Towards a unified asymptotic theory for autoregression

By P. C. B. PHILLIPS

*Cowles Foundation for Research in Economics, Yale University, Yale Station, New Haven,
Connecticut 06520-2125, U.S.A.*

SUMMARY

This paper develops an asymptotic theory for a first-order autoregression with a root near unity. Deviations from the unit root theory are measured through a noncentrality parameter. When this parameter is negative we have a local alternative that is stationary; when it is positive the local alternative is explosive; and when it is zero we have the standard unit root theory. Our asymptotic theory accommodates these possibilities and helps to unify earlier theory in which the unit root case appears as a singularity of the asymptotics. The general theory is expressed in terms of functionals of a simple diffusion process. The theory has applications to continuous time estimation and to the analysis of the asymptotic power of tests for a unit root under a sequence of local alternatives.

Some key words: Autoregression; Brownian motion; Diffusion; Near-integrated process; Noncentrality parameter; Unit root.

1. INTRODUCTION

There has recently been a growing interest in the asymptotic theory of autoregressive time series with roots on or near the unit circle. Fuller (1976) and Dickey & Fuller (1979, 1981) developed statistical tests for detecting the presence of a unit root in a first-order autoregressive process, AR(1). Subsequent papers (Solo, 1984; Said & Dickey, 1984; Phillips, 1987) have extended these procedures to quite general integrated time series of the ARIMA(p, 1, q) type. The limiting distributions of the various test statistics proposed in these papers are known and are all nonnormal. However, these limiting distributions may usually be represented quite simply in terms of functionals of standard Brownian motion. Moreover, tables of the relevant distributions have been obtained by Monte Carlo methods for the asymptotic case, as well as for a range of finite sample sizes. These are reported by Fuller (1976) and Dickey & Fuller (1979, 1981).

Autoregressive time series with roots that are near unity have also been studied. Evans & Savin (1981, 1984) found in extensive simulation experiments that the statistical properties of the coefficient estimator and associated t test in a stationary AR(1) with a root near unity are close to those that apply when the model is a random walk, even when the sample size is as large as 100. Similar results were found when the AR(1) is mildly explosive. In related work, Ahtola & Tiao (1984) recently studied the sampling behaviour of the score function in an AR(1) as the autoregressive coefficient approaches

The present paper deals with a closely related subject. We shall consider a time series that is generated by the model

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

with

$$a = \exp(c/T) \quad (-\infty < c < \infty), \tag{2}$$

where T is the sample size. Since the coefficient in this autoregression depends on T, time series generated by (1) formally constitute a triangular array of the type $\{y_{tT}: t = 1, \ldots, T; T = 1, 2, \ldots\}$. However, with the exception of § 5, this formality is not essential to our discussion and we shall simply refer to time series generated by (1) as $\{y_t\}$. Initial conditions for (1) are set at $t = 0$ and y_0 may be any random variable, including a constant, whose distribution is fixed and independent of T.

It is convenient to treat the parameter c in (2) as a noncentrality parameter. When $c = 0$ the model (1) has a unit root. When $c < 0$ and T is finite, $0 < a < 1$ and the model is evidently stable over a finite stretch of data. Similarly, when $c > 0$ and T is finite it is clear that $a > 1$ and the model has explosive characteristics in finite samples of data. When c is close to zero, a is close to unity and the model may be thought of as having a root that is local to unity. In this case (1) and (2) comprise a nearly nonstationary AR(1) of the type considered by Ahtola & Tiao (1984). However, Ahtola & Tiao deal only with stable alternatives ($a < 1$) and they require the sequence of innovations $\{u_t\}$ in (1) to be independently and identically distributed as $N(0, \sigma^2)$, so that their model is a conventional Gaussian AR(1). Much weaker conditions on $\{u_t\}$ will be employed in this paper. As a result, the asymptotic theory that we develop will apply to rather a wide class of nearly nonstationary processes. In particular, our approach is to require that the sequence of innovations $\{u_t\}$ in (1) satisfy some rather general moment and weak dependence conditions. Under these conditions $\{u_t\}$ may be generated by a wide variety of models, including all Gaussian and many other finite-order ARMA models.

If $T \to \infty$ while the noncentrality parameter is held fixed we see that $a \to 1$. Thus, in the limit as $T \to \infty$ the model (1) has a unit root. Note that the rate of approach to unity is controlled at $O(T^{-1})$. This is the order of consistency of the least-squares estimator of the coefficient in (1) when there is a unit root. It might be anticipated that the main effect of the hypothesis (2) is to induce noncentrality in the limiting distribution theory. It turns out that the asymptotic theory is indeed noncentral. The relevant limiting distributions are most conveniently represented as functionals of a first-order diffusion process, rather than standard Brownian motion. The coefficient in the diffusion process is the noncentrality parameter c. This parameter may be used to measure the effects of the departures from the hypothesis of a unit root in (1) on the limiting distribution theory. Moreover, the resulting noncentral limiting distribution theory yields the asymptotic power functions of statistical tests for a unit root under a sequence of local alternatives to unity.

The paper is organized as follows. Some preliminary theory is presented in § 2. Section 3 develops the general theory which accommodates autoregressive roots in the vicinity of unity. In § 4 we examine how these results change as the noncentrality parameter approaches the boundaries of its domain of definition. Section 5 applies the results to

2. SOME PRELIMINARY THEORY

Throughout this paper we assume that the innovation sequence $\{u_t\}$ satisfies the following general conditions:

(i) $E(u_t) = 0$ for all t;

(ii) $\sup_t E|u_t|^{\beta+\varepsilon} < \infty$ for some $\beta > 2$ and $\varepsilon > 0$;

(iii) as $T \to \infty$, $\sigma^2 = \lim E(T^{-1}S_T^2)$ exists and $\sigma^2 > 0$, where $S_t = u_1 + \ldots + u_t$;

(iv) $\{u_t\}$ is strong mixing with mixing coefficients α_m that satisfy

$$\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty. \tag{3}$$

Condition (iv) imposes a form of asymptotic weak dependence on the sequence of innovations $\{u_t\}$. The reader is referred, for example, to Hall & Heyde (1980, p. 132) for the definition of strong mixing and the mixing coefficients α_m. The summability requirement (3) on the mixing coefficients is satisfied when $\alpha_m = O(m^{-\lambda})$ for some $\lambda > \beta/(\beta - 2)$. Condition (ii) controls the allowable heterogeneity in the sequence $\{u_t\}$ in relation to the mixing decay rate prescribed by (3). Thus, as β declines towards 2 the moment condition (ii) weakens and the probability of outliers rises. On the other hand, the mixing decay rate $\beta/(\beta - 2)$ increases as β approaches 2 and the effect of outliers is required by condition (iii) to wear off more quickly. Condition (iii) is a convergence condition on the average variance of the partial sum S_T. It is a common requirement in much central limit theory. If $\{u_t\}$ is weakly stationary then

$$\sigma^2 = E(u_1^2) + 2 \sum E(u_1 u_k), \tag{4}$$

where the sum is over $k = 2, \ldots, \infty$, and the convergence of the series is implied by the mixing condition (iii), as proved in Theorem 18.5.3 of Ibragimov & Linnik (1971, p. 346). As is conventional, we still require $\sigma^2 > 0$ to exclude degenerate results.

Conditions (i)-(iv) are quite weak. They permit the innovation sequence $\{u_t\}$ to be heterogeneously distributed and weakly dependent over time. This includes a wide variety of possible data generating mechanisms such as all Gaussian and many other finite-order ARMA models under very general conditions on the underlying errors (Withers, 1981).

From the sequence of partial sums $\{S_t\}$ we construct

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

where $[Tr]$ denotes the integral part of Tr. The random element $X_T(r)$ lies in $D = D[0, 1]$, the space of real-valued functions on the interval $[0, 1]$ that are right continuous and have finite left limits. 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 $f, g \in D$.

Under conditions (i)-(iv) the random element $X_T(r)$ obeys a central limit theory on the function space D. In particular, from Herrndorf (1984) we have, as $T \to \infty$,

$$X_T(r) \Rightarrow W(r). \tag{5}$$

The symbol '$\Rightarrow$' here signifies weak convergence of the associated probability measures. Thus the probability measure of $X_T(r)$ converges weakly to the probability measure, namely Wiener measure, of the random function $W(r)$. The result is known as a functional central limit theorem or invariance principle and the limit process $W(r)$ is nevertheless

Definition. A time series $\{y_t\}$ that is generated by (1) and (2) with $c \neq 0$ and where $\{u_t\}$ satisfies (i)–(iv) is called near-integrated. When $c = 0$, in (2), $\{y_t\}$ is called an integrated process.

The terminology we employ here for an integrated process corresponds to the usage popularized by Box & Jenkins (1976) when $\{u_t\}$ is generated by a stationary ARMA model. The above definition actually extends the terminology to include time series whose first differences are not necessarily stationary processes and may be generated, for example, by finite-order ARMA models whose innovations are nonidentically distributed. When $c \neq 0$, the specification (2) allows us to introduce the closely related concept of a near-integrated process. The latter includes alternatives which are strongly autoregressive $(c < 0)$ or mildly explosive $(c > 0)$ in finite samples of data.

The following functional will play a central role in our theory:

$$J_c(r) = \int_0^r e^{(r-s)c}\, dW(s).$$

Here $J_c(r)$ is a Gaussian process which, for fixed $r > 0$, has the distribution

$$J_c(r) \equiv N(0, \tfrac{1}{2}(e^{2rc} - 1)/c), \tag{6}$$

where we use the symbol '$\equiv$' to signify equality in distribution. Actually, $J_c(r)$ is an Ornstein–Uhlenbeck process, generated by the stochastic differential equation

$$dJ_c(r) = cJ_c(r)\, dr + dW(r), \tag{7}$$

with initial condition $J_c(0) = 0$. It is simple to establish that

$$J_c(r) = W(r) + c \int_0^r e^{(r-s)c} W(s)\, ds,$$

and by stochastic differentiation of $\{\int e^{-sc}\, dW(s)\}^2$, where the integral is over $(0, r)$, we deduce the following useful relationship:

$$\{J_c(1)\}^2 = 1 + 2c \int_0^1 \{J_c(r)\}^2\, dr + 2 \int_0^1 J_c(r)\, dW(r). \tag{8}$$

When $c = 0$, equation (8) reduces to the commonly-used formula

$$\int_0^1 W(r)\, dW(r) = \tfrac{1}{2}[\{W(1)\}^2 - 1].$$

In these expressions, $\int J_c\, dW$ and $\int W\, dW$, where the integrals are over $(0, 1)$, are interpreted as stochastic integrals.

3. ASYMPTOTICS FOR NEAR-INTEGRATED PROCESSES

Our first step is to find the relevant asymptotic theory for the sample moments of data generated by (1) and (2). For integrated time series, the limiting distribution theory is

The following lemma has all the results we need for the development of our regression theory.

LEMMA 1. *If* $\{y_t\}$ *is a near-integrated time series generated by* (1) *and* (2), *then, as* $T \to \infty$:

(a) $T^{-\frac{1}{2}} y_{[Tr]} \Rightarrow \sigma J_c(r)$;

(b) $T^{-3/2} \sum y_t \Rightarrow \sigma \int J_c(r) \, dr$, *where the integral is over* (0, 1);

(c) $T^{-2} \sum y_t^2 \Rightarrow \sigma^2 \int J_c(r)^2 \, dr$, *where the integral is over* (0, 1);

(d) $T^{-1} \sum y_{t-1} u_t \Rightarrow \sigma^2 \int J_c(r) \, dW(r) + \frac{1}{2}(\sigma^2 - \sigma_u^2)$, *where the integral is over* (0, 1);

with $\sigma_u^2 = \lim T^{-1} \sum E(u_t^2)$. *Joint weak convergence of* (a)-(d) *also applies.*

Proof. The approach we follow is based on Phillips (1987). To prove (a) we first note from (1) and (2) that

$$y_t = \sum_{j=1}^{t} e^{(t-j)c/T} u_j + e^{tc/T} y_0$$

and thus

$$T^{-\frac{1}{2}} y_{[Tr]} = \sigma \sum_{j=1}^{[Tr]} e^{([Tr]-j)c/T} \int_{(j-1)/T}^{j/T} dX_T(s) + O_p(T^{-\frac{1}{2}})$$

$$= \sigma \sum_{j=1}^{[Tr]} \int_{(j-1)/T}^{j/T} e^{(r-s)c} \, dX_T(s) + O_p(T^{-\frac{1}{2}})$$

$$= \sigma \int_0^r e^{(r-s)c} \, dX_T(s) + O_p(T^{-\frac{1}{2}}).$$

We now use integration by parts on the first term, which is valid since $e^{(r-s)c}$ is continuous and $X_T(s)$ is increasing and of bounded variation. We obtain, as $T \to \infty$,

$$\sigma \left\{ X_T(r) + c \int_0^r e^{(r-s)c} X_T(s) \, ds \right\} + O_p(T^{-\frac{1}{2}}) \Rightarrow \sigma W(r) + c \int_0^r e^{(r-s)c} W(s) \, ds = \sigma J_c(r)$$

by (5) and the continuous mapping theorem (Billingsley, 1968, p. 30). This proves (a). The proofs of (b) and (c) are entirely similar. To prove (d) note by squaring (1) and summing over t that

$$T^{-1} y_T^2 = 2cT^{-2} \sum y_{t-1}^2 + T^{-1} \sum u_t^2 + 2T^{-1} \sum y_{t-1} u_t + O_p(T^{-\frac{1}{2}}).$$

Now $T^{-1} \sum u_t^2 \to \sigma_u^2$ almost surely, by the strong law of large numbers for weakly dependent sequences; see, in particular, McLeish (1975, Th. 2.10). From parts (a) and (c) and the continuous mapping theorem we now deduce that, as $T \to \infty$,

$$2T^{-1} \sum y_{t-1} u_t \Rightarrow \sigma^2 \{J_c(1)\}^2 - 2c\sigma^2 \int_0^1 \{J_c(r)\}^2 \, dr - \sigma_u^2 = 2\sigma^2 \int_0^1 J_c(r) \, dW(r) + \sigma^2 - \sigma_u^2$$

in view of (8). Part (d) of the lemma follows immediately. To establish joint convergence of (a)-(d) we need only note that the vector of sample moments may also be expressed as a continuous functional of $X_T(s)$ up to an error of $O_p(T^{-\frac{1}{2}})$. The required result then follows.   □

sample moments of nearly nonstationary time series. Thus, since $J_c(r)$ is Gaussian it is easy to show by elementary calculations that

$$\int_0^1 J_c(r) \, dr \equiv N(0, v),$$

where $v = 1/c^2 + \frac{1}{2}(1/c^3)(e^{2c} - 4e^c + 3)$. Parts (a) and (b) of Lemma 1 therefore yield

$$T^{-\frac{1}{2}} y_T \Rightarrow \sigma J_c(1) \equiv N\{0, \tfrac{1}{2}(\sigma^2/c)(e^{2c} - 1)\}, \tag{9}$$

$$T^{-3/2} \sum y_t \Rightarrow \sigma \int_0^1 J_c(r) \, dr \equiv N(0, \sigma^2 v). \tag{10}$$

When $c = 0$, expression (9) is $\sigma J_0(1) \equiv N(0, \sigma^2)$, which is the limiting distribution of the standardized sum $T^{-\frac{1}{2}} \sum u_t$ of the innovations in (1). The variance of this limiting distribution is $\sigma^2 = \lim T^{-1} E(S_T^2)$. When $\{u_t\}$ is stationary we have $\sigma^2 = 2\pi f_u(0)$, where $f_u(\lambda)$ is the spectral density of $\{u_t\}$. In this special case of (9) we therefore find that

$$T^{-\frac{1}{2}} \sum u_t \Rightarrow N\{0, 2\pi f_u(0)\}.$$

This is, of course, a general central limit theorem for stationary time series; see, for example, Hannan (1970, Th. 11, p. 221).

When $c = 0$ in (10) a simple calculation gives $\sigma \int J_0(r) \, dr \equiv N(0, \frac{1}{3} \sigma^2)$, where the integral is over $(0, 1)$, which is the limiting distribution of the standardized sample mean of an integrated process (Phillips, 1987).

Perhaps the most useful application of these results is to the theory of regression for near-integrated time series. Suppose (1) is estimated by least-squares giving the regression coefficient $\hat{a} = \sum y_t y_{t-1} / \sum y_{t-1}^2$, and associated t statistic $t_a = (\sum y_{t-1}^2)^{\frac{1}{2}}(\hat{a} - a)/s$, where $s^2 = T^{-1} \sum (y_t - \hat{a} y_{t-1})^2$. The asymptotic theory of these regression statistics is as follows.

THEOREM 1. *If* $\{y_t\}$ *is a near-integrated time series generated by* (1) *and* (2), *then, as* $T \to \infty$:

   (a) $T(\hat{a} - a) \Rightarrow \{\int J_c(r) \, dW(r) + \frac{1}{2}(1 - \sigma_u^2/\sigma^2)\} / \int \{J_c(r)\}^2 \, dr$, *where the integrals are over* $(0, 1)$;

   (b) $\hat{a} \to 1, s^2 \to \sigma_u^2$ *in probability*;

   (c) $t_a \Rightarrow (\sigma/\sigma_u)\{\int J_c(r) \, dW(r) + \frac{1}{2}(1 - \sigma_u^2/\sigma^2)\} / [\int \{J_c(r)\}^2 \, dr]^{\frac{1}{2}}$, *where the integrals are over* $(0, 1)$.

Proof. To prove part (a) we note that

$$T(\hat{a} - a) = (T^{-2} \sum y_{t-1}^2)^{-1}(T^{-1} \sum y_{t-1} u_t)$$

$$\Rightarrow \left[\int_0^1 \{J_c(r)\}^2 \, dr \right]^{-1} \left\{ \int_0^1 J_c(r) \, dW(r) + \tfrac{1}{2}(1 - \sigma_u^2/\sigma^2) \right\}$$

by direct application of the continuous mapping theorem and Lemma 1. Moreover, this implies that $\hat{a} = a + O_p(T^{-1}) = 1 + O_p(T^{-1})$ so that part (b) also follows. Part (c) is an immediate consequence of Lemma 1, the continuous mapping theorem and part (b). □

When $c = 0$ and $\{u_t\}$ is independently and identically distributed with zero mean and

1979). In particular, (a) and (c) become

$$T(\hat{a}-1)\Rightarrow\left[\int_0^1 \{W(r)\}^2\,dr\right]^{-1}\left\{\int_0^1 W(r)\,dW(r)\right\},$$

$$t_a\Rightarrow\left[\int_0^1 \{W(r)\}^2\,dr\right]^{-\frac{1}{2}}\left\{\int_0^1 W(r)\,dW(r)\right\}.$$

In this case, we have $\sigma^2 = \sigma_u^2$ and $J_c(r) = W(r)$ in the formulae of Theorem 1.

When $c \neq 0$ Theorem 1 delivers the noncentral asymptotic theory for the regression statistics $\hat{a}$ and t_a for a very general class of innovations in (1). The theory is particularly useful in studying the asymptotic power of tests for a unit root under the sequence of local alternatives given by $a = e^{c/T} \sim 1 + c/T$. Dickey & Fuller (1979) suggested the test statistics $T(\hat{a}-1)$ and t_1. These statistics are appropriate when the errors in (1) are independent and identically distributed, in which case we have $\sigma^2 = \sigma_u^2$ and

$$T(\hat{a}-1)\Rightarrow c+\left[\int_0^1 \{J_c(r)\}^2\,dr\right]^{-1}\left\{\int_0^1 J_c(r)\,dW(r)\right\},$$

$$t_1\Rightarrow c\left\{\int_0^1 J_c(r)\,dr\right\}^{\frac{1}{2}}+\left[\int_0^1 \{J_c(r)\}^2\,dr\right]^{-\frac{1}{2}}\left\{\int_0^1 J(r)\,dW(r)\right\}.$$

The Dickey-Fuller tests have recently been extended by the author to accommodate rather general time series with a unit root. The new statistics Z_a and Z_t (Phillips, 1987) are based on $T(\hat{a}-1)$ and t_1 but they employ a nonparametric correction for serial correlation. Under the sequence of alternatives $a = e^{c/T}$, Theorem 1 provides asymptotic power functions for these new tests as well; Z_a and Z_t have the same asymptotic local power as the Dickey-Fuller tests $T(\hat{a}-1)$ and t_1 given above, yet the new tests allow for a much wider class of error processes.

4. LIMIT DISTRIBUTIONS AS $c \to \pm\infty$

It is interesting to study the limiting behaviour of the asymptotic theory of § 3 as the noncentrality parameter approaches the boundaries of its domain of definition. As might be expected, the results for the sample moments are different at the two boundaries and two different normalizations are required to eliminate degeneracies. The central results we shall use are contained in the following lemma which is proved in the Appendix.

LEMMA 2. *As* $c \to -\infty$:

(a) $(-2c)\int\{J_c(r)\}^2\,dr\to 1$ *in probability, where the integral is over* $(0,1)$;

(b) $(-2c)^{\frac{1}{2}}\int J_c(r)\,dW(r)\Rightarrow N(0,1)$, *where the integral is over* $(0,1)$;

As $c \to +\infty$,

(c) $[(2c)\,e^{-c}\int J_c(r)\,dW(r),(2c)^2\,e^{-2c}\int\{J_c(r)\}^2\,dr]\Rightarrow(\xi\eta,\eta^2)$, *where the integrals are over* $(0,1)$;

where ξ and η are independent $N(0,1)$ *variates.*

We now define

$$K_2(c) = g(c)^{\frac{1}{2}}\left[\int_0^1 \{J_c(r)\}^2\,dr\right]^{-1}\left\{\int_0^1 J_c(r)\,dW(r)+\tfrac{1}{2}(1-\sigma_u^2/\sigma^2)\right\},$$

where

$$g(c) = E\left[\int_0^1 \{J_c(r)\}^2\, dr\right] = -(2c)^{-1}\{1 + (2c)^{-1}(1 - e^{2c})\}.$$

Here $K_2(c)$ and $K_3(c)$ are functionals which represent the limiting distributions of the standardized regression coefficient $g(c)^{\frac{1}{2}} T(\hat{a} - a)$ and the associated t ratio t_a. The behaviour of $K_2(c)$ and $K_3(c)$ for large $|c|$ is now a simple consequence of Lemma 2.

THEOREM 2. *As* $c \to -\infty$:

  (a) $K_2(c) \Rightarrow N(0, 1)$ *if* $\sigma^2 = \sigma_u^2$ *and diverges otherwise,*
  (b) $K_3(c) \Rightarrow N(0, 1)$ *if* $\sigma^2 = \sigma_u^2$ *and diverges otherwise.*

As $c \to +\infty$:

  (c) $K_2(c) \Rightarrow C$,
  (d) $K_3(c) \Rightarrow N(0, \sigma^2/\sigma_u^2)$,

where C denotes the standard Cauchy distribution.

Proof. Parts (a) and (b) follow directly from Lemma 2 upon appropriate standardization of numerator and denominator. To prove part (c), we write

$$K_2(c) = (2c)^{-1} e^c \left[\int_0^1 \{J_c(r)\}^2 dr\right]^{-1} \left\{\int_0^1 J_c(r)\, dW(r) + \tfrac{1}{2}(1 - \sigma_u^2/\sigma^2)\right\} + o_p(1)$$

$$= \left[(2c)^2\, e^{-2c} \int_0^1 \{J_c(r)\}^2\, dr\right]^{-1} \left\{2c e^{-c} \int_0^1 J_c(r)\, dW(r) + c e^{-c}(1 - \sigma_u^2/\sigma^2]\right\} + o_p(1)$$

$$\Rightarrow \xi\eta/\eta^2 = \xi/\eta \equiv C,$$

as $c \to \infty$ by Lemma 2 and the continuous mapping theorem. In a similar way we find

$$K_3(c) = (\sigma/\sigma_u)\left[\int_0^1 \{J_c(r)\}^2\, dr\right]^{-\frac{1}{2}} \left\{\int_0^1 J_c(r)\, dW(r) + \tfrac{1}{2}(1 - \sigma_u^2/\sigma^2)\right\}$$

$$= (\sigma/\sigma_u)\left[(2c)^2 e^{-2c} \int_0^1 \{J_c(r)\}^2\, dr\right]^{-\frac{1}{2}}$$

$$\times \left\{(2c)\, e^{-c} \int_0^1 J_c(r)\, dW(r) + c e^{-c}(1 - \sigma_u^2/\sigma^2)\right\}$$

$$\Rightarrow (\sigma/\sigma_u)\xi\eta/|\eta| = (\sigma/\sigma_u)\xi \operatorname{sgn}(\eta) \equiv N(0, \sigma^2/\sigma_u^2)$$

as $c \to \infty$, proving part (d).   □

The above results are obtained by studying asymptotic behaviour in successive limits: first as $T \to \infty$ with c taken to be a fixed constant; and secondly as $c \to \pm\infty$. Heuristically, one might expect the asymptotic results of Theorem 2 to provide reasonable approximations in finite samples for which both T and $|c|$ are large. Note, in particular, from (2) that $c = T \log a$, and thus finite-sample configurations with large T and large $|c|$ may be associated with stable or explosive AR(1)'s depending on the sign of c. In this sense the results of the theorem are suggestive. In fact, parts (a) and (b) correspond with known asymptotic theory for the stationary AR(1); and parts (c) and (d) correspond with asymptotic results obtained by White (1958, 1959) and Anderson (1959) for the explosive

with a fixed coefficient. This point has been emphasized by a referee. The reason is that, in general and without further conditions, one cannot deduce rigorous asymptotic results that apply for $T \to \infty$ with the coefficient a fixed by telescoping the limits as $T \to \infty$ and $c \to \pm\infty$.

Theorem 2 relates to some recent independent work by N. H. Chan and C. Z. Wei in a University of Maryland research report. These authors take the case of independent and identically distributed innovations $\{u_t\}$ and by an approach that is quite different from that used here they obtain parts (b) and (d) of Theorem 2, both for the special case in which $\sigma^2 = \sigma_u^2$.

5. Regressions with continuous time observations

Models such as (1) and (2) have another interesting interpretation and application. Let $\{y_{nt} : t = 1, \ldots, T_n; n = 1, \ldots\}$ be an autoregressive array generated for each row n by

$$y_{nt} = a_n y_{nt-1} + u_{nt} \quad (t = 1, \ldots, T_n), \quad y_{n0} = y(0), \tag{11}$$

in which the innovations u_{nt} are independent and identically distributed with zero mean and variance $\sigma^2 h_n$. Now let $h_n = 1/T_n$ and define

$$a_n = e^{ch_n} = e^{c/T_n}. \tag{12}$$

Each row of the triangular array $\{y_{nt}\}$ may be interpreted as an autoregression in discrete time with sampling interval h_n. We require $T_n \to \infty$ as $n \to \infty$, so that $h_n \to 0$. The array then represents a sequence of autoregressions with sampling intervals that decrease as we get deeper into the array. For each value of n, expressions (11) and (12) are just a special case of the earlier model (1) and (2).

Note that by definition $T_n h_n = 1$ so that $\{y_{nt}; t = 1, \ldots, T_n\}$ may be regarded as equispaced observations of a continuous stochastic process over the interval $[0, 1]$. In fact, by methods analogous to those of the proof of Lemma 1(a), it is easy to show that, as $T_n \to \infty$,

$$y_{n[T_n r]} \Rightarrow \sigma J_c(r) + e^{rc} y(0). \tag{13}$$

Thus, setting $\sigma = 1$ and initial conditions $y(0) = 0$, we find that the triangular array $\{y_{nt}\}$ converges weakly as $n \to \infty$ to the diffusion $J_c(r)$ over the unit interval $0 \le r \le 1$. In this way (11) and (12) may be regarded as a discrete time autoregression whose natural limit as $n \to \infty$ is the first-order stochastic differential equation (7). Note that it is not necessary to treat $\{y_{nt}\}$ as a Gaussian process for this interpretation to be valid since (13) applies provided the innovations in (11) are independent and identically distributed. Of course, if $\{y_{nt}\}$ is Gaussian then we may go further and treat (11) and (12) as the discrete time equivalent of (7), that is the discrete time model that is satisfied almost surely by equispaced observations generated from (7).

When continuous time observations are available a natural estimator of the coefficient in (7) is the least-squares estimator

$$\hat{c} = \int_0^1 J_c(r) \, dJ_c(r) \Big/ \int_0^1 \{J_c(r)\}^2 \, dr.$$

This estimator was originally suggested by Bartlett (1946, 1978) and Grenander (1950)

Since

$$\hat{c} - c = \int_0^1 J_c(r)\, dW(r) \Big/ \int_0^1 \{J_c(r)\}^2\, dr, \tag{14}$$

the finite-sample distribution of $\hat{c} - c$ is identical to the asymptotic distribution of the corresponding statistic $T_n(\hat{a}_n - a_n)$ that is based on a discrete time record, taken as $T_n \to \infty$. Similarly we introduce the continuous record t statistic

$$t_c = \int_0^1 J_c(r)\, dW(r) \Big/ \left[\int_0^1 \{J_c(r)\}^2\, dr \right]^{\frac{1}{2}} \tag{15}$$

whose distribution is identical to that of the asymptotic distribution of the regression t statistic t_{a_n} from discrete time data.

From Theorem 2 we now deduce the asymptotic behaviour of these continuous record statistics as c approaches the limits of its domain of definition. Thus

$$g(c)^{\frac{1}{2}}(\hat{c} - c) \Rightarrow N(0, 1), \quad t_c \Rightarrow N(0, 1), \tag{16}$$

as $c \to -\infty$; and

$$g(c)^{\frac{1}{2}}(\hat{c} - c) \Rightarrow C, \quad t_c \Rightarrow N(0, 1), \tag{17}$$

as $c \to +\infty$, where C is the standard Cauchy distribution.

These results together with (14) and (15) completely characterize the distributional behaviour of the continuous-time regression statistics. Note that (14) and (15) apply whether the stochastic differential equation (7) is stable with $c < 0$, explosive with $c > 0$ or simply a continuous-time random walk with $c = 0$. The asymptotic distributions given by (16) and (17) for the boundary cases $c = \pm\infty$ complete the theory.

6. SUPPLEMENTARY REMARKS

Many observed time series in the physical and social sciences are well modelled by integrated processes of the ARIMA type. However, we are frequently uncertain whether the process has a root of unity or a root in the vicinity of unity, and the discriminatory power of tests for the presence of a unit root is rather low against such alternatives. The present paper develops an asymptotic theory of autoregression covering the possibility of a root near unity. This theory has several advantages.

First, it may be used to construct distributional approximations for regression statistics in mildly explosive, strongly autoregressive or unit root models. Secondly it provides a mechanism by which we can obtain asymptotic power functions for unit root tests under a sequence of local alternatives. The local departures from the unit root theory are then measured through a noncentrality parameter which figures in the asymptotic theory. Finally, the theory enables us to obtain a very convenient unification of the asymptotic theory for autoregressions with roots in the vicinity of unity. This is useful because in previous work the unit root case has been viewed as a singularity of the asymptotic theory. In the new theory for near integrated processes that is developed in this paper it becomes a simple special case, the case where the noncentrality parameter is zero.

then $y_t = \mu(a^t - 1)/(a - 1) + y_t^0$, where y_t^0 is driven by a model such as (1). When $a = 1$, $y_t = \mu t + y_t^0$ and y_t is dominated by a deterministic trend. Similarly, when $a \doteq 1$, for example $a = \exp(c/T^{3/2})$, we still find that $y_t \sim \mu t$ when the drift $\mu \neq 0$. In this case conventional normal asymptotics obtain for regression statistics such as $T^{3/2}(\hat{a} - a)$ and t_a. When $\mu = 0$ and (18) is estimated with a fitted drift, results analogous to those of the present paper apply. One needs only to make simple modifications to the formulae which account for the fitted mean. For example, we have

$$T^{-2} \sum (y_t - \bar{y})^2 \Rightarrow \sigma^2 \left[\int_0^1 \{J_c(r)\}^2 \, dr - \left\{ \int_0^1 J_c(r) \, dr \right\}^2 \right]$$

in place of (c) in Lemma 1. Sometimes (18) is fitted with a trend as well as a drift in order to discriminate between processes which are stationary in differences rather than stationary about a trend. Such models may also be extended in the manner of the present paper to accommodate roots in the vicinity of unity and closely related results are again obtained.

ACKNOWLEDGMENT

APPENDIX

Proof of Lemma 2

To prove parts (a) and (b) we employ the invariance principle. We first find the limiting distribution of $(T^{-1} \sum y_{t-1} u_t, T^{-2} \sum y_{t-1}^2)$ when the innovation sequence $\{u_t\}$ is independent and identically distributed as $N(0, 1)$. In this case, of course, $\sigma^2 = \sigma_u^2 = 1$ and by Lemma 1 the limiting distribution is that of the functional $[\int J_c(r) \, dW(r), \int \{J_c(r)\}^2 \, dr]$, where the integrals are over $(0, 1)$. However, by the invariance principle this distribution is not dependent on the normality assumption made about the innovation sequence. The assumption is simply a device which facilitates the extraction of the mathematical form of the distribution. Moreover, relaxation of the independence assumption about the u_t leads only to the additional presence of the constants σ^2 and σ_u^2 in the limiting distributions; see, in particular, parts (c) and (d) of Lemma 1. Thus, we may extrapolate easily from the limiting distribution obtained under independent $N(0, 1)$ innovations to the general case.

When $\{u_t\}$ is independent $N(0, 1)$, $(T^{-1} \sum y_{t-1} u_t, T^{-2} \sum y_{t-1}^2)$ is a pair of quadratic forms in normal variates. The joint moment generating function of these forms may be obtained following White (1958), allowing for the representation $a = e^{c/T}$. The limit of this function as $T \to \infty$ is then the moment generating function of $[\int J_c(r) \, dW(r), \int \{J_c(r)\}^2 \, dr]$, where the integrals are over $(0, 1)$. Simple calculations along these lines yield the joint moment generating function

$$M_c(w, z) = (\tfrac{1}{2}(c^2 + 2cw - 2z)^{-\frac{1}{2}} e^{c+w}[\{(c^2 + 2cw - 2z)^{\frac{1}{2}} - (c + w)\} \exp\{(c^2 + 2cw - 2z)^{\frac{1}{2}}\}$$
$$+ \{(c^2 + 2cw - 2z)^{\frac{1}{2}} - (c + w)\} \exp\{-(c^2 + 2cw - 2z)^{\frac{1}{2}}\}])^{-\frac{1}{2}}. \tag{A1}$$

This expression holds for all c and will be used later in our derivations for explosive $(c \to \infty)$ alternatives. For our present purpose, with $c < 0$, we note that the joint moment generating function of $[(-2c)^{\frac{1}{2}} \int J_c(r) \, dW(r), (-2c) \int \{J_c(r)\}^2 \, dr]$, where the integrals are over $(0, 1)$, is

$$L_c(p, q) = M_c\{(-2c)^{\frac{1}{2}} p, -2cq\}. \tag{A2}$$

Using (A3) in (A2) we deduce that, as $c \to -\infty$, $\lim L_c(p, q) = \exp(\frac{1}{2}p^2 + q)$. It follows that

$$(-2c)^{\frac{1}{2}} \int_0^1 J_c(r) \, dW(r) \Rightarrow N(0, 1), \quad (-2c) \int_0^1 \{J_c(r)\}^2 \, dr \to 1,$$

the latter in probability as $c \to -\infty$, proving parts (a) and (b).

To prove (c) we first deduce from (A1) that the joint moment generating function of

$$\left[2c \, e^{-c} \int_0^1 J_c(r) \, dW(r), (2c)^2 \, e^{-2c} \int_0^1 \{J_c(r)\}^2 \, dr \right]$$

is

$$K_c(p, q) = M_c\{2c \, e^{-c} p, (2c)^2 \, e^{-2c} q\}. \tag{A4}$$

Now

$$\{c^2 + (2c)^2 e^{-c} p - 2(2c)^2 \, e^{-2c} q\}^{\frac{1}{2}} = c\{1 + 2e^{-c}p - 4e^{-2c}q - 2(e^{-c}p - 2e^{-2c}q)^2 + O(e^{-3c})\} \tag{A5}$$

for large positive c. Substituting (A5) into (A4) we deduce after a little calculation that, as $c \to \infty$,

$$K_c(p, q) \to (1 - p^2 - 2q)^{-\frac{1}{2}}. \tag{A6}$$

Setting $p = 0$ in (A6) we have $K_\infty(0, q) = (1 - 2q)^{-\frac{1}{2}}$, corresponding to a χ_1^2 variate. Setting $q = 0$ in (A7) we have $K_\infty(p, 0) = (1 - p^2)^{-\frac{1}{2}}$, which is the moment generating function of a product of independent $N(0, 1)$ variates. Moreover, a simple calculation shows that $K_\infty(p, q) = (1 - p^2 - 2q)^{-\frac{1}{2}}$ is the joint moment generating function of $(\xi\eta, \eta^2)$, where ξ and η are independent $N(0, 1)$ variates. Thus we also have joint weak convergence as required for part (c) of the lemma.

REFERENCES

AHTOLA, J. & TIAO, G. C. (1984). Parameter inference for a nearly nonstationary first order autoregressive model. *Biometrika* **71**, 263-72.

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30**, 676-87.

BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series, *J.R. Statist. Soc. B.* **7**, 27-41.

BARTLETT, M. S. (1978). *An Introduction to Stochastic Processes*, 3rd ed. Cambridge University Press.

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

BOX, G. E. P. & JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed. San Francisco: Holden Day.

BROWN, B. M. & HEWITT, J. I. (1975). Asymptotic likelihood theory for diffusion processes. *J. Appl. Prob.* **12**, 228-38.

DICKEY, D. A. & FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Assoc.* **74**, 427-31.

DICKEY, D. A. & FULLER, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49**, 1057-72.

EVANS, G. B. A. & SAVIN, N. E. (1981). Testing for unit roots: 1. *Econometrica* **49**, 753-79.

EVANS, G. B. A. & SAVIN, N. E. (1984). Testing for unit roots: 2. *Econometrica* **52**, 1241-69.

FEIGIN, P. D. (1979). Some comments about a curious singularity. *J. Appl. Prob.* **16**, 440-4.

FULLER, W. A. (1976). *Introduction to Statistical Time Series.* New York: Wiley.

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

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

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

HERRNDORF, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Prob.* **12**, 141-53.

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

SAID, E. S. & DICKEY, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599-607.

SOLO, V. (1984). The order of differencing in ARIMA models. *J. Am. Statist. Assoc.* 79, 916-21.

WHITE, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* 29, 1188-97.

WHITE, J. S. (1959). The limiting distribution of the serial correlation coefficient in the explosive case II. *Ann. Math. Statist.* 30, 831-4.

WITHERS, C. S. (1981). Conditions for linear processes to be strong mixing. *Z. Wahr. verw. Geb.* 57, 477-80.