A CONTINUOUS TIME APPROXIMATION TO THE STATIONARY FIRST-ORDER AUTOREGRESSIVE MODEL

Pierre Perron*
Princeton University and C.R.D.E.

Econometric Research Program
Research Memorandum No. 337

October 1988

*Assistant Professor, Department of Economics, Princeton University, and Research Associate, Centre de Recherche et Développement en Economique (Université de Montréal). I would like to thank Jean-Marie Dufour for helpful discussion. Research support is acknowledged from S.S.H.R.C. (Canada), N.S.E.R.C. (Canada), and F.C.A.R. (Québec).

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, NJ 08544-1017, USA

ABSTRACT

Consider the stationary first-order autoregressive process $y_t = \alpha y_{t-1} + e_t$, $e_t \sim IN(0,\sigma^2)$, $y_0 \sim N(0, \sigma^2(1-\alpha^2)^{-1})$ and let $\hat{\alpha}$ be the least-squares estimator of α based on a sample of size (T+1) sampled at frequency h. Consider also the continuous time diffusion process $dy_t = \Theta y_t dt + \sigma dw_t$, $y_0 \sim N(0, -\sigma^2/2\Theta)$, w_t a Wiener process, and let $\hat{\Theta}$ be the continuous time maximum likelihood (conditional upon y_0) estimator of Θ based upon a single path of data of length N. As shown in Perron (1988a), the asymptotic distribution of $T(\hat{\alpha} - \alpha)$, as the sampling interval converges to zero, is the same as the exact distribution of $N(\hat{\Theta} - \Theta)$. This distribution permits explicit consideration of the effect of the initial condition y_0 on the distribution of $\hat{\alpha}$. While our earlier work concentrated on the case where y_0 is fixed, we consider here the stationary case. The moment-generating function of $N(\hat{\Theta} - \Theta)$ is derived and used to tabulate the distribution and probability density functions. We also investigate the mean and mean-squared error of $\hat{\Theta}$ as well as the power function. For each case, the adequacy of the approximation to the finite sample distribution of $\hat{\alpha}$ is assessed. The approximations are found to be excellent for values of α not too far from one, where the usual asymptotic distributional theory is inadequate.

Key words: diffusion processes, moment-generating function, near-integrated processes, distribution theory.

Address : Pierre Perron
Department of Economics
Princeton University
Princeton, New-Jersey
U.S.A., 08540

1. INTRODUCTION

Consider the stationary first–order stochastic difference equation:

(1) $$y_t = \alpha y_{t-1} + e_t \qquad t=1,...,T$$

where $|\alpha| < 1$, $e_t \sim IN(0, \sigma_*^2)$ and $y_0 \sim N(0, \sigma_*^2 (1 - \alpha^2)^{-1})$. The least–squares estimator of α based on a sequence of observations of size (T + 1), $\{y_t\}_0^T$ is:

$$\hat{\alpha} = \Sigma_{t=1}^T y_t y_{t-1} (\Sigma_{t=1}^T y_{t-1}^2)^{-1}.$$

The distribution of $\hat{\alpha}$ has been extensively studied. One perennial topic of concern has been the adequacy of various asymptotic approximations to the finite sample distribution. The standard asymptotic result, derived by Mann and Wald (1943) and Rubin(1950), is that $T^{1/2}(\hat{\alpha} - \alpha)(1 - \alpha^2)^{-1}$ has a limiting N(0, 1) distribution when $|\alpha| < 1$. A fact of interest is that this limiting result is valid whether y_0 has the distribution specified in (1) or whether it is any fixed constant. Furthermore, the conclusion from previous studies is that the limiting normal distribution is an inadequate approximation to the finite sample distribution, especially for values of α near one (see Basmann et al. (1974), Evans and Savin (1981, 1984) among others). Asymptotic approximations to higher orders, such as the Edgeworth expansions, also do not yield adequate approximations as shown by Phillips (1977).

Recently, a new class of models which specifically deal with the presence of a root close to one has been studied. Phillips (1988) introduced the concept of a near–integrated random process where the autoregressive parameter is defined by :

(2) $$\alpha = \exp(c/T).$$

Here, the real–valued constant c is a measure of the deviation from the unit root case. The model (1) and (2) may also be described as having a root local to unity (see Cavanagh (1986)): as the sample size increases, the autoregressive parameter converges to unity. An expression for the limiting distribution of $T(\hat{\alpha} - \alpha)$ under (2) has been derived by Phillips (1988), Cavanagh (1986) and Chan and Wei (1987). Tabulations of this limiting distribution have been obtained by Nabeya and Tanaka (1987), Cavanagh (1986) and Perron (1988b) using different procedures. These studies also show the approximation to the finite sample distribution to be quite good in the case where $y_0 = 0$.

An important feature of this work is that the limiting distribution of $T(\hat{\alpha} - \alpha)$, under the near–integrated process (2), is invariant to the value of the initial observation, whether it be a random variable or a fixed constant. By contrast, results of Nankervis and Savin (1988) show that the exact distribution of $\hat{\alpha}$ in the stationary case is quite different from the fixed start–up case where the initial condition is $y_0 = 0$. There is accordingly a need to consider an alternative asymptotic framework which permits a distinction between the stationary and fixed start–up cases.

More recently, Perron (1988a) showed how the 'continuous record' asymptotic framework is useful in this regard. Here, the asymptotic analysis is performed not by letting the sample size increase to infinity keeping a fixed sampling interval, but rather by letting the sample size increase while keeping a fixed span of data, i.e. by letting the sampling interval converge to zero. We derived the limiting distribution of $T(\hat{\alpha} - \alpha)$ under general specifications for the initial condition y_0 and showed how it is directly influenced by different assumptions on its distribution.

We also established how this 'continuous record' asymptotic distribution can be related to the exact distribution of the continuous time estimator of the diffusion parameter in a suitably defined diffusion process. This result, in turn, permitted the derivation of a moment–generating function appropriate for the calculations of various distributional quantities under different assumptions about the initial condition.

While our earlier work analysed the case where the initial condition is any fixed value, the purpose of this paper is to consider in detail the stationary case as specified by (1). The paper is organized as follows: Section 2 outlines the framework adopted and describes the relevant distributional results. The main theoretical contribution is contained in Section 3 where we derive the appropriate joint moment–generating function necessary to calcute, by numerical integration, the distributional quantities of interest. Section 4 presents such computations concerning the c.d.f., p.d.f., moments and power functions. In each case, the adequacy of the asymptotic approximation to the finite sample distribution is assessed. Finally, Section 5 offers concluding comments and suggestions for future research.

2. SOME DISTRIBUTIONAL RESULTS

In this section, we briefly outline the particular stochastic processes of interest with their distributional properties. A more detailed treatment is presented in Perron (1988a). We begin with an observable process $\{y_t,\ t \geq 0\}$ defined on a probability space $(\Omega,\ F,\ \mu_y^\Theta)$. The measure μ_y^Θ is induced by the following diffusion process postulated for y_t:

(3) $$dy_t = \Theta y_t dt + \sigma\, dw_t \quad ; \quad y_0 \sim N(0, -\sigma^2/2\Theta); \quad t \geq 0.$$

Here, Θ and σ are unknown parameters with $-\infty < \Theta < 0$ and $\sigma > 0$. w_t is the standard unit Wiener process. According to (3), the stochastic process $\{y_t\}$ is second–order stationary (see, e.g., Arnold (1974)). Our concern is the estimation of Θ given a single sample path of observations $\{y_t,\ 0 \leq t \leq N\}$ where N is the span of the data. The analog of the least–squares procedure in continuous time yields the estimator:

(4) $$\hat{\Theta}_N(y) = \int_0^N y_t\, dy_t \,/\, \int_0^N y_t^2\, dt.$$

For simplicity of notation, we write $\hat{\Theta} \equiv \hat{\Theta}_N(y)$ and analyze the distributional properties of the standardized estimator $N(\hat{\Theta} - \Theta)$.

The discrete–time representation of the process y_t is easily shown to be given by :

(5) $$y_{th} = \exp(\Theta h)\, y_{(t-1)h} + u_{th}$$

where $u_{th} \sim N(0,\ \sigma^2\,(e^{2\Theta h} - 1)/2\Theta)$, $y_0 \sim N(0, -\sigma^2/2\Theta)$ and h is the sampling interval. This process is equivalent to (1) by specifying $\alpha \equiv \alpha_h \equiv \exp(\Theta h)$ and $\sigma^2 = \sigma_*^2\ 2\Theta/(e^{2\Theta h} - 1)$. In this discrete–time framework, the goal is to estimate the unknown quantity $\alpha_h = \exp(\Theta h)$ given a sequence of observations $\{y_{th};\ t = 0,1,2,\ldots,T\}$ where $T \equiv N/h$. The least–squares estimator is

(6) $$\hat{\alpha}_h = \Sigma_{t=1}^T\, y_{th}\, y_{(t-1)h} \,/\, \Sigma_{t=1}^T\, y_{(t-1)h}^2.$$

We focus on the asymptotic distribution of $T(\hat{\alpha}_h - \alpha_h)$ as $h \to 0$ given a fixed span

N. For simplicity, we consider a limiting sequence $\{h = h_1, h_2, \ldots, h_n\}$ such that $T = N/h$ is integer–valued and we require $h_n \to 0$ as $n \to \infty$.

The following result, proved in Perron (1988a), characterizes the exact distribution of $N(\hat{\theta} - \theta)$ given a sample path of observations $\{y_t\}_0^N$ and the asymptotic distribution of $T(\hat{\alpha}_h - \alpha_h)$ with the 'continuous record' asymptotic framework.

<u>Proposition 1</u>

Let $A(\gamma, c) \equiv \gamma\int_0^1 \exp(cr)dw_r + \int_0^1 J_c(r)\, dw_r$

and $B(\gamma, c) \equiv \gamma^2(\exp(2c) - 1)/2c + \gamma\int_0^1 \exp(cr)J_c(r)dr + \int_0^1 J_c(r)^2 dr$

where $J_c(r) \equiv \int_0^r \exp(c(r-s))dw_s$ and w_r is the Wiener process defined on C(0, 1).

i) Let $\{y_t, t \geq 0\}$ be a continuous time stochastic process generated by (3) and let $\hat{\theta}$ be the estimator of θ defined by (4). Then,

$$N(\hat{\theta} - \theta) \overset{d}{=} A(\gamma, c)/B(\gamma, c) \equiv Z(\gamma, c)$$

where $\gamma = y_0/\sigma N^{1/2}$ and $c = \theta N$; and where $\overset{d}{=}$ signifies equality in distribution.

ii) Let $\{y_t, t \geq 0\}$ be a continuous time stochastic process generated by (3) and let $\hat{\alpha}_h$ be defined by (6) with $\{y_{th}, t = 0, \ldots, T\}$ generated by (5). Then as $h \to 0$ with $T \to \infty$ and N fixed:

$$T(\hat{\alpha}_h - \alpha_h) \to Z(\gamma, c)$$

with, again, $\gamma = y_0/\sigma N^{1/2}$ and $c = \theta N$; and where '$\to$' denotes weak convergence in distribution.

<u>Remark</u>: The distributional result (ii) is valid under more general conditions upon the innovation sequence $\{u_{th}\}$. Indeed, the result remains valid if $\{u_{th}\}$ is a martingale difference sequence, thereby allowing possible heterogeneity and non–normal distributions.

Proposition 1 establishes that the small–h (or continuous records) asymptotic distribution of $T(\hat{\alpha}_h - \alpha_h)$ is the same as the exact distribution of the normalized continuous time estimator $N(\hat{\Theta} - \Theta)$. This result justifies studying the distributional properties of $N(\hat{\Theta} - \Theta)$ as an approximation to the exact distribution of the discrete time normalized least–squares estimator $T(\hat{\alpha} - \alpha)$.

More importantly, the stochastic representation of the variable $Z(\gamma, c)$ is explicitly affected by the specification of the initial conditon y_0. In particular, it is readily seen that the distribution of $Z(\gamma, c)$ is different when considering the fixed start–up case, $y_0 = 0$, and the stationary case, $y_0 \sim N(0, \sigma_*^2 (\alpha^2 - 1)^{-1})$. In this asymptotic context, one can at least hope to achieve a better approximation to the exact distribution in finite samples then in the usual $(T \to \infty)$ framework. Although it is possible to study the distributional properties of $Z(\gamma, c)$ under any assumption on the distribution of y_0, the natural case of interest is the one stipulated above where $\{y_t\}$ is a stationary process and we shall restrict our attention to this important case.

In order to use these distributional results as approximations to the finite sample behavior, we need a method to compute the distributional quantities of the random variable $Z(\gamma, c)$, or equivalently $N(\hat{\Theta} - \Theta)$. To this effect, the next section presents a moment–generating function which can be numerically integrated to obtain the various quantities of interest.

3. THE EXACT DISTRIBUTION OF $N(\hat{\Theta} - \Theta)$ IN THE STATIONARY CASE

Given that, in the stationary case, the exact distribution of $N(\hat{\Theta} - \Theta)$ depends only on the parameters $c = \Theta N$, it is useful to transform the original model. The transformation $t \in (0, N) \to r \in (0, 1)$ with $t \equiv Nr$ yields :

$$N(\hat{\Theta} - \Theta) = \sigma N^{-1} \int_0^N y_t \, dw_t / N^{-2} \int_0^N y_t^2 \, dt = \sigma N^{-1/2} \int_0^1 y_r dw_r / N^{-1} \int_0^1 y_r^2 \, dr$$

where

$$y_r = \exp(cr) y_0 + \sigma N^{1/2} \int_0^r \exp(c(r - s)) dw_s = \exp(cr) y_0 + \sigma N^{1/2} J_c(r) \; ; \; 0 \leq r \leq 1 \, ,$$

with initial condition $y_0 \sim N(0, -\sigma^2/2\Theta)$. Now, let $x_r = y_r / \sigma N^{1/2}$. Then

$$N(\hat{\Theta} - \Theta) = \int_0^1 x_r dw_r \Big/ \int_0^1 x_r^2 \, dr \tag{7}$$

where

$$x_r = \exp(cr)\gamma + J_c(r). \tag{8}$$

The expression (8) is the solution x_r of the stochastic differential equation:

$$dx_r = cx_r dr + dw_r \qquad 0 \leq r \leq 1 \tag{9}$$

$$x_0 = \gamma \sim N(0, -1/2c)$$

where $c = \Theta N$.

To study the exact distribution of $N(\hat{\Theta} - \Theta)$, we derive the joint moment–generating function of $\left[\int_0^1 x_r dw_r, \int_0^1 x_r^2 \, dr \right]$ given that x_r is a random variable in the probability space $(\Omega, F, \mu_y^\Theta \equiv \mu_x^c)$ generated according to the diffusion process (9). We denote this joint moment–generating function as :

$$M_c(v, u) = E\left[\exp(v\int_0^1 x_r dw_r + u\int_0^1 x_r^2\, dr)\right]$$

where the expectation is taken with respect to the measure $\mu_y^{\Theta} \equiv \mu_x^c$. The main result of this section is an expression for $M_c(v, u)$ which is presented in the following theorem.

Theorem 1

Let $\lambda = (c^2 + 2cv - 2u)^{1/2}$.

Then :

$$M_c(v, u) = \left[\frac{4c\lambda\exp(-(v + c + \lambda))}{4c\lambda\exp(-2\lambda) + (1 - \exp(-2\lambda))(v^2 - (c - \lambda)^2)}\right]^{1/2}.$$

Proof: The first part of the proof is the same as in the case of a fixed y_0 considered in Perron (1988a) and is omitted. The relevant result is that

(10) $$M_c(v, u) = E\left[\exp\{(a/2)((x_1^\lambda)^2 - (x_0^\lambda)^2)\}\right]$$

where $a = v + c - \lambda$ and x_r^λ is a stochastic process generated by the following stochastic differential equation :

(11) $$dx_r^\lambda = \lambda x_r^\lambda dr + dw_r \qquad 0 \leq r \leq 1$$

with $\quad x_0^\lambda = \gamma \sim N(0, -1/2c)$ and where $\lambda = (c^2 + 2cv - 2u)^{1/2}$.

Now the stochastic process x_r^λ given by (11) has the following unique (μ_x^λ measure) solution :

(12) $$\begin{aligned} x_r^\lambda &= \exp(\lambda r)\gamma + \int_0^r \exp(\lambda(r - s))dw_s \\ &= \exp(\lambda r)\gamma + J_\lambda(r) \end{aligned}$$

where $J_\lambda(r) \sim N(0, (e^{2\lambda r} - 1)/2\lambda)$. Hence, x_1^λ is a normal random variable with mean 0 and variance given by

$$\mathrm{Var}(x_1^\lambda) = \exp(2\lambda)\ \mathrm{Var}(\gamma) + \mathrm{Var}(J_\lambda(1))\ ,$$

since by assumption $\gamma = x_0^\lambda$ is independent of $w_s(0 \le s \le 1)$. Simple calculations yield

$$\mathrm{Var}(x_1^\lambda) = \frac{\exp(2\lambda)}{2}\left[\frac{1}{\lambda} - \frac{1}{c}\right] - \frac{1}{2\lambda} \equiv d$$

and

$$E[x_0^\lambda\ x_1^\lambda] = \exp(\lambda)(-1/2c).$$

Let $x' = (x_1^\lambda\ x_0^\lambda)$. The components of x are jointly normal and $(x_1^\lambda)^2 - (x_0^\lambda)^2$ is a quadratic form in normal random variables. Hence, we can write

$$M_c\ (v, u) = E[\exp\{(a/2)(x'Ax)\}]$$

where $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $x \sim N(0, \Omega)$;

and $\Omega = (-1/2c)\begin{bmatrix} h & \exp\{\lambda\} \\ \exp(\lambda) & 1 \end{bmatrix} \equiv (-1/2c)\Psi$

with $h = (-2c)d = \exp(2\lambda)[1 - c/\lambda] + c/\lambda$.

Let $g = \Omega^{-1/2}x \sim N(0, I_2)$,then $x'Ax = (-1/2c)g'\Psi^{1/2}A\ \Psi^{1/2}g$ is a quadratic form in N(0, 1) independent variates. $\Psi^{1/2}A\ \Psi^{1/2}$ is symmetric and can therefore be diagonalized as

$$P'\Psi^{1/2}A\ \Psi^{1/2}\ P = \Lambda$$

where PP'= I and Λ is a diagonal matrix with the characteristic roots of $\Psi^{1/2}A\ \Psi^{1/2}$ as elements. Now let $Z = P'g$ then

$$\begin{aligned} x'Ax &= (-1/2c)Z'P'\Psi^{1/2}\Lambda\ \Psi^{1/2}PZ \\ &= (-1/2c)Z'\Lambda\ Z \\ &= (-1/2c)(\lambda_1 Z_1^2 + \lambda_2 Z_2^2) \end{aligned}$$

where $Z' = (Z_1, Z_2) \sim N(0,I_2)$.

The moment–generating function can then be expressed as :

$$M_c(v, u) = E\,[\exp\{(-a/4c)(\lambda_1 Z_1^2 + \lambda_2 Z_2^2)\}]$$
$$= E\,[\exp\{(-a/4c)\lambda_1 Z_1^2\}]E[\exp\{(-a/4c)\lambda_2 Z_2^2\}]$$

(13) $$= \left[\frac{1}{1 + (a/2c)\lambda_1}\right]^{1/2} \left[\frac{1}{1 + (a/2c)\lambda_2}\right]^{1/2}$$

since Z_1^2 and Z_2^2 are independent central chi–square variates.

It remains to find the explicit form of the eigenvalues λ_1 and λ_2. These are analysed in the appendix and are given by :

(14) $$\lambda_1, \lambda_2 = (h-1)/2 \pm [(h-1)/2 + h - \exp(2\lambda)]^{1/2}$$

where $$h = \exp(2\lambda)(1 - c/\lambda) + c/\lambda\,.$$

Simple but tedious manipulation of (13) and (14) yields the desired form of the moment–generating function . □

The moment–generating function derived in Theorem 1 allows us to compute the cumulative distribution function as well as other distributional quantities using numerical integration. Various computational procedures are discussed in Section 4 and the numerical results presented in Section 5.

4. NUMERICAL INTEGRATION OF THE MOMENT-GENERATING FUNCTION

The moment–generating function derived in Section 3 can be used directly to derive the moments of $N(\hat{\theta} - \theta)$. Using Mehta and Swamy's (1978) results, we have:

$$(15) \qquad E[N(\hat{\theta} - \theta)]^r = \frac{1}{\Gamma(r)} \int_0^\infty u^{r-1} \left[\frac{\partial^r M_c(v, -u)}{\partial v^r}\right]_{v=0} du$$

where $M_c(v, u) = E[\exp(v\int_0^1 x_r dw_r + u\int_0^1 x_r^2\, dr)]$ is the joint moment–generating function of $\left[\int_0^1 x_r dw_r, \int_0^1 x_r^2\, dr\right]$.

Section 5 presents results for the first two moments of $N(\hat{\theta} - \theta)$ evaluated at various values of c. In practice, the integral in (15) is evaluated in a range $(\epsilon, \bar{U})$ where $\bar{U}$ is set such that the integrand is less than ϵ when evaluated at $\bar{U}$. ϵ was fixed at 1.0E – 07. The subroutine DCADRE of the International Mathematical and Statistical Library (IMSL) was used to evaluate the integral (the error criterion for this routine integration was also set at 1.0E – 07).

To compute values for the cumulative and probability density functions we must consider the joint characteristic function of $\left[\int_0^1 x_r dw_r, \int_0^1 x_r^2\, dr\right]$, denoted by $cf_c(v, u)$. Then :

$$cf_c(v, u) = M_c(iv, iu) = E[\exp(iv\int_0^1 x_r dw_r + iu \int_0^1 x_r^2\, dr)].$$

The distribution function of $N(\hat{\theta} - \theta)$ can be obtained as follows. Let $F_c(z) = P[N(\hat{\theta} - \theta) \leq z]$ and recall that $N(\hat{\theta} - \theta) = \int_0^1 x_r dw_r \,/ \int_0^1 x_r^2\, dr$. Then, by Theorem 1 of Gurland (1948), we have :

$$F_c(z) = \frac{1}{2} - \frac{1}{2\pi i} \lim_{\substack{\epsilon_1 \to 0 \\ \epsilon_2 \to \infty}} \int_{\epsilon_1 < |v| < \epsilon_2} \left[\frac{cf_c(v, -vz)}{v}\right] dv$$

$$(16) \qquad = \frac{1}{2} - \frac{1}{2\pi} \int_0^\infty \text{AIMAG}\left[\frac{cf_c(v, -vz)}{v}\right] dv$$

where AIMAG $(\cdot)$ denotes the imaginary part of the complex number. Further, the density function is obtained as follows:

$$(17) \qquad f_c(z) = \frac{d}{dz} F_c(z) = \frac{1}{2\pi i} \lim_{\substack{\epsilon_1 \to 0 \\ \epsilon_2 \to \infty}} \int_{\epsilon_1 < |v| < \epsilon_2} \left.\frac{\partial cf_c(v, u)}{\partial u}\right|_{u=-vz} dv.$$

The expressions (16) and (17) can be numerically integrated to obtain values for the cumulative distribution function and the probability density function. When calculating these values, we again evaluate the integrals in a range $(\epsilon, \bar{V})$ where $\bar{V}$ is an upper bound set such that the integrand evaluated at $\bar{V}$ is less than ϵ. We again use the subroutine DCADRE of IMSL and all error criteria are set at 1.0E − 07. The integration of the c.d.f. and p.d.f. are, however, quite different in practice than the integration involving the moments. Here, the integrand involves the square root of a complex valued function. The use of the principal value of the square root may not ensure the continuity of the integrand. We must, therefore, integrate over the Reimann surface consisting of two planes. The method used is described in more detail in the Appendix to Perron (1988b).

The cumulative distribution function (16) can also be used to analyze the power function for tests of the null hypothesis

$$(18) \qquad H_0 : \theta = \theta_0$$

using the statistic $N(\hat{\theta} - \theta_0)$ against various alternatives θ $(-\infty < \theta < 0, \theta \neq \theta_0)$. Denote by z^* the value such that $P_{\theta_0}[N(\hat{\theta} - \theta_0) \leq z^*] = \beta$. Then the power function of a one-sided test with size β for testing $\theta = \theta_0$ against $\theta < \theta_0$ is given by:

$$(19) \qquad P_\theta[N(\hat{\theta} - \theta_0) < z^*] = P_\theta[N(\hat{\theta} - \theta) < z^* + (c_0 - c)]$$

with $c = N\theta$ and $c_0 = N\theta_0$. Expression (16) can be used to evaluate the power function (19) for various values of c, c_0 and significance level β. A similar analysis carries over for testing $\theta = \theta_0$ against $\theta > \theta_0$ or for two-tailed tests.

5. DISCUSSION OF THE NUMERICAL RESULTS

Figure 1 presents the graph of the cumulative distribution function of $N(\hat{\theta} - \theta)$ for c = –5.0, –1.0, –0.5, –0.1 and –0.01. Each curve is constructed by evaluating the integral (16) at 160 equidistant points. Figure 2 presents the corresponding probability density functions and is constructed in a similar way (we have, however, omitted the case c = –0.01 from the graph due to a sharply different scale from the other cases). Several features are worth noting. First, there are marked differences in the distribution functions for different values of c. The probability density function is very flat for large (negative) values of c. It gets more concentrated around zero as c gets small (in absolute value). The p.d.f. is also sharply non–symmetric when c is large, however, the skewness decreases as c gets smaller. From this evidence and the results of other experiments not reported, it appears that the distribution function becomes degenerate at zero as c approaches zero.

Table 1 presents selected percentage points of the distribution function of $N(\hat{\theta} - \theta)$ for c = –10.0, –5.0, –1.0, –0.1 and –0.01. These are presented in the infinity rows. Table 2 presents the mean, median and standard deviation of $N(\hat{\theta} - \theta)$ for the same values of c. These tables also presents finite sample results to assess the adequacy of the asymptotic approximation. The finite sample results were obtained by simulation methods. We used 20,000 replications of model (1) with $\alpha = \exp(c/T)$ to obtain our estimates. N(0, 1) random deviates were used to construct the series. These were obtained from the subroutine GASDEV in Press et al. (1986).

Overall, the results show our approximation to be adequate. The approximation is better when 1) c is closer to zero, 2) T increases and 3) for the right–hand tail of the distribution. When c is as large as –10.0 and T is small (e.g., 25), the approximation is not very good. It becomes noticeably better as T increases. To get a better grasp of the relationship between the finite sample distribution and its asymptotic counterpart, Figure 3 graphs the c.d.f. for c = –10.0 and T = 25, 50 , 100 and infinity. One can see from this graph that the approximation improves rapidly when T increases.

When c= –0.01, the approximation is excellent for even very small sample sizes, and even in the extreme tail of the distribution. This feature is to be expected since our asymptotic framework is one that is local to $\alpha = 1$ and $\alpha_h \to 1$ as h, the sampling interval, converges to zero.

To see more clearly how our approximation fares in terms of the exact distribution of $\hat{\alpha}$, for various α, Tables 3 and 4 present some additional evidence with respect to the mean and standard deviations of $\hat{\alpha}$ for various values of α and sample sizes T. These tables present the exact mean and standard deviation of $\hat{\alpha}$ as calculated by Nankervis and Savin (1988a, b) using Sawa's (1978) method. The selected values of α and T are $\alpha = 0.99$, 0.95, 0.90, 0.80 and 0.60 and T = 10, 20, 25, 30, 50 and 100. Our approximation for the expected value of $\hat{\alpha}$ is also compared with White's (1961) approximation as reported in Nankervis and Savin (1988b). Our approximation for the standard deviation is also compared with those of White (1961) and Bartlett (1946), again as reported in Nankervis and Savin (1988b). It should be noted that our approximations to the mean and standard deviations of $\hat{\alpha}$ are obtained using (15) with r = 1, 2 and specifying $c = T \ln(\alpha)$.

The results of Table 3 show our approximation to the expected value of $\hat{\alpha}$ to be better than White's approximation for $\alpha = 0.99$ and $\alpha = 0.95$ (for all values of T). The approximation is, in general, excellent even for small values of T. For values of α at 0.90, 0.80 and 0.60, White's approximation is better. Our approximation is imprecise for small sample sizes but becomes quite precise as the sample size increases. In general we find that the smaller the value of α , the slower is the approach between the asymptotic approximation and the exact distribution as T increases.

Much of the same comments apply to the results concerning the standard deviation of $\hat{\alpha}$ presented in Table 4. Our approximation is better than White's approximation for $\alpha = 0.99$ and $\alpha = 0.95$; almost as good for $\alpha = 0.90$ and less good for $\alpha = 0.80$ and 0.60. An interesting feature is that our approximation overstates the exact result while White's and Bartlett's understate it.

To obtain a clearer view of the behavior of the first two moments of $N(\hat{\theta} - \theta)$, Figures 4 and 5 present a graph of these two quantities for values of c in the −20 to 0 range. To see the effect induced by the stationarity assumption we also compare the results with those obtained in the fixed start-up case where $y_0 = 0$ (see Perron (1988a)).

As can be seen from Figure 4, the normalized bias of $\hat{\theta}$ (the mean of $N(\hat{\theta} - \theta)$) decreases rapidly towards zero as c approaches zero. This is unlike the fixed start-up case where the decrease in the bias is small as c approaches zero (the decrease is quite rapid, however, as c increases on the positive side, see Perron (1988a)). Although not shown in the graph, the bias in the stationary case eventually approaches the value of the bias in the fixed start–up case for very large (negative) values of c.

Much of the same comments apply to the normalized mean–squared error of $\hat{\theta}$ (the second moment of $N(\hat{\theta} - \theta)$) presented in Figure 5. In the fixed start–up case the curve is basically linear in c. The mean–squared error in the stationary case follows closely that in the fixed start–up case for large (negative) values of c (again for very large c, the two curves eventually join). However, in the stationary case, the mean–squared error function approaches zero as c approaches zero.

The results presented in Figures 4 and 5 help to explain the result of Nankervis and Savin (1988) concerning the finite sample properties of $\hat{\alpha}$. First, if α is much lower than one, the mean and standard deviation of $\hat{\alpha}$ are basically the same in the stationary and fixed start–up cases. However, for value of α closer to one, the two cases yield quite different behavior, the stationary case showing less bias and variability.

The final computational exercise using the results of Section 4 concerns the power function of a test of the null hypothesis (18) obtained using the function (19). Figure 5 presents the power functions of a 5 % one–tailed test using the statistic $N(\hat{\theta} - \theta_0)$ for $\theta_0 = -5.0$, -2.0, -1.0, -0.5 and -0.1. Several features from this graph are worth mentioning. First, the power function against left–sided alternatives ("more stationary") rises faster as θ_0 is closer to 0. For instance, the power of a test that $c_0 = -0.5$ against an alternative that $c = -5.0$ is approximately 0.50 which is close to the power of a test that $c_0 = -5.0$ against an alternative that $c = -14.0$.

The behavior of the power function for right–sided alternatives is, however, quite different. For $\theta_0 = -5.0$ and -2.0, the power function is monotonic and increases faster than for left–sided alternatives. Interestingly, the power functions in the cases $\theta_0 = -1.0$ and -0.5 exhibit non–monotonic behavior. For the null hypotheses $\theta_0 = -1.0$ and $\theta_0 = -0.5$, the power function initially increases but eventually starts decreasing as the alternative value approaches zero. The tests even become biased for values of c close to zero. In the case of a null hypothesis that $\theta_0 = -0.1$, the power function is monotonic–decreasing as c approaches zero. Accordingly the test is biased for all values of c between –0.1 and 0.0.

This non–monotonic behavior of the power function can be understood by looking at the graph of the cumulative distribution function of $N(\hat{\Theta} - \Theta)$ presented in Figure 1. What emerges from this graph is that the distribution function of $N(\hat{\Theta} - \Theta)$ shows little displacement as c changes for low values of c. It simply shows a greater concentration around zero with little horizontal shift. Now, from (19), the power function can be decomposed as:

$$(20) \qquad P_{\Theta}[N(\hat{\Theta} - \Theta) < z^{*} + (c_0 - c)]$$

$$= P_{\Theta}[N(\hat{\Theta} - \Theta) < z^{*}] - P_{\Theta}[z^{*} + (c_0 - c) < N(\hat{\Theta} - \Theta) < z^{*}].$$

The first term in (20) will be so much above 0.05 (the size of the test) depending on the displacement of the distribution of $N(\hat{\Theta} - \Theta)$ as c (or Θ) changes. The second term in (20) tends to reduce the overall power and the amount of the reduction is greater as c approaches zero. The power function will exhibit non–monoticity when the second term eventually outweighs the first. This happens for small values of c essentially because in this case changes in c cause only a small displacement in the distribution function. Hence the first term in (20) is little above the size of the test. On the other hand, as c approaches zero there is even more concentration in the distribution causing the second term in (20) to be larger as c approaches zero. In the case of the null hypothesis $\Theta_0 = -0.1$, the displacement in the distribution as c gets closer to zero is so small and the increased concentration so large that the second term in (20) outweighs the first for all values of c between –0.1 and 0.0, and results in a monotonic decreasing power function.

6. CONCLUDING COMMENTS

This paper has presented an asymptotic approximation to the least–squares estimator of the parameter in a stationary first–order autoregressive model. Our approach is different from the previous literature in that we consider a framework using the continuous record asymptotic analysis ($h \to 0$) instead of the usual ($T \to \infty$) asymptotic theory. The main advantage of our method is that it allows explicit consideration of the effects of different assumptions concerning the initial condition.

Our approach yields interesting results concerning the distributional properties of the continuous time estimator of the diffusion parameter θ. These are interesting theoretical results per se but their usefulness lies in their adequacy to provide satisfactory approximations to finite sample estimators in discrete–time models. To this effect, we have presented extensive evidence that the approximation is indeed adquate for a certain range of parameter values near unity. This is particularly useful because it is precisely this range of parameter values where the usual asymptotic theory yields inadequate approximations.

While our framework is particularly simple, the method can be extended to more complex models where exact results are too difficult or expensive to obtain. These include models with an intercept and/or time trend. Such analyses are left for future research.

APPENDIX

Derivation of the eigenvalues λ_1, λ_2

Let $\Psi^{1/2} \equiv \begin{bmatrix} p & q \\ q & r \end{bmatrix}$ and $X = \Psi^{1/2} A \, \Psi^{1/2}$, then :

$$X = \begin{bmatrix} p^2 - q^2 & pq - qr \\ pq - qr & q^2 - r^2 \end{bmatrix}$$

and the eigenvalues of X are the roots λ_1 , λ_2 of $|X - \lambda I_2| = 0$ or equivalently the roots λ_1 , λ_2 of

$$\lambda^2 + \lambda\,(r^2 - p^2) - (pr - q^2)^2 = 0.$$

Hence,

$$\lambda_1, \lambda_2 = (p^2 - r^2)/2 \pm [(p^2 - r^2)^2/4 + (pr - q^2)^2\,]^{1/2}.$$

To solve for λ_1 and λ_2, we need an expression for the quantities $(p^2 - r^2)$ and $(pr - q^2)^2$. Using the identity $\Psi^{1/2}\,\Psi^{1/2} = \Psi$, we have the following system of equations :

(A.1) $$p^2 + q^2 = h$$

(A.2) $$pq + qr = \exp(\lambda)$$

(A.3) $$q^2 + r^2 = 1.$$

Using (A.1) and (A.3) we have : $p^2 - r^2 = h - 1$. Now,

(A.4) $$(pr - q^2)^2 = p^2r^2 + q^4 - 2prq^2.$$

From (A.1) and (A.3) : $(p^2 + q^2)\,(q^2 + r^2) = h$ or

(A.5) $$p^2r^2 + q^4 = h - (p^2q^2 + q^2r^2).$$

Now, from (A.2) :

(A.6) $$(p^2q^2 + q^2r^2) = \exp\,(2\lambda) - 2prq^2.$$

Substituting (A.6) in (A.5) yields :

$$p^2r^2 + q^4 = h - \exp(2\lambda) + 2\,prq^2 \tag{A.7}$$

and substituting (A.7) in (A.4) yields :

$$(pr - q^2)^2 = h - \exp(2\lambda) .$$

Hence :

$$\lambda_1 , \lambda_2 = (h-1)/2 \pm [\,(h-1)^2/4 + h - \exp(2\lambda)]^{1/2} .$$

BIBLIOGRAPHY

Arnold, L.,1974, *Stochastic Differential Equations: Theory and Applications*, John Wiley and Sons, New York.

Bartlett, M. S.,1946, "On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series", *Journal of the Royal Statistical Society* 8, 27–41.

Basmann, R.L., D.H. Richardson and R.J. Rohr,1974, "Finite Sample Distributions Associated with Stochastic Difference Equations – Some Experimental Evidence", *Econometrica* 42, 825–839.

Cavanagh, C.,1986, "Roots Local to Unity", mimeo, Harvard University.

Chan, N.H. and C.Z. Wei,1986, "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *Annal of Statistics* 15,1050–1063.

Evans, G.B.A. and N.E. Savin,1981, "Testing for Unit Roots: 1", *Econometrica* 49, 753–779.

Evans, G.B.A. and N.E. Savin,1984, "Testing for Unit Roots: 2", *Econometrica* 52, 1241–1269.

Gurland,J.,1948, "Inversion Formulae for the Distribution of Ratios", *Annals of Mathematical Statistics* 19, 228–237.

Mann, H.B. and A. Wald,1943, "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica* 11, 173–220.

Mehta, J.S. and P.A.V.B. Swamy,1978, "The Existence of Moments of Some Simple Bayes Estimators of Coefficients in a Simultaneous Equation Model", *Journal of Econometrics* 7, 1–14.

Nabeya, S. and K. Tanaka,1987, "Asymptotic Distribution of the Estimator for the Nonstable First–order Autoregressive Model", mimeo, Hitotsubashi University.

Nankervis, J. and N.E. Savin,1988a, "The Exact Moments of the Least–Squares Estimator for the Autoregressive Model: Corrections and Extensions", *Journal of Econometrics* 37, 381–388.

Nankervis, J. and N.E. Savin,1988b, "The Exact Moments of the Least–Squares Estimator for the Autoregressive Model : Corrections and Extensions", Working Paper Series No. 86–23 (Department of Economics, University of Iowa, Iowa City,Ia).

Perron, P., 1988a, "A Continuous–time Approximation to the Unstable First–order Autoregressive Model: The Case without an Intercept", mimeo, Princeton University.

Perron, P., 1988b, "The Calculation of the Limiting Distribution of the Least–Squares Estimator in a Near–Integrated Model", *Econometric Theory* (forthcoming).

Phillips, P.C.B.,1977, "Approximations to Some Finite Sample Distributions Associated with a First–order Stochastic Difference Equation", *Econometrica* 45, 463–485.

Phillips, P.C.B.,1988, "Regression Theory for Near–Integrated Time Series", *Econometrica* 56, 1021–1044.

Press, W.H., B.P. Flanery, S.A. Teukolsky and W.T. Vetterling,1986, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge.

Rubin, H.,1950, "Consistency of Maximum Likelihood Estimator in the Explosive Case", in T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Wiley and Sons, New York.

Sawa, T.,1978, "The Exact Moments of the Least–Squares Estimator for the Autoregressive Model", *Journal of Econometrics* 8, 159–172.

White, J.S.,1961, "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient", *Biometrika* 48, 85–95.

Table 1

Percentage Points of the Distribution of $T(\hat{\alpha} - \alpha)$

		1 %	2.5 %	5 %	10 %	90 %	95 %	97.5 %	99 %
c = –10.0	T = 25	–12.531	–10.144	–8.416	–6.523	3.387	4.200	4.876	5.513
	50	–14.871	–12.117	–9.861	–7.439	3.551	4.443	5.087	5.807
	100	–16.687	–13.382	–10.935	–8.241	3.630	4.567	5.201	5.769
	∞	–18.478	–14.728	–11.834	–8.846	3.793	4.729	5.436	6.154
c = –5.0	T = 25	–11.909	–9.675	–7.959	–6.004	2.272	2.835	3.273	3.781
	50	–13.288	–10.863	–8.733	–6.530	2.318	2.886	3.327	3.829
	100	–14.931	–11.676	–9.302	–6.926	2.324	2.901	3.337	3.788
	∞	–15.735	–12.365	–9.808	–7.232	2.380	2.971	3.417	3.905
c = –1.0	T = 25	–10.023	–7.810	–6.085	–4.344	0.988	1.356	1.692	2.119
	50	–11.103	–8.386	–6.484	–4.655	0.981	1.334	1.653	2.059
	100	–11.319	–8.830	–6.702	–4.742	0.970	1.308	1.623	2.028
	∞	–11.848	–8.942	–6.831	–4.818	0.971	1.322	1.639	2.042
c = –0.1	T = 25	–7.196	–5.045	–3.564	–2.300	0.500	0.744	1.006	1.372
	50	–7.365	–5.213	–3.688	–2.302	0.510	0.739	0.989	1.373
	100	–8.041	–5.652	–3.834	–2.368	0.497	0.723	0.958	1.281
	∞	–8.168	–5.634	–3.906	–2.405	0.495	0.727	0.971	1.312
c = –0.01	T = 25	–4.702	–2.856	–1.662	–0.762	0.254	0.409	0.602	0.894
	50	–4.714	–2.801	–1.618	–0.762	0.251	0.404	0.589	0.919
	100	–5.155	–2.798	–1.616	–0.751	0.249	0.396	0.580	0.846
	∞	–5.051	–2.942	–1.672	–0.776	0.249	0.402	0.588	0.878

Table 2

Summary Statistics of the Distribution of $T(\hat{\alpha} - \alpha)$

		Mean	Median	Standard Deviation
c = –10	T = 25	–1.187	–0.647	3.927
	50	–1.498	–0.798	4.455
	100	–1.722	–0.864	4.842
	∞	–1.904	–0.939	5.246
c = –5.0	T = 25	–1.382	–0.724	3.381
	50	–1.579	–0.807	3.688
	100	–1.720	–0.870	3.923
	∞	–1.821	–0.888	4.144
c = –1.0	T = 25	–1.255	–0.603	2.454
	50	–1.355	–0.640	2.603
	100	–1.410	–0.648	2.697
	∞	–1.431	–0.651	2.781
c = –0.1	T = 25	–0.626	–0.170	1.592
	50	–0.638	–0.182	1.620
	100	–0.678	–0.181	1.730
	∞	–0.688	–0.179	1.766
c = –0.01	T = 25	–0.236	–0.027	0.972
	50	–0.233	–0.029	0.965
	100	–0.246	–0.028	1.043
	∞	–0.253	–0.028	1.062