^{-1/2} \tag{16}$$

where

$$s^2 = T^{-1}\Sigma_1^T(y_t - \hat{\alpha}y_{t-1})^2 \ . \tag{17}$$

t is the conventional regression t-statistic that is often used for testing hypotheses about α . The distribution of this statistic under the null hypothesis $\alpha = 1$ and under certain alternatives $\alpha \neq 1$ has recently been studied extensively in a series of articles by Dickey and Fuller [12, 13] and Savin and Nankervis [31]. This work concentrates altogether on the special case in which the innovation sequence $\{u_t\}$ is i.i.d. $(0, \sigma^2)$. In related work Solo [38] has recently studied the asymptotic distribution of the closely related Lagrange multiplier (LM) statistic in a general ARIMA setting; once again, his results are established under the assumption that i.i.d. innovations drive the model.

Our approach here is close to that of Solo's but we allow for a much more general class of error processes. The following result characterizes the limiting distribution of the t-statistic (16):

THEOREM 4.1. If $\{u_t\}_1^\infty$ *satisfies the conditions of* Theorem 3.1 *then*

$$t \Rightarrow \frac{(1/2)\left\{W(1)^2 - \sigma_u^2/\sigma^2\right\}}{(\sigma_u/\sigma)\left\{\int_0^1 W(t)^2 dt\right\}^{1/2}} \tag{18}$$

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

As shown in the Appendix the limiting distribution (18) is the same whether we employ the variance estimate s^2 as in (17) or $s'^2 = T^{-1}\Sigma_1^T(y_t - y_{t-1})^2$ as in the LM approach. Writing $LM = t'^2$ for the latter case, we deduce from Theorem 4.1 and the continuous mapping theorem that

$$LM \Rightarrow \frac{(1/4)\left[W(1)^2 - \sigma_u^2/\sigma^2\right]^2}{(\sigma_u/\sigma)^2\int_0^1 W(t)^2 dt} .$$

This specializes to Solo's result in [38] when $\sigma_u^2 = \sigma^2$.

5. ESTIMATION OF (σ_u^2, σ^2)

The limiting distributions given in Theorems 3.1 and 4.1 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) . The new statistics (given by (21) and (22) below) provide a very general test of the random walk hypothesis (1).

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

$$(19) \qquad s_u^2 = T^{-1}\Sigma_1^T(y_t - y_{t-1})^2 = T^{-1}\Sigma_1^T u_t^2$$

which is consistent for σ_u^2 under the null hypothesis (2).

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 solved by White and Domowitz [40]. An excellent treatment is available in Chapter VI of [39].

We follow the approach of [39] but allow for the more general dependence conditions due to McLeish [29] that are discussed in Section 2. We also correct an error of Theorem 3.5 of [40] and Theorem 6.20 of [39].[2]

---

[1] This is most easily seen by noting that $S_T/\sqrt{T} \Rightarrow N(0, \sigma^2)$ by the invariance principle Lemma 2.1.

[2] This error was detected independently by Whitney Newey and myself.

Start by defining

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

and the approximant

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

We call ℓ the lag truncation number. The following result is given by White [39, Lemma 6.17, p. 149]:

LEMMA 5.1. If the sequence $\{u_t\}_1^\infty$ *satisfies:*

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

(b) $\sup_t E|u_t|^{2+2\eta} < \infty$, *for some* $\eta > 0$;

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

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

$$\sigma_T^2 - \sigma_{T\ell}^2 \to 0$$

as $T \uparrow \infty$.

This result 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 in [39] and [40]. We define

$$(20) \qquad s_{T\ell}^2 = T^{-1}\Sigma_1^T u_t^2 + 2T^{-1}\Sigma_{\tau=1}^{\ell}\Sigma_{t=\tau+1}^{T} u_t u_{t-\tau} .$$

The following result establishes that this is a consistent estimator of σ^2 .

THEOREM 5.2. If the sequence $\{u_t\}_1^\infty$ *satisfies the conditions of either* Lemma 2.1 *or* 2.2 *and if*

(a) $\sup_t E|u_t|^{4(r+\delta)} < \infty$ *for some* $\delta > 0$ *and* $r > 1$;

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

(c) $\ell \uparrow \infty$ *as* $T \uparrow \infty$ *such that* $\ell = o(T^{1/4})$

then $s_{T\ell}^2 \xrightarrow[p]{} \sigma^2$ *as* $T \uparrow \infty$.

According to this result, if we allow the number of estimated autocorrelations 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 provide some guidelines for the selection of ℓ in [40]. Inevitably 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 autocorrelations 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.

6. NEW TESTS OF THE RANDOM WALK HYPOTHESIS

The consistent estimates s_u^2 and $s_{T,\ell}^2$ may be used to develop new tests of the random walk hypothesis that apply under very general conditions. Define the statistics

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

and

$$(22) \qquad Z_t = \frac{\hat{\alpha}-1}{s_{T\ell}(\Sigma_1^T y_{t-1}^2)^{-1/2}} - \frac{(s_{T\ell}^2 - s_u^2)/2}{s_{T\ell}(T^{-2}\Sigma_1^T y_{t-1}^2)^{1/2}}$$

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

The limiting distributions of Z_α and Z_t are given by:

THEOREM 6.1. If the conditions of Theorem 5.2 *are satisfied then*

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

and

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

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

Theorem 6.1 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 $\{u_t\}$. Moreover, the limiting distribution of Z_α is identical to that of $T(\hat{\alpha}-1)$ when $\sigma_u^2 = \sigma^2$ (see (9) above). The latter distribution has recently been computed by Evans and Savin [17, 18] using numerical methods. Tabulations and graphical plots of both the limiting pdf are presented in [17]; [17] 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 the tables in [17].

The limiting distribution of Z_t given in Theorem 6.1 is identical to that of the regression t-statistic when $\sigma^2 = \sigma_u^2$ (see Theorem 4.1). This is, in fact, the limiting distribution of the t-statistic when the innovation sequence $\{u_t\}$ is i.i.d. $(0, \sigma^2)$. The latter distribution has been calculated using Monte Carlo methods by Dickey [11] and tabulations of percentage points of the distribution are reported in Fuller [19, table 8.5.2, p. 373].

Theorem 6.1 shows that much of the work of these authors on the distribution of the OLS estimator $\hat{\alpha}$ and the regression t-statistic under i.i.d. innovations remains relevant for a very much larger class of models. In fact, their results appear to be relevant in almost any time series with a unit root provided one makes the simple modifications contained in (21) and (22) to the usual statistics.

7. REGRESSIONS WITH A CONTINUUM OF OBSERVATIONS

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. For these reasons it is of some intrinsic interest to study the limiting behavior of econometric estimators and test statistics as the time interval between sampled observations (which we call the sampling interval and denote by h) goes to zero. Problems of this type have already received a good deal of attention in the econometric literature (see [4] for a collection of articles in the field and [5] for an excellent recent review). However, none of the existing research in this field has considered unstable systems or regressions with unit roots. It turns out, as we show below, that there is a very interesting relationship 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 the following section, help to explain some of the recent Monte Carlo results reported by Evans and Savin [17, 18], in particular, their most interesting result that when $\alpha = 1$ the finite sample distribution of $\hat{\alpha}$ is very well approximated by its asymptotic distribution even for quite small samples $(T \geq 20)$.

We begin by considering the stochastic differential equation

$$\dot{y}(t) = \theta y(t) + \zeta(t) \tag{24}$$

where $\dot{y}$ denotes dy/dt, $\theta \leq 0$ and $y(t)$ and $\zeta(t)$ are random functions of continuous time over $0 \leq t \leq N$. When $\zeta(t)$ in (24) is a pure noise process in continuous time (which we may rigorously treat as a generalized stochastic process[1]) (24) is more usually written in the form:

$$dy(t) = \theta y(t)dt + \zeta(dt) \tag{25}$$

where $\zeta(dt)$ is a σ-additive random measure defined on all subsets of $-\infty < t < \infty$ with finite Lebesgue measure with the properties

$$E(\zeta(dr)) = 0 \ , \quad E(\zeta(dr)^2) = \sigma^2 dr \cdot \tag{26}$$

for some $\sigma^2 > 0$ (e.g., see [36], pp. 1157-1163). In this case, the stochastic integral $\int_0^t \zeta(dr)$ has zero mean, variance $\sigma^2 t$ and uncorrelated

[1]E.g., see [42].

increments; if its distribution is Gaussian then $\sigma^{-1}\int_0^t \zeta(dr)$ is a Wiener process. Given the initial condition $y(0)$, (25) has solution [5, p. 1163]:

$$(27) \qquad y(t) = \int_0^t e^{\theta(t-r)}\zeta(dr) + e^{\theta t}y(0) .$$

If we consider (in a purely formal way) the problem of estimating the parameter θ in (24) from the continuous record $\{y(t);\ 0 \leq t \leq N\}$ least squares suggests the criterion

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

leading to the estimator

$$(28) \qquad \hat{\theta} = \frac{\int_0^N y\dot{y}dt}{\int_0^N y^2 dt} = \frac{\int_0^N ydy}{\int_0^N y^2 dt} .$$

(28) was originally suggested by Bartlett in [3]. Its properties were subsequently studied in [22] and more recently in [9]. When $y(t)$ is generated by (25) then $\int_0^N ydy$ in the numerator of (28) is understood to be a stochastic integral.

The discrete time model corresponding to (24) or (25) is (e.g., see [5], pp. 1154, 1163):

$$(29) \qquad y(th) = e^{\theta h}y(th-h) + u(th) \ ; \quad t = 1, 2, \dots ,$$

where

$$(30)\qquad u(th) = \int_{th-h}^{th} e^{\theta(th-r)}\zeta(r)dr$$

$$= \int_{th-h}^{th} e^{\theta(th-r)}\zeta(dr)$$

with the latter stochastic integral in (30) applying when the model is (25). Note that (30) takes the form

$$(31)\qquad y_t = \alpha y_{t-1} + u_t$$

of (1) with $\alpha = e^{\theta h}$, $y_t = y(th)$, $u_t = u(th)$; and this reduces to the random walk with $\alpha = 1$ when $\theta = 0$.

We shall concentrate on the case where there is a fixed span $[0,N]$ over which observations are taken at discrete intervals determined by h . We set $N = Th$ and we shall consider below the asymptotic distribution of $\hat{\alpha}$ as $h \downarrow 0$. Since N is fixed we have $T = N/h \uparrow \infty$ as $h \downarrow 0$. It will be convenient (although not essential) in what follows to assume that T is integer, for instance, corresponding to the sequence $h = 1, 1/2, 1/3, \ldots$. We shall also fix attention on the model (25) as the generating mechanism for $y(t)$.[1]

Let $S_i = \sum_{j=1}^{i} u(jh)$ with $S_0 = 0$ as usual. Define the random element of $D[0,1]$:

$$(32)\qquad Y_h(t) = \sigma^{-1}S_{i-1}\ ; \quad (ih-h)/N \leq t < ih/N \quad (i = 1, \ldots, N/h)$$

$$Y_h(1) = \sigma^{-1}S_{N/h}$$

---

[1]A more general treatment will be pursued in later work. It is not needed for our present purpose.

and the process

$$(33) \qquad \xi_i = h^{-1/2}u(ih) = h^{-1/2}\int_{ih-h}^{ih} \zeta(dr) \ .$$

Then

LEMMA 7.1. If $\{\xi_i\}$ *satisfies the conditions of* Lemma 2.2 *then* $Y_h(t) \Rightarrow N^{1/2}W(t)$ *as* $h \downarrow 0$.

Note that from (26) $E(\xi_i) = 0$ and $E(\xi_i^2) = \sigma^2$. Moreover, if the process $\{\xi_i\}$ is Gaussian (as is frequently assumed), then the mixing conditions of Lemma 2.2 hold automatically since the process is i.i.d.

THEOREM 7.2. If $y(t)$ *is generated by* (25) *with* $\theta = 0$ *and if* $\{\xi_i\}$ *satisfies the conditions of* Lemma 2.2 *then as* $h \downarrow 0$

(a) $h\Sigma_1^T y(th-h)^2 \Rightarrow N^2\sigma^2\{\int_0^1 W(t)^2dt + 2(y_0/\sigma\sqrt{N})\int_0^1 W(t)dt + y_0^2/\sigma^2 N\}$,

(b) $\Sigma_1^T y(th-h)\{y(th) - y(th-h)\} \Rightarrow (N\sigma^2/2)\{W(1)^2 - 1 + 2(y_0/\sigma\sqrt{N})W(1)\}$,

(c) $h^{-1}(\hat{\alpha}-1) \Rightarrow \left(\frac{1}{N}\right)\dfrac{(1/2)\{W(1)^2 - 1)\} + (y_0/\sigma\sqrt{N})W(1)}{\int_0^1 W(t)^2dt + 2(y_0/\sigma\sqrt{N})\int_0^1 W(t)dt + y_0^2/\sigma^2 N}$

where $W(t)$ *is the Wiener process on* $C[0,1]$.

Theorem 7.2 shows that for small h the distribution of $\hat{\alpha}$ may be approximated by a suitable functional of Brownian motion. This functional involves the initial condition y_0 , which may be either constant (see (3a)) or random (3b). For large N this distribution may be approximated by

$$\left(\frac{1}{N}\right)\frac{(1/2)(W(1)^2 - 1)}{\int_0^1 W(t)^2dt}$$

corresponding to a special case of our earlier result in Theorem 3.1 (in which $\sigma_u^2 = \sigma^2$) and to the conventional large T asymptotics. However, in other cases, the initial value y_0 and more particularly the ratio y_0/σ have an important effect on the limiting distribution of $h^{-1}(\hat{\alpha}-1)$. When y_0/σ is large the effect is likely to be substantial. Indeed, as $c = y_0/\sigma$ becomes very large the limiting distribution of $h^{-1}(\hat{\alpha}-1)$ given in (c) is approximately $N^{-3/2}c^{-1}W(1)$ or $N(0, 1/c^2N^3)$.

THEOREM 7.3. If $y(t)$ *is generated by* (25) *with* $\theta = 0$ *and if* $\{\xi_i\}$ *is Gaussian then* $\hat{\theta}$ *has the same distribution in finite samples (i.e., finite* N *) as the functional*

$$(34) \qquad \frac{1}{N}\cdot\frac{(1/2)\{W(1)^2 - 1\} + (y(0)/\sigma\sqrt{N})W(1)}{\int_0^1 W(t)^2 dt + 2(y(0)/\sigma\sqrt{N})\int_0^1 W(t)dt + y(0)^2/\sigma^2 N}$$

of the Wiener process $W(t)$.

Theorems 7.2 and 7.3 help to explain several of the phenomena discovered in the experimental investigation of Evans and Savin [17]. These authors found: (i) that the finite sample distribution of $\hat{\alpha}$ was very well approximated by its asymptotic distribution (as $T \to \infty$ in the conventional large sample asymptotics with h fixed) even for quite small samples when the initial value $y_0 = 0$; (ii) that changes in $c = y_0/\sigma$ precipitate substantial changes in the distribution of $\hat{\alpha}$; specifically the distribution noticeably concentrates as c increases.

The fact that the asymptotic distribution of $\hat{\alpha}$ is a very good approximation in finite samples when $y_0 = 0$ is well explained by Theorems 7.2 and 7.3. First, we observe that this asymptotic distribution applies not only as $T \uparrow \infty$ in the conventional sense with h fixed (see our Theorem 3.1) but also as $h \downarrow 0$ with a fixed data span N (our Theorem 7.2(c) with $y_0 = 0$). Thus, the limiting distribution theory operates in two different directions with identical results when $y_0 = 0$. Moreover, as shown in Theorem 7.3 this limiting distribution is actually the finite sample distribution of the continuous record estimator $\hat{\theta}$ when the process is Gaussian (as in the experimental investigation [17]).

Theorem 7.2 also demonstrates that $c = y_0/\sigma = y(0)/\sigma$ is an important parameter in the limiting distribution of $h^{-1}(\hat{\alpha}-1)$ as $h \downarrow 0$. 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/σ . The second phenomena noted by Evans and Savin, that the distribution of $T(\hat{\alpha}-1)$ concentrates as $c = y_0/\sigma$ increases, is a natural consequence of the fact that (34) $\xrightarrow[p]{} 0$ as $c \uparrow \infty$.

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

A general refinement of the functional central limit theorem of Lemma 2.1 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 $\{u_i\}$ with stronger properties.

We will consider here the special case in which the u_i are i.i. $N(0, \sigma^2)$. In this case it is easy to see that, for fixed $t \in [0,1]$, $X_T(t)$ is $N(0, [Tt]/T)$. In fact, we may write for $0 < t \leq 1$

$$X_T(t) = W(t)([Tt]/Tt)^{1/2}$$

$$= W(t)\{1 - (Tt - [Tt])/Tt\}^{1/2}$$

$$(35) \qquad = W(t)\left\{1 - \frac{1}{2}\frac{Tt - [Tt]}{Tt}\right\} + O_p(T^{-2})$$

while $X_T(0) = W(0) = 0$. (35) provides a simple asymptotic expansion for the finite dimensional (in fact, one dimensional) distribution of $X_T(t)$ with t fixed. Note that since $W(t)/t \xrightarrow[a.s.]{} 0$ as $t \downarrow 0$ (e.g., see [26, p. 57] the expansion (35) remains well defined in the neighborhood of $t = 0$. Higher order finite dimensional distributions of $X_T(t)$ may be treated in a similar way. The error on the approximation $X_T(t) \sim W(t)$ is seen to be of $O(T^{-1})$ in (35). This suggests that certain functionals of $X_T(t)$ may be expected to differ from the same functionals of $W(t)$ by quantities of the same order. In particular

$$(36) \qquad \int_0^1 X_T(t)^2 dt = \int_0^1 W(t)^2 dt + O_p(T^{-1})$$

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

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

$$(37) \qquad T^{-1}\Sigma_1^T y_{t-1}(y_t - y_{t-1}) = (\sigma^2/2)X_T(1)^2 - (2T)^{-1}\Sigma_1^T u_i^2 + y_0\bar{u}$$

from A3 in the Appendix. We shall confine our attention to the case where the initial value $y_0 = 0$. Since $X_T(1)^2 = W(1)^2 + O_p(T^{-1})$ (37) becomes

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

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

where ξ is $N(0,1)$ and is independent of $W(t)$. The distribution of ξ follows directly from the Lindeberg-Lévy theorem. Note that ξ is dependent on a quadratic function of the u_i whereas $W(t)$ depends on partial sums which are linear in the u_i . Hence, ξ and $W(t)$ are uncorrelated and, being normal, are therefore independent.

We deduce the following result.

THEOREM 8.1. 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* n.i.d. $(0, \sigma^2)$ *then*

$$T(\hat{\alpha}-1) = \frac{(1/2)(W(1)^2 - 1) - (1/\sqrt{2T})\xi}{\int_0^1 W(t)^2 dt} + O(T^{-1}) \tag{38}$$

where $W(t)$ *is the Wiener process on* $C[0,1]$ *and* ξ *is* $N(0,1)$ *and independent of* $W(t)$.

(38) 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 [33]. 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 (38) 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 of [17]. It will be of interest to discover the extent to which (38) 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.

9. 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

$$(39) \qquad a(L)(1-L)y_t = b(L)e_t$$

for given finite order lag polynomials $a(L)$ and $b(L)$ in the lag operator L. Then, upon inversion, (39) becomes

$$(40) \qquad y_t = y_{t-1} + u_t \;, \quad u_t = a(L)^{-1}b(L)e_t$$

and u_t will satisfy the weak dependence and heteroskedasticity assumptions of Lemma 2.2 under very general conditions on the innovations and lag polynomials of (39). In the latter case, we may write

$$(41) \qquad a(L)(1-L)y_t = b(L)x_t + c(L)e_t$$

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 (39) and (41) 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 statistic (21) and use the appropriate limiting distributions (given by Theorems 3.1 and 6.1); this approach applies under conditions that are of even wider applicability than (39) and (41). In a certain sense, this idea is already implicit in the Box-Jenkins modeling approach. However, none of the traditional theory in this research (e.g., see [8] or [21]) allows for estimation or testing procedures that have anything approaching the range of applicability of our own approach.

More analysis of the statistical properties, including power, of the procedures devised here is needed. The present paper is only an exploratory study and more research on these and other aspects of time series with unit roots is planned for future work.

MATHEMATICAL APPENDIX

Proof of Lemma 2.1. This Lemma is Theorem 3.8 of McLeish [29] and is proved on pp. 176-177 of [29].

Proof of Lemma 2.2. This Lemma is Corollary 3.9 of McLeish [29] and is proved on p. 169 of [29].

Proof of Lemma 2.3. See Theorem 5.1, pp. 30-31 of [7].

Proof of Theorem 3.1

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

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

$$= T^{-2}\Sigma_1^T(S_{i-1}^2 + 2y_0S_{i-1} + y_0^2)$$

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

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

$$\text{(A1)} \rightarrow \sigma^2\int_0^1 W(t)^2dt \ .$$

by Lemmas 2.1-2.3. Note that (A1) holds whether y_0 is a constant (see (3a)) or is random (see (3b)).

In the above derivation

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

(see condition (a) of Lemma 2.1). When the process $\{u_t\}_1^\infty$ is weakly stationary as in Lemma 2.2 the convergence of the sequence $\{E(T^{-1}S_T^2)\}$ to σ^2 is assured by the mixing decay condition (b) of Lemma 2.2 (see [29], p. 169).

In the case of (a) we have:

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

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

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

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

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

Under the conditions of the theorem (in particular, the requirement that $\sup_t E|u_t|^\gamma < \infty$ for some $\gamma > 2$) we deduce that:

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

and

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

by the strong law of McLeish [28, Theorem 2.10 with condition (2.12)].

Now $X_T(1) \Rightarrow W(1)$ by either Lemma 2.1 or 2.2. It therefore follows from (A3)-(A5) and the continuous mapping theorem (Lemma 2.3) that:

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

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.

Proof of Theorem 3.2

Write

$$\hat{\alpha} = 1 + T^{-1}\xi_T$$

where

$$\xi_T = \frac{T^{-1}\Sigma_1^T y_{t-1}u_t}{T^{-2}\Sigma_1^T y_{t-1}^2} .$$

Since ξ_T has a limiting distribution (given by (c) of Theorem 3.1) it follows directly that $\hat{\alpha} \xrightarrow[p]{} 1$ as $T \uparrow \infty$.

Proof of Theorem 4.1

Write

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

Clearly, under the conditions of the theorem,

$$s^2 \xrightarrow[p]{} \sigma_u^2 \quad \text{as} \quad T \uparrow \infty .$$

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 statistic (13). The result given in (15) now follows directly from the earlier results of Theorem 3.1.

Proof of Lemma 5.1. The proof follows the same line as the proof of Lemma 6.17, pp. 149-152 of [39]. We need only note that this proof applies also to the weaker dependence measures φ_m and α_m since McLeish's Lemma 3.5 in [29] may be used in place of Lemma 6.15 of [39] to establish the decay rates of the autocorrelation sequence $\{E(u_t u_{t-\tau})\}$. Note also that under our modified definition of size (see Section 2) we also have $\Sigma_{n=0}^{\infty}\varphi_n^{1/2} < \infty$ and $\Sigma_{n=0}^{\infty}\alpha_n^{\eta/(2+2\eta)} < \infty$, which are required in the proof of [39].

Proof of Theorem 5.2

The proof follows the same lines as the proof of Theorem 6.20, pp. 155-159 of [39], although we have no need to treat estimated residuals here.

Note first that the moment condition (a) of the Theorem implies the moment conditions of Lemmas 2.1 and 2.2. In addition, the mixing decay conditions (b) are stronger than those of Lemmas 2.1 and 2.2 and therefore imply the latter. Thus, if φ_m is of size -2 then it is certainly of size $-\beta/(2\beta-2)$ for any $\beta \geq 2$, thereby satisfying the condition of Lemma 2.1. Similarly, if α_m is of size $-2(r+\delta)/(r+\delta-1)$ for $r > 1$ and some $\delta > 0$ then α_m is of size $-(2\eta+2)/\eta = -2 - 2/\eta$ for $\eta = r+\delta-1 > 0$; it follows that α_m is necessarily of size $-\beta/(\beta-2) = -1 - 2/\gamma$ for $\gamma = \beta-2 = \eta$ i.e. for $\beta = 2+\eta = r+\delta+1$. (Note that the moment condition of Lemma 2.1 is still implied by (a).) It is also simple to verify that the mixing decay conditions (in summation form) of Lemma 2.2 are also necessarily satisfied by assumption (b) of the Theorem. Finally, we note that the assumptions of the Theorem ensure that the conditions of Lemma 5.1 hold and, therefore, $\sigma_T^2 \to \sigma_{T,\ell}^2 \to 0$ as $T \uparrow \infty$.

Now

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

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 [39, pp. 155-157]. It is necessary, however, to correct the error that occurs on page 156 of [39] in the use of Lemma 6.19 of [39]. When one allows for the fact that s may increase with T the conclusion of Lemma 6.19 should be amended to (the details are omitted but may be obtained from the author on request):

$$E(\Sigma_{t=s+1}^{T}Z_{ts})^2 \leq s(T-s)\Delta < sT\Delta$$

for a suitable constant Δ . It follows that (see p. 156 of [39] for details):

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

$$\leq \Sigma_{s=1}^{\ell} sT\Delta\ell^2/\varepsilon^2T^2$$

$$= \Delta\ell^3(\ell+1)/2\varepsilon^2T$$

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 6.1

By Theorems 3.1 and 5.2 we have as $T \uparrow \infty$:

$$T(\hat{\alpha}-1) \Rightarrow \frac{(1/2)(W(1)^2 - \sigma_u^2/\sigma^2)}{\int_0^1 W(t)^2 dt} ,$$

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

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

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

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

and (b) follows as required.

Proof of Lemma 7.1.

$$\begin{aligned} Y_h(t) &= \sigma^{-1}\Sigma_1^{i-1} u(jh). , \quad (ih-h)/N \le t < ih/N \\ &= h^{1/2}\sigma^{-1}\Sigma_1^{i-1}\xi_j \\ &= N^{1/2}(1/\sqrt{T}\sigma)\Sigma_1^{i-1}\xi_j \\ &= N^{1/2}X_T(t) \end{aligned}$$

for $(i-1)/T \le t < i/T$ where $T = N/h$. $X_T(t) \Rightarrow W(t)$ by Lemma 2.2 and the result follows.

Proof of Theorem 7.2

(a) $$h\Sigma_1^T y(th-h)^2 = N\Sigma_{i=1}^T T^{-1}(\Sigma_1^{i-1}u(jh) + y_0)^2$$

$$= N\Sigma_{i=1}^T T^{-1}\sigma^2(Y_h(t) + y_0/\sigma)^2$$

$$= N^2\sigma^2\int_0^1 X_T(t)^2 dt + 2N^{3/2}\sigma y_0\int_0^1 X_T(t)dt + Ny_0^2$$

$$\Rightarrow N^2\sigma^2\int_0^1 W(t)^2 dt + 2N^{3/2}\sigma y_0\int_0^1 W(t)dt + Ny_0^2 .$$

as required. In the above the random element

$$X_T(t) = \Sigma_1^{i-1}\xi_j/\sqrt{T}\sigma , \quad (j-1)/T \leq t < j/T$$

lies in $D[0,1]$ and the ξ_j satisfy the assumptions of Lemma 2.2, so that $X_T(t) \Rightarrow W(t)$.

(b) $$\Sigma_1^T y(th-h)\{y(th) - y(th-h)\}$$

$$= \Sigma_{i=1}^T(\Sigma_0^{i-1}u(jh) + y_0)u(ih)$$

$$= h\Sigma_1^T(\Sigma_0^{i-1}\xi_j)\xi_i + y_0 h^{1/2}\Sigma_1^T\xi_i$$

$$= N\sigma^2\Sigma_1^T(\Sigma_0^{i-1}\xi_j/\sqrt{T}\sigma)(\xi_i/\sqrt{T}\sigma) + y_0N^{1/2}\sigma\Sigma_1^T\xi_i/\sqrt{T}\sigma$$

$$= (N\sigma^2/2)\{\Sigma_1^T[X_T(t)^2]_{(i-1)/T}^{i/T} - T^{-1}\Sigma_1^T\xi_i^2/\sigma^2\} + y_0N^{1/2}\sigma X_T(1)$$

$$= (N\sigma^2/2)X_T(1)^2 - (N/2T)\Sigma_1^T\xi_i^2 + y_0N^{1/2}\sigma X_T(1)$$

$$\Rightarrow (N\sigma^2/2)W(1)^2 - (N/2)\sigma^2 + y_0N^{1/2}\sigma W(1)$$

as required. (c) follows directly from the expression

$$\hat{\alpha}-1 = \Sigma_1^T y(th-h)\{y(th) - y(th-h)\}/\Sigma_1^T y(th-h)^2$$

and the arguments above involving the continuous mapping theorem and Lemma 7.1.

<u>Proof of Theorem 7.3.</u> From the solution (27) we have (under the null hypothesis $\theta = 0$)

$$y(t) = \int_0^t \zeta(dr) + y(0)$$

and thus

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

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

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

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

and

$$\int_0^N y\,dy = \sigma^2\int_0^N W\,dW + \sigma y(0)\int_0^N dW$$

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

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

and the required result follows.

REFERENCES

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

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

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

[4] Bergstrom, A. R. *Statistical Inference in Continuous Time Economic Models.* North Holland: Amsterdam, 1976.

[5] Bergstrom, A. R. "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, 1984.

[6] Bhargava, A. "On the theory of testing for unit roots in observed time series," mimeographed, LSE, 1983.

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

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

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

[10] Chung, K. L. *A Course in Probability Theory.* Academic Press: New York, 1974.

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

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

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

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

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

[16] Erdös, P. and M. Kac. "On certain limit theorems in the theory of probability," *Bulletin of the American Mathematical Society*, 52 (1946), 292-302.

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

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

[19] Fuller, Wayne A. *Introduction to Statistical Time Series*. John Wiley & Sons: New York, 1976.

[20] Fuller, Wayne A. and D. P. Hasza. "Predictors for the first order autoregressive process," *Journal of Econometrics*, 13 (1980), 139-157.

[21] Granger, C. W. J. and P. Newbold. *Forecasting Economic Time Series*. Academic Press: New York, 1977.

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

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

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

[25] Hasza, D. P. and W. A. Fuller. "Estimation for autoregressive processes with unit roots," *Annals of Statistics*, 7 (1979), 1106-1120.

[26] Hida, T. *Brownian Motion*. Springer Verlag: New York, 1980.

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

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

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

[30] McLeish, D. L. "On the invariance principle for nonstationary mixingales," *Annals of Probability*, 5 (1977), 616-621.

[31] Nankervis, J. C. and N. E. Savin. "Testing the autoregressive parameter with the t statistic," *Journal of Econometrics* (to appear).

[32] Nankervis, J. C. and N. E. Savin. "The student's t approximation in a stationary first order autoregressive model," mimeographed, Trinity College, 1984.

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

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

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

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

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

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

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

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

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

[42] Yaglom, A. M. *An Introduction to the Theory of Stationary Random Functions.* Dover: New York, 1973.