THE ADEQUACY OF LIMITING DISTRIBUTIONS IN THE AR(1) MODEL WITH DEPENDENT ERRORS

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

Econometric Research Program
Research Memorandum No. 349

November 1989
Revised June 1990

Some material presented in Section 3 was initially investigated as part of a joint project with Peter C. B. Phillips in the Spring of 1986; I wish to thank him for his help and comments during this initial stage. This research was supported by grants from the Social Sciences and Humanities Research Council of Canada, the Natural and Engineering Council of Canada and the Fonds pour la Formation de Chercheurs et l'aide à la Recherche du Québec.

Econometric Research Program
Princeton University
204 Fisher Hall
Princeton, NJ 08544-1021, USA

ABSTRACT

We consider the normalized least squares estimator of the parameter in a nearly integrated first–order autoregressive model with dependent errors. In a first step we consider its asymptotic distribution as well as asymptotic expansion up to order $O(T^{-1})$. We derive a limiting moment generating function which enables us to calculate various distributional quantities by numerical integration. We provide an extensive simulation study to assess the adequacy of the asymptotic distribution when the errors are correlated. We focus our attention on two leading cases : a) MA(1) errors and , b) AR(1) errors. The asymptotic approximations are shown to be inadequate as the MA root gets close to minus one and as the AR root gets close to either minus one or one. We discuss the cause of this poor performance by investigating the adequacy of the underlying functional central limit theorem. For each of the cases where the asymptotic approximation is inadequate we provide an alternative asymptotic framework . This is achieved by considering the limiting behavior of the least squares estimator in this nearly integrated model with errors that are "local" to the boundary of the permissible region, i.e. when the root of the error process approaches one (or minus one) at a suitable rate. Several interesting by–products of our analyses are outlined.

Key Words : Near–integrated model, functional weak convergence, simulation experiment, unit root process, nearly stationary model, nearly twice integrated model, nearly seasonally integrated model.

Address : Pierre Perron
Department of Economics
Princeton University
Princeton, NJ, 08544

1. INTRODUCTION

In an attempt to cover more general time series structures, it has become popular in econometric methodology to consider models which permits that both the regressors and the errors have substantial heterogeneity and dependence over time. On a theoretical level, this advance has become possible due to a new class of central limit theorems (or functional central limit theorems) which provides asymptotic results allowing both substantial heterogeneity and dependence. An integrated treatment can be found in White (1984). This approach has made possible the analysis of a wide class of models with substantial relaxation of the standard conditions. Examples include time series models with unit roots (e.g.,Phillips (1987a)), testing for structural change in a general nonlinear framework (e.g., Andrews and Fair (1988)), and cointegration (e.g., Phillips and Ouliaris (1987)). However, very little is known about the adequacy of the limiting distributions as an approximation to the finite sample distribution in such a general framework. This paper is a first step in a careful examination of this issue. We consider the leading case of a dynamic first–order autoregressive model when the errors are allowed to be dependent and provide a detailed analysis of the behavior of the associated ordinary least squares estimator. To be more precise, we consider the following first–order stochastic difference equation :

$$y_t = \alpha y_{t-1} + u_t \qquad (t = 1,\dots,T) \tag{1.1}$$

where y_0 is a fixed constant. The least–squares estimator of α based on a sequence of observations $\{y_t\}_0^T$ is given by :

$$\hat{\alpha} = \Sigma_{t=1}^{T} y_t y_{t-1} (\Sigma_{t=1}^{T} y_{t-1}^2)^{-1} . \tag{1.2}$$

The distribution of $\hat{\alpha}$ has been extensively studied, especially under the case where the errors $\{u_t\}$ are uncorrelated. Mann and Wald (1943) and Rubin (1950) showed that $T^{1/2}(\hat{\alpha} - \alpha)(1 - \alpha^2)^{-1/2}$ has a limiting N(0,1) distribution when $|\alpha| < 1$. White (1958) showed that when $|\alpha| > 1$, the limiting distribution of $|\alpha|^T(\alpha^2 - 1)^{-1} (\hat{\alpha} - \alpha)$ is Cauchy provided that $y_0 = 0$. White also considered the cased $|\alpha| = 1$ and showed that the limiting distribution of $T(\hat{\alpha} - 1)$ can be expressed in terms of the ratio of two functionals of a Wiener process (see also Phillips (1987a)) .The case of the unit root, $\alpha = 1$, has attracted a great deal of attention. The asymptotic distribution of $T(\hat{\alpha} - 1)$ has been tabulated by

Dickey (1976) via simulation methods (see also Fuller (1976)) and by Evans and Savin (1981a) using numerical integration. Evans and Savin (1981b) showed how the standard limiting distributions fail to provide an adequate approximation to the exact distribution when α is close to but not equal to one.

Recently, a new class of models which specifically deal with the presence of a root close to, but not necessarily equal to one, has been studied. Consider a near–integrated process where the autoregressive parameter is defined by :

$$\alpha = \exp(c/T). \tag{1.3}$$

Here, the constant c is a measure of the deviation from the unit root case. The model (1.1) and (1.3) may also be described as having a root local to unity : as the sample size increases, the autoregressive parameter converges to unity. When $c < 0$, the process $\{y_t\}$ is said to be (locally) stationary and when $c > 0$, it is said to be (locally) explosive. An expression for the limiting distribution of $T(\hat{\alpha} - \alpha)$ under (1.3) has been derived by Phillips (1987b), Cavanagh (1986) and Chan and Wei (1987). This framework has been quite useful in studying various problems such as the power of tests of a unit root under local alternatives (Phillips (1987b), Phillips and Perron (1988) and Perron (1990a)), the derivation of confidence intervals when α is near unity (Cavanagh (1986)) and the power of tests of a unit root with a continuum of observations (Perron (1989b)).

In the near–integrated context, with errors that are weakly dependent, Phillips (1987b) showed that (under some conditions to be made precise later) :

$$T(\hat{\alpha} - \alpha) \Rightarrow \left\{ \int_0^1 J_c(r)dW(r) + \lambda \right\} \left\{ \int_0^1 J_c(r)^2 dr \right\}^{-1}, \tag{1.4}$$

where $\lambda = (\sigma^2 - \sigma_u^2)/(2\sigma^2)$, $\sigma^2 = \lim_{T\to\infty} E(T^{-1}S_T^2)$, $S_t = \Sigma_{j=1}^t u_j$, $\sigma_u^2 = \lim_{T\to\infty} T^{-1}E(\Sigma_{t=1}^T u_t^2)$, $J_c(r) = \int_0^r \exp((r-s)c)dW(r)$; and W(r) is the unit Wiener process (or standard Brownian motion) on C[0,1], the space of real–valued continuous functions on the [0,1] interval. This type of asymptotic distribution provides a useful framework to analyze models with dependent errors.

Tabulations of the limiting distribution (1.4) with $\lambda = 0$ have been obtained by Nabeya and Tanaka (1987), Cavanagh (1986) and Perron (1989a) using different procedures. These studies also provide measures of the adequacy of this limiting distribution as an approximation to the finite sample distribution of $\hat{\alpha}$ when α is in the vicinity of 1. They show the approximation to be quite good in the case where $y_0 = 0$. Perron (1988a,b) also considers a continuous time approximation which performs well even in the case where the initial condition is non–zero. These asymptotic frameworks provide a substantial improvement over the traditional asymptotic distribution theory, when α is in the vicinity of one, essentially because the asymptotic distributions obtained are continuous with respect to the autoregressive parameter α. However, most of the available evidence about the adequacy of the approximation pertains to the case where $\lambda = 0$, i.e when there is no correlation in the residuals.

The purpose of this paper is to investigate the adequacy of such an asymptotic framework in approximating the exact distribution in finite samples when the errors are dependent. Section 2 presents results concerning the limiting distribution of $T(\hat{\alpha} - \alpha)$ and extends Phillips' (1987c) $O(T^{-1})$ expansion to the near–integrated setting. The results of Perron (1988a) are used to derive the limiting joint moment–generating function of $\{T^{-1}\Sigma_1^T y_{t-1}u_t,\ T^{-2}\Sigma_1^T y_{t-1}^2\}$ in both the $O(1)$ and $O(T^{-1})$ frameworks. This limiting moment–generating function permits the calculation of the cumulative distribution function and the moments of the asymptotic distribution. In Section 3, we present an extensive simulation experiment to compare the asymptotic results with their finite sample counterparts. We concentrate our analysis on two leading cases, namely :

(1.5) MA(1) errors : $u_t = e_t + \theta e_{t-1}$,

(1.6) AR(1) errors : $u_t = \rho u_{t-1} + e_t$,

where $\{e_t\}$ is a sequence of i.i.d. $N(0,\sigma_e^2)$ random variables. The results show that the asymptotic distribution is a very poor guide to the finite sample distribution , even for quite large sample sizes, when either θ (in the MA case) or ρ (in the AR case) is close to -1. When ρ is close to $+1$, the approximation is not as bad but the approach to the limiting values is quite slow.

Section 4 discusses the behavior of the underlying functional central limit theorem used to prove results such as (1.4). We show how the approach to the limiting values can be quite slow in an important part of the parameter space. Sections 5 through 7 are devoted to providing an alternative asymptotic framework in each of the cases where the usual asymptotic distribution fails to be a sensible guide to the finite sample distribution. In Section 5 we consider the limiting behavior of the normalized least–squares estimator allowing the MA parameter θ to approach -1 at a suitable rate. This provides an asymptotic framework which we label as "nearly white noise – nearly integrated process". Here again we derive a limiting joint moment generating function which allows calculation of distributional quantities. The adequacy of this local framework is assessed.

Section 6 considers the case where ρ, the AR parameter, approaches 1. Our asymptotic analysis provides a limiting distribution for processes with nearly two unit roots. Section 7 finally considers the case where ρ approaches -1. Here the framework is shown to be related to a nearly integrated seasonal model of period 2. These asymptotic analyses help to understand the differing behavior of the normalized least squares estimator as ρ approaches plus or minus one. Section 8 provides some concluding comments and an appendix contains mathematical derivations.

2. THE LIMITING DISTRIBUTION OF $T(\hat{\alpha} - \alpha)$

This section considers the limiting distribution of the normalized least–squares estimator $T(\hat{\alpha} - \alpha)$ in the near–integrated model with possibly dependent errors. We also consider its asymptotic expansion up to order $O(T^{-1})$. As a matter of notation , we denote, throughout the paper, weak convergence in distribution by '$\Rightarrow$' and equality in distribution by '$\stackrel{d}{=}$'. The asymptotic analysis to be considered can be obtained under various conditions upon the error structure. These sets of conditions vary according to the measure of temporal dependence used. For our purposes it does not really matter which of these conditions are used as most of them permit stationary and invertible Gaussian ARMA processes of finite order. For the sake of simplicity we consider those of Herrndorf (1984) involving the concept of strong mixing. These are the same conditions as used in Phillips (1987a) and Phillips and Perron (1988) and are stated as follows :

ASSUMPTION 1 : *(a) $E(u_t) = 0$ for all t; (b) $\sup_t E|u_t|^{\beta+\epsilon} < \infty$ for some $\beta > 2$ and $\epsilon > 0$; (c) $\sigma^2 = \lim_{T\to\infty} T^{-1}E(S_T^2)$ exists and $\sigma^2 > 0$, where $S_t = \Sigma_1^t u_j$; (d) $\{u_t\}_1^\infty$ is strong mixing with mixing numbers that satisfy : $\Sigma_1^\infty \alpha_m^{1-2/\beta} < \infty$.*

When the sequence $\{u_t\}$ is strictly stationary condition (c) is implied by (a), (b) and (d) and $\sigma^2 = 2\pi f_u(0)$, where $f_u(0)$ is the non–normalized spectral density function of $\{u_t\}$ evaluated at frequency zero. When considering the asymptotic expansions of order $O(T^{-1})$ the following additional restriction will be imposed on the sequence of errors $\{u_t\}$.

ASSUMPTION 2 : *$\{u_t\}_1^\infty$ is a Gaussian weakly stationary sequence.*

Consider the construction of random elements $X_T(r)$ lying in the space D[0,1] of real–valued functions on the interval [0,1] that are right continuous and have finite left limits, and endowed with the uniform metric. $X_T(r)$ is defined as :

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

From Herrndorf (1984) we have the following functional central limit theorem valid under Assumption 1 :

(2.2) $$X_T(r) \Rightarrow W(r) \, ,$$

where W(r) is the unit Wiener process on C[0,1]. Phillips (1987b) proved (1.4) under the conditions of Assumption 1 using the result (2.2). We consider here an extension of his (1987c) result concerning the asymptotic expansion up to order $O(T^{-1})$ of the normalized least–squares estimator $T(\hat{\alpha} - \alpha)$. Our result is contained in the following Theorem proved in the Appendix.

THEOREM 1 : *Let $J_c(r) = \int_0^r \exp((r-s)c)dW(r)$, $\lambda = (\sigma^2 - \sigma_u^2)/(2\sigma^2)$ where $\sigma_u^2 = \lim_{T\to\infty} T^{-1}E(\Sigma_{t=1}^T u_t^2)$ and σ^2 is as defined in Assumption 1; $\gamma = y_0/(\sigma T^{1/2})$; $\nu^2 = 2\pi f_{u^2}(0)$, where $f_{u^2}(0)$ is the non–normalized spectral density function of $\{u_t^2 - E(u_t^2)\}$ evaluated at frequency 0 ; η is a N(0,1) random variable independent of the Wiener process W(r) ; then under Assumptions 1 and 2 :*

$$T(\hat{\alpha} - \alpha) \stackrel{d}{=} H(c,\gamma)/K(c,\gamma) + O_p(T^{-1})$$

where $$H(c,\gamma) = \int_0^1 J_c(r)dW(r) + \lambda + \gamma\int_0^1 \exp(cr)dW(r) - (\nu/(2\sigma^2 T^{1/2}))\eta$$

and $$K(c,\gamma) = \int_0^1 J_c(r)^2 dr + 2\gamma\int_0^1 \exp(cr)J_c(r)dr.$$

Using the normalization $\sigma_e^2 = 1$, we have the following specifications for the various variables in the MA(1) and AR(1) processes as specified in (1.5) and (1.6). For the MA(1) case $\sigma_u^2 = 1 + \theta^2$, $\sigma^2 = (1 + \theta)^2$, hence $\lambda = \theta/(1 + \theta)^2$ and $\gamma = y_0/[T^{1/2}(1 + \theta)]$. Also $\nu^2 = 2(1 + 4\theta^2 + \theta^4)$. In the AR(1) case we have : $\sigma_u^2 = (1 - \rho^2)^{-1}$, $\sigma^2 = (1 - \rho)^{-2}$; hence $\lambda = \rho/(1 + \rho)$ and $\gamma = y_0(1 - \rho)/T^{1/2}$. Also $\nu^2 = 2(1 + \rho^2)/(1 - \rho^2)^3$.

The asymptotic expansion of $T(\hat{\alpha} - \alpha)$ is directly affected by the value of the initial condition. This is also the case in the continuous records asymptotic distribution considered in Perron (1988a). With i.i.d. errors we showed that as $T \to \infty$ keeping a fixed span (N) of the data (or letting the sampling interval h = N/T converge to zero at the same rate as $T \to \infty$) :

$$T(\hat{\alpha} - \alpha) \Rightarrow A(c,\gamma)/B(c,\gamma)$$

where $$A(c,\gamma) = \int_0^1 J_c(r)dW(r) + \gamma\int_0^1 \exp(cr)dW(r)$$

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

and where in this context c is related to the coefficient of the underlying Ornstein–Uhlenbeck process. These distributions are similar in a way that will be interesting for purposes of obtaining the limiting joint moment–generating function of $\{T^{-1}\Sigma_1^T y_{t-1}u_t, T^{-2}\Sigma_1^T y_{t-1}^2\}$ that allows computation , by numerical integration, of distributional quantities related to the asymptotic distribution in Theorem 1.

To this effect consider the exact joint moment–generating function of $\{A(c,\gamma), B(c,\gamma)\}$, denoted by $M_{c,\gamma}(v,u) = E[\exp(vA(c,\gamma) + uB(c,\gamma)]$. It was shown in Perron (1988a) that :

(2.3) $$M_{c,\gamma}(v,u) = \psi_c(v,u)\exp\{-(\gamma^2/2)(v + c - \lambda)(1 - \exp(v + c + \lambda)\psi_c^2(v,u))\}$$

where

(2.4) $$\psi_c(v,u) = \left\{2\lambda\exp(-(v + c))/[(\lambda + v + c))\exp(-\lambda)+(\lambda - v - c))\exp(\lambda)]\right\}^{1/2} ;$$

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

Using (2.3) it is easy to obtain the joint moment–generating function of $\{H(c,\gamma), K(c,\gamma)\}$, which we denote by

$$MGF(v,u) = E[\exp\{vH(c,\gamma) + uK(c,\gamma)\}] .$$

The result is summarized in the following Theorem.

THEOREM 2 : *Let $g = \nu/(2\sigma^2 T^{1/2})$, $d = \gamma^2(exp(2c) - 1)/2c$, and $M_{c,\gamma}(v,u)$ be as defined in (2.3) and (2.4) ; then the joint moment–generating function of $\{H(c,\gamma), K(c,\gamma)\}$ is given by :*

$$MGF(v,u) = exp(v\lambda - ud + v^2 g^2/2)\, M_{c,\gamma}(v,u) .$$

Proof : We can write $H(c,\gamma) = A(c,\gamma) + \lambda - g\eta$, and $K(c,\gamma) = B(c,\gamma) - d$. Hence :

$$MGF(v,u) = E[\exp\{v(A(c,\gamma) + \lambda - g\eta) + u(B(c,\gamma) - d)\}]$$

$$= \exp(v\lambda - ud)E[\exp(-vg\eta)]E[\exp\{vA(c,\gamma) + uB(c,\gamma)\}]$$

given that η is independent of the Wiener process W(r) and that λ and d are fixed constants. With η a N(0,1) random variable, $E[\exp(-vg\eta)] = \exp(v^2 g^2/2)$ and the result follows using the definition of $M_{c,\gamma}(v,u)$. □

The result in Theorem 2 allows direct computation, by numerical integration, of the cumulative distribution and probability density functions as well as the moments of the asymptotic distribution. More precisely , let the joint characteristic function of $\{H(c,\gamma), K(c,\gamma)\}$ be given by :

$$CF(v,u) = MGF(iv,iu) = E[\exp\{ivH(c,\gamma) + iuK(c,\gamma)\}] .$$

Using Theorem 1 of Gurland (1948) , the limiting distribution function of $T(\hat{\alpha} - \alpha)$ is given by :

(2.5) $$F(z) = (1/2) - (1/2\pi i)\int_0^\infty [CF(v,-vz)/v]dv$$

$$= (1/2) - (1/2\pi)\int_0^\infty AIMAG[CF(v,-vz)/v]dv$$

where AIMAG( ) denotes the imaginary part of the complex number. Similarly, the probability density function is given by :

(2.6) $$f(z) = dF(z)/dz = (1/2\pi i)\int_0^\infty \left\{\partial CF(v,u)/\partial u\right\}_{u=-vz} dv .$$

The moments of the asymptotic distribution can be obtained using Mehta and Swamy's (1978) result, which in our case implies :

$$(2.7) \qquad E[H(c,\gamma)/K(c,\gamma)]^r = \Gamma(r)^{-1}\int_0^\infty u^{r-1}\left\{\partial^r MGF(v,-u)/\partial v^r\right\}_{v=0} du .$$

These results allow computation of distributional quantities for quite a variety of processes. In particular, quantities for the usual asymptotic distribution (1.4) can be obtained by simply letting $g = \gamma = 0$. We then have :

$$(2.8) \qquad MGF(v,u) = \exp(v\lambda)\psi_c(v,u) ,$$

where $\psi_c(v,u)$ is defined in (2.4). The next section makes extensive use of Theorem 2, (2.5) and especially (2.8), to calculate the asymptotic distribution of $T(\hat{\alpha} - \alpha)$ when the errors are either MA(1) or AR(1) processes.

3. APPROXIMATING THE EXACT DISTRIBUTION OF $T(\hat{\alpha} - \alpha)$

This section considers the adequacy of the asymptotic distribution in approximating the exact distribution of $T(\hat{\alpha} - \alpha)$ in finite samples. The asymptotic values are obtained using the results of Section 2 and the finite sample values are obtained via simulations. As discussed in the introduction, we consider two leading cases where the error sequence is either an MA(1) or an AR(1) process. As we will see, these cases are sufficient to provide a rich characterization of the relationship between the finite sample results and their asymptotic counterparts. Given that our aim is mainly oriented towards studying the effect of correlation in the errors we consider only the case where $y_0 = 0$.

The setup of the experiment is as follows. For each of the MA and AR cases we consider three values of c, namely c = 0.0, – 5.0 and 2.0. When c = 0.0 , we have a unit root process. When c = – 5.0 , the process is locally stationary and when c = 2.0 it is locally explosive. For each value of c we consider the following specifications for the errors[1] : a) MA(1) case , $\theta = -0.9, -0.7, -0.5, -0.3$ and 0.5 ; for the AR(1) case , $\rho = -0.9, -0.5, 0.5, 0.9$ and 0.95. For each of these values we consider sample sizes of length T = 25, 50, 100, 500, 1000 (except $\theta = -0.3, 0.5$ and $\rho = -0.5, 0.5, 0.9$), and 5000 (for $\theta = -0.9$). The finite sample results are obtained using 10,000 replications (5,000 when T = 5,000). The asymptotic results for the cumulative distribution function are obtained using (2.5) with $\gamma = g = 0$ for the O(1) asymptotic (see (2.8)). The numerical integrations were performed using the subroutine QDAG of the IMSL library. The bounds of integration are given by $(\epsilon, \bar{V})$ where $\bar{V}$ is chosen such that the square of the integrand evaluated at $\bar{V}$ is less than ϵ. The error criterion for the numerical integration was also set at ϵ. For each experiment we set ϵ at 1.0E–07. Special care must be taken with the numerical integration since it involves the square root of a complex valued quantity. Use of the principal value may not ensure the continuity of the integrand. The numerical integration must be performed over Reimann surfaces consisting here of two planes. The method is described in more detail in Perron (1989a). This latter paper also present some evidence that, in the case were the errors are uncorrelated, the asymptotic distribution is a quite satisfactory guide to the finite sample distribution.

Tables I and II present the percentage points of the distribution of $T(\hat{\alpha} - \alpha)$ in each of the cases discussed above. Consider first the MA(1) case presented in Tables I.A (c = 0.0) , I.B (c = – 5.0) and I.C (c = 2.0). When $\theta = 0.5$, the asymptotic approximation is excellent in each case, even for very small samples sizes (e.g. T = 25). Further experiments (not reported) show this adequacy for any process with positively autocorrelated MA(1) errors.

When θ is negative the picture is rather different. Here the adequacy of the approximation depends very much upon the magnitude of θ and deteriorates quite rapidly as θ approaches -1. Consider the case where c = 0.0 (for which some of the results discussed are presented in Table I.A). For different values of θ the asymptotic distribution adequately approximates the finite sample distribution for the following sample sizes : $\theta = -0.3$, $T \geq 500$; $\theta = -0.5$, $T \geq 1000$; $\theta = -0.7$, $T \geq 5000$; and when $\theta = -0.9$, the asymptotic distribution is still quite far away from the exact distribution when T = 5000. The differences are quite substantial. For example, the first percentage point of the exact distribution when $\theta = -0.9$ and T = 50 corresponds approximately to the 95th percentage point of the asymptotic distribution.

Tables I.B and I.C show the same qualitative results for the cases where c = – 5.0 and c = 2.0. The main difference is that the approximation is marginally better when the process is explosive (c = 2.0) and marginally worse when it is stationary (c = – 5.0).

Consider now the case where the errors are AR(1). The results are presented in Tables II (A,B,and C). Table II.A presents the case of a unit root process where c = 0.0. The picture is quite different from the MA(1) case. For $\rho < 0$, the approximation is again inadequate and worsens as ρ approaches -1. However, for comparable values of ρ and θ the approach of the finite sample distribution to the asymptotic distribution is faster in the AR(1) case than it is in the MA(1) case. For $\rho = -0.5$, the asymptotic approximation is adequate for $T \geq 500$, and for $\rho = -0.9$ when $T \geq 1000$. For smaller sample sizes there are important discrepancies especially when $\rho = -0.9$. However, these discrepancies are not as severe as in the MA(1) case. For example, when T = 25, the first percentage point of the exact distribution with $\rho = -0.9$ corresponds roughly to the median of the asymptotic distribution (in the MA(1) case the first percentage point of the exact distribution with $\theta = -0.9$ and T = 25 corresponds to the 99th percentage point of the asymptotic distribution). As shown in Tables II.B and II.C , similar qualitative results hold in the stationary and explosive cases. When c = – 5.0, the adequacy is marginally inferior and when c = 2.0 it is marginally superior.

The AR(1) case with positive autocorrelation offers yet a different picture. First, unlike all the cases considered so far the approach of the finite sample distribution to its asymptotic counterpart is from a density with a larger spread to one with a smaller spread. Secondly, the differences between the finite sample percentage points and the asymptotic percentage points are not substantial. For example when $\rho = 0.9$ and $T = 25$, the tenth percentage point of the exact distribution corresponds roughly to the fifth percentage point of the asymptotic distribution. Nevertheless, what is interesting , and different from the MA(1) case with positive coefficient, is the fact that the approach to the asymptotic value is quite slow. The tail of the exact distribution is not well approximated until $T = 500$.

Tables II.B and II.C present the results when $c = -5.0$ and $c = 2.0$ respectively. The same qualitative behavior remains as in the case where $c = 0.0$. However several interesting differences emerge. Again the approximation is marginally better when $c = 2.0$ and marginally worse when $c = -5.0$. More interestingly , when $c = -5.0$ we notice a difference from the cases where $\rho > 0$ and $c = 0.0$ or 2.0. The right tail seems to be much better approximated by the asymptotic distribution even for quite small sample sizes (e.g. $\rho = 0.9$ and $T = 25$). Secondly, when compared to the MA case, there is much more movement in the percentage points as c varies.

The results presented in this section suggest that the quality of the asymptotic approximation is heavily dependent on the nature and extent of the correlation in the residuals. When there is negative autocorrelation, the approximation becomes rapidly useless as the magnitude of this correlation increases, both in the MA(1) and AR(1) cases. In the AR(1) case with positive autocorrelation, the discrepancies are not as severe but the approach to the asymptotic distribution remains quite slow. Only in the MA(1) case with positive autocorrelation is the approximation adequate, indeed as good as in the case with no correlation.

We also performed also an extensive analysis of the behavior of the mean and variance of the exact and asymptotic distributions. For reason of space constraint, we report only the figures for values of c and θ or ρ presented in Tables I and II. The asymptotic results were obtained using numerical integration of the function in (2.7). The specifications are basically the same as in the numerical integration of the cumulative distribution function, except that here the integrand does not involve complex valued quantities, so only straightforward numerical integration routines are needed.

Consider first the case of the mean of the distribution when $c = 0.0$. For the MA case and $\theta \geq 0$, the mean of the asymptotic distribution is basically equal to the mean of the exact distribution when $T \geq 100$. When $T < 100$, the discrepancies are minor. When $\theta < 0.0$, the picture is rather different. As θ approaches one, it takes an increasingly larger sample size to have the mean of the exact distribution correspond to that of the asymptotic distribution. When $\theta = -0.1$, a sample of size 100 still appear enough but there is a larger discrepancy with smaller sample sizes compared to the case where $\theta \geq 0$. When θ is between -0.2 and -0.5, a sample of size 500 is needed to ensure a satisfactory approximation. When $\theta = -0.6$, a sample of size 1000 is needed and when $\theta = -0.7$ or -0.8, the corresponding figure is $T = 5000$. Finally, when $\theta = -0.9$ or -0.95 even a sample size as large as 5,000 is not sufficient to provide an adequate approximation.

Of particular interest is the fact that when $\theta \leq -0.5$, the mean of the finite sample distribution changes very rapidly as T increases. Hence, for this part of the parameter space , the asymptotic distribution provides a very bad approximation to the mean of the exact distribution when the sample size is not very large.

Consider now the AR(1) case. When $\rho \leq 0$, the picture is similar to that of the MA case, except that the discrepancies between the mean of the asymptotic and exact distributions are not as severe. When ρ is between -0.1 and -0.5, the exact mean attains its asymptotic value when T reaches somewhere between $T = 100$ and $T = 500$. When ρ is between -0.6 and -0.95 , the correspondence is achieved with a sample size somewhere between $T = 500$ and $T = 1000$. When $\rho > 0$, the picture is different. With ρ between 0.1 and 0.5, a sample size as small as 50 is enough to provides a adequate approximation. When $\rho = 0.7$, a sample of size 100 is needed, and with $\rho = 0.9$ or 0.95, the corresponding figure is $T = 500$. Also of interest is the fact that when ρ is between 0.5 and 0.95, the mean is positive (unlike all the other cases considered). Finally, also to be noted is the fact that in all cases considered so far, the mean of the exact distribution approaches its asymptotic counterpart in a monotonically decreasing way [2].

With $c = -5.0$, a locally stationary process, the asymptotic approximation is, in general, less good than in the case where $c = 0.0$ for both the MA and AR cases ; i.e., for a given θ or ρ and sample size T, the discrepancy between the exact mean and its asymptotic counterpart is greater. A feature that is different, however, is the fact that when θ or ρ is greater than 0.3, the approach to the asymptotic value is achieved in a monotonically *increasing* way. When $c = 2.0$, the locally explosive case, the general features are similar but the discrepancies between the exact and asymptotic results are not as severe, compared to both cases where $c = 0.0$ or -5.0. As was the case with $c = 0.0$, the approach to the asymptotic value as T increases is monotonically increasing. A small difference from earlier cases is that, unless T is small and $\rho = 0.9$ or 0.95, the mean of the distribution is negative.

Note that the use of the $O(T^{-1})$ asymptotic expansion described in Theorem 1 does not provide any improvement over the usual O(1) asymptotic distribution. Indeed both of them yields the same mean. This is due to the fact that the $O(T^{-1})$ expansion does not provide any location adjustment given that the extra term η has mean zero and is independent of the Wiener process W(r) present in the other components.

Tables I and II also present the results for the variance of the distribution. Consider first the case of a unit root process, $c = 0.0$. For the MA(1) case (Table I.A), the general features are similar as for the mean of the distribution. An interesting difference is that when $\theta < 0$ the approach to the asymptotic value appears to be slower. For example, when $\theta = -0.7$ and $T = 1000$, the mean of the exact distribution is quite close to the mean of the asymptotic distribution, but the variance is still quite far away. Of particular interest is the fact that, for a given θ, the exact variance approaches its asymptotic value in a monotonically increasing way. However, for a given sample size, the variance does not increase monotonically as θ approaches -1. For example, with $T = 50$ the variance of the distribution of $T(\hat{\alpha} - \alpha)$ is 66.38 ($\theta = -0.5$), 142.37 ($\theta = -0.7$) and 119.64 ($\theta = -0.9$). These features will be further discussed in later sections. In particular, the alternative asymptotic framework to be derived in Section 5 is able to explain such a phenomenon.

Much of the same features apply for the AR case (Table II.A); in particular the discrepancies between the exact and asymptotic results are not as severe as for the MA(1) case. Some interesting features are, however, different from the MA(1) case. First, for a given value of the sample size, the variance is monotonically increasing as ρ approaches -1. More importantly, when $\rho \geq 0.3$ the exact variance approaches its asymptotic value in a monotonically decreasing way unlike the case with MA(1) errors or with negatively correlated AR(1) errors. This feature will prove of interest when considering the $O(T^{-1})$ asymptotic expansions later.

The behavior of the variance when $c = -5.0$ or 2.0 is similar to that when $c = 0.0$. As was the case for the mean, the discrepancies between the exact and asymptotic results are more severe with $c = -5.0$ and less severe with $c = 2.0$. Apart from this fact, the only notable difference is that in the MA(1) case with $\theta \geq 0.1$ and $c = 2.0$, the exact variance now appears to decrease monotonically towards its asymptotic value.

Unlike for the mean of the distribution, the $O(T^{-1})$ asymptotic expansion provides an adjustment to the variance of the asymptotic distribution. However, given the independence of the variable η and the Wiener process $W(r)$, the $O(T^{-1})$ expansion yields a *higher* variance than the $O(1)$ asymptotic distribution. As we saw, for most cases the exact variance is *smaller* than the asymptotic variance. This implies that , for most cases, the $O(T^{-1})$ asymptotic expansion provides a *less* accurate approximation to the exact variance than the $O(1)$ asymptotic distribution. Given that the $O(T^{-1})$ asymptotic expansion does not provide any adjustment to the mean of the $O(1)$ asymptotic distribution, it follows that, in most cases, this asymptotic expansion provides a *poorer* approximation than the usual $O(1)$ asymptotic distribution. The only case where the asymptotic expansion could yield a better approximation to the exact variance is when the variance approaches its asymptotic value in a monotonically decreasing way. This is the case with AR(1) errors and $\rho \geq 0.3$. Even in those cases the $O(T^{-1})$ expansion will provide a better approximation to the exact distribution only if the mean is fairly stable as T increases. In our experimental setting this is the case only when ρ is near 0.5 and $c = 0.0$ or 2.0.

We performed a number of calculations related to the distribution of the asymptotic expansion to order $O(T^{-1})$. The results were as outlined above. Not only does it not provide a better approximation, but the extra $O(T^{-1/2})$ term seems in almost all cases to worsen the approximation. For these reasons we decided not to report these results, but they are available upon request. The only case where the asymptotic expansion provided an interesting improvement is the AR(1) case with positive coefficient. We postpone the presentation and discussion of the corresponding results to Section 6.

4. A LOOK AT THE FUNCTIONAL CENTRAL LIMIT THEOREM WITH DEPENDENT ERRORS

As stated in Section 2 the limiting distribution of Theorem 1 is obtained using a functional central limit theorem for dependent variables. Consider a special case of (2.2) with $r = 1$. Under the conditions of Assumption 1 we have :

$$X_T(1) \Rightarrow W(1) , \tag{4.1}$$

where $X_T(1) = T^{-1/2}\sigma^{-1}S_T$, $S_T = \Sigma_1^T u_t$ and $W(1) \sim N(0,1)$. Given the simple structure of our framework it is possible to provide an exact analysis of the adequacy of such a functional central limit theorem when applied to the MA(1) and AR(1) processes defined by (1.5) and (1.6).

First let us note that when $\theta > -1$ and $-1 < \rho < 1$, the conditions of Assumption 1 are satisfied for the sequence $\{u_t\}$. In particular given that the basic innovations $\{e_t\}$ are Gaussian with mean 0 , so are the sequence $\{u_t\}$ and the sum S_T in both the MA(1) and AR(1) cases. Hence, the only difference between the exact distribution of $X_T(1)$ and $W(1)$ is that the variance of $X_T(1)$ need not equal 1 in finite samples. The adequacy of the functional central limit theorem can then be assessed by determining to what extent the finite sample variance is different from 1.

Consider first the MA(1) case. Using the fact that $\sigma^2 = (1 + \theta)^2$ it is easy to deduce that :

$$\mathrm{Var}(X_T(1)) = 1 + k/T \text{ , where } k = 1/(1 + \theta)^2 - 1 . \tag{4.2}$$

Similarly in the AR(1) case , $\sigma^2 = (1 - \rho)^{-2}$ and :

$$\mathrm{Var}(X_T(1)) = 1 + \rho^2(1 - \rho^{2T})/(1 - \rho^2) - 2\rho(\rho^T - 1)/(1 - \rho) . \tag{4.3}$$

It is interesting to note that the exact variance of $X_T(1)$ in the AR(1) case is bounded between 0 and 2 for values of ρ between -1 and 1. Such is not the case with MA(1) errors as the variance is unbounded as θ approaches – 1. Table III presents the percentage error between the exact and asymptotic variance of $X_T(1)$ for a range of values of T and θ (in the MA(1) case) and ρ (in the AR(1) case).

The results are quite striking. Consider first the MA(1) case. When $\theta > 0$, the error is small and vanishes rapidly (less than 1 % for T $\geq$ 100). On the other hand when θ is negative , the error is large and more so as θ approaches – 1. In particular, it remains quite high for θ close to – 1 even with a sample size as large as 5,000. These observations may go quite a way in explaining the inadequacy of the asymptotic approximation when θ has a large negative value , and the fact that the approximation is good when θ is positive. Basically the underlying functional central limit theorem is unreliable when θ is "close" to minus one.

In the case of AR(1) errors the picture is different. First, as noted above, the absolute discrepancy is never more than 100 percent. Secondly it becomes negligible more quickly for all parameter values (e.g. it is less than 1 % for T = 5,000 and any value of ρ presented). The most important difference is, however, that for small sample sizes it is large for both negative and positive values of ρ. More interestingly, for a given absolute value of ρ, the error is larger if ρ is positive than if it is negative. These features are helpful because, as our experimental study showed, the discrepancies between the exact and asymptotic distributions are larger with negatively correlated MA(1) errors than with negatively correlated AR(1) errors. But more puzzling is the fact that the discrepancies between the exact and asymptotic distributions of $T(\hat{\alpha} - \alpha)$ are very much *smaller for positively correlated* AR(1) errors, than they are with negatively correlated AR(1) errors. Yet the functional central limit theorem (5.1) performs comparatively *worse if* ρ *is positive* than if ρ is negative. Hence the reason for the differing performance of the asymptotic distribution must be sought elsewhere.

Some insights can be gained by looking at the behavior of the parameter λ in the asymptotic distribution given in Theorem 1. In the MA case $\lambda = \theta/(1 + \theta)^2$. λ becomes unbounded when θ approaches – 1, and decreases to 0 as θ increases. In the AR(1) case $\lambda = \rho/(1 + \rho)$. It again diverges to $-\infty$ as ρ approaches – 1 but at a smaller rate than in the MA(1) case. When ρ approaches 1, λ approaches 1/2.

These considerations lead to the following conjectures. First the asymptotic distribution is a bad approximation when θ approaches -1 because the asymptotic distribution of $T(\hat{\alpha} - \alpha)$ letting θ approach -1 is unbounded . However when θ approaches 1 the asymptotic distribution is still valid and the functional central limit theorem is an adequate approximation for relatively small sample sizes. In the AR(1) case the asymptotic distribution of $T(\hat{\alpha} - \alpha)$ is again unbounded if $\rho \to -1$ as $T \to \infty$. However, the rate at which ρ may approach -1 to obtain a non–degenerate local asymptotic distribution is higher in the AR(1) case than it is in the MA(1) case. This would explain the relatively smaller discrepancies in the AR(1) case for a given equal value of θ and ρ. On the other hand, when ρ approaches $+1$ as $T \to \infty$, $T(\hat{\alpha} - \alpha)$ still has a non–degenerate asymptotic distribution but different from that given by (1.4).

The rest of this paper is devoted to making precise these conjectures. We derive local asymptotic distribution results letting θ approach -1 as $T \to \infty$, and letting ρ approach -1 or $+1$ in the AR(1) case.

5. A NEARLY WHITE NOISE NEARLY INTEGRATED PROCESS

In this Section, we propose an alternative asymptotic framework that is intended, on the one hand, to provide an asymptotic distribution which better approximates the exact distribution of $T(\hat{\alpha} - \alpha)$ when the errors have an MA(1) structure with large negative correlation, i.e. when θ is close to -1. On the other hand, our intention, using this alternative approach, is also to explain some of the finite samples phenomena described in Section 3. Consider the following parameterization of the nearly integrated process with MA(1) errors :

(5.1) $$y_t = \exp(c/T)y_{t-1} + e_t + \theta_T e_{t-1} \ ,$$

where

(5.2) $$\theta_T = -1 + \delta/T^{1/2}.$$

For simplicity we assume, as in the finite sample simulation experiments, that $e_t \sim$ i.i.d. $N(0,\sigma_e^2)$. The process defined by (5.1) and (5.2) is an ARMA(1,1) where the autoregressive root approaches 1 and the moving average root approaches -1 as T converges to infinity. In the limit, the roots cancel and the process $\{y_t\}$ is white noise provided the sequence $\{e_t\}$ is white noise. However, in any finite sample, $\{y_t\}$ is nearly integrated, hence the expression "nearly white noise – nearly integrated model". A variant of this specification, with $c = 0.0$, has been considered by Pantula (1988) in a different context. Our aim, in this section, is to study the asymptotic distribution of $\hat{\alpha}$ under the specification (5.1) and (5.2). The next Theorem, proved in the Appendix, characterizes this asymptotic distribution.

THEOREM 3 : *Let $\{y_t\}$ be a sequence of random variables generated by (5.1) and (5.2), then as $T \to \infty$:*

$$\hat{\alpha} \Rightarrow \left\{\delta^2 \int_0^1 J_c(r)^2 dr\right\} \left\{1 + \delta^2 \int_0^1 J_c(r)^2 dr\right\}^{-1} ;$$

where $J_c(r) = \int_0^r exp((r-s)c)dW(r)$, and $W(r)$ is the unit Wiener process on $C[0,1]$.

Theorem 3 shows that under this nearly white noise setting the asymptotic distribution of $\hat{\alpha}$ is degenerate unless $\delta = 0$, in the sense that $\hat{\alpha}$ converges to a random variable instead of fixed constant. Hence $\hat{\alpha}$ is not a consistent estimator of α. If $\delta = 0$, we have that $\hat{\alpha} \to 0$ in probability as expected ; and as $\delta \to \infty$ the limit of $\hat{\alpha}$ tends to 1.

This result helps to explain some of the findings in Section 3. Note first that, under the present setting, $T(\hat{\alpha} - \alpha)$ is unbounded and converges to $-\infty$. Hence , on the one hand, we would expect the distribution of $T(\hat{\alpha} - \alpha)$ to shift leftward as θ decreases. On the other hand, we would also expect the usual asymptotic approximation to be inadequate for values of θ close to -1.

Theorem 3 presents an alternative distributional theory that could provide a more adequate approximation to the exact distribution of $T(\hat{\alpha} - \alpha)$ for values of T and θ where the usual asymptotic theory fails to provide a useful guide. To investigate this issue the next Theorem presents the limiting joint moment–generating function of $\{T^{-1}\Sigma_1^T y_t y_{t-1}$, $T^{-1}\Sigma_1^T y_{t-1}^2\}$ that will allow computation of distributional quantities related to $\hat{\alpha}$, in a manner similar to the methods used in Section 3.

THEOREM 4 : *Denote the joint moment–generating function of* $\{\delta^2\int_0^1 J_c(r)^2 dr,$ $1 + \delta^2\int_0^1 J_c(r)^2 dr\}$ *by* $MG_{\delta,c}(v,u) = E[exp\{v(\delta^2\int_0^1 J_c(r)^2 dr) + u(1 + \delta^2\int_0^1 J_c(r)^2 dr)\}]$, *then :*

$$MG_{\delta,c}(v,u) = exp(u - (c - \lambda)/2)[1 - (c - \lambda)(exp(2\lambda) - 1)/2\lambda]^{-1/2}$$

where $\lambda = (c^2 - 2\delta^2(v + u))^{1/2}$.

Proof : The stochastic process $J_c(r)$ is defined on a probability space (Ω, F), say, with probability measure μ_c that is induced by the following diffusion process :

$$(5.3) \qquad dJ_c(r) = cJ_c(r)dr + dW(r) \;\; ; \;\; J_c(0) = 0 \qquad (0 \leq r \leq 1) .$$

Consider now an alternative diffusion process z(r) defined on the same probability space (Ω, F) but with probability measure μ_λ induced by the diffusion process :

(5.4) $dz(r) = \lambda z(r)dr + dW(r) \; ; \; z(0) = 0 \qquad (0 \le r \le 1)$.

From Theorem 7.19 of Liptser and Shiryayev (1978) the measures μ_c and μ_λ are equivalent and the density or Radon–Nykodym derivative $d\mu_c/d\mu_\lambda$ evaluated with z(r) is given by :

(5.5) $$d\mu_c/d\mu_\lambda\,(z) = \exp\left\{(c-\lambda)\int_0^1 z(r)dz(r) - (1/2)(c^2-\lambda^2)\int_0^1 z(r)^2 dr\right\}.$$

We have :

$$MG_{\delta,c}(v,u) = E[\exp\{v(\delta^2\int_0^1 J_c(r)^2 dr) + u(1+\delta^2\int_0^1 J_c(r)^2 dr)\}]$$

$$= \exp(u)E[\exp\{\delta^2(v+u)\int_0^1 J_c(r)^2 dr\}]$$

$$= \exp(u)E[\exp\{\delta^2(v+u)\int_0^1 z(r)^2 dr\}(d\mu_c/d\mu_\lambda(z))]$$

$$= \exp(u)E[\exp\{[\delta^2(v+u) - (c^2-\lambda^2)/2]\int_0^1 z(r)^2 dr + (c-\lambda)\int_0^1 z(r)dW(r)\}]$$

using (5.5). Now let $\lambda = (c^2 - 2\delta^2(v+u))^{1/2}$, then :

$$MG_{\delta,c}(v,u) = \exp(u)E[\exp\{(c-\lambda)\int_0^1 z(r)dW(r)\}].$$

By Ito's Lemma $\int_0^1 z(r)dW(r) = (z(1)^2-1)/2$, given that $z(0) = 0$. Hence :

$$MG_{\delta,c}(v,u) = \exp(u)E[\exp\{(c-\lambda)(z(1)^2-1)/2\}]$$

(5.6) $$= \exp(u-(c-\lambda)/2)E[\exp\{(c-\lambda)z(1)^2/2\}].$$

Note that $z(1) \sim N(0,(\exp(2\lambda)-1)/2\lambda)$. Let $s^2 = (\exp(2\lambda)-1)/2\lambda$, then $z(1)^2/s^2 \sim \chi_1^2$ and

$$E[\exp\{(c-\lambda)z(1)^2/2\}] = (1-(c-\lambda)s^2)^{-1/2}.$$

The result follows upon substitution in (5.6) and some rearrangements. □

Theorem 4 allows the computation, by numerical integration, of the c.d.f. , p.d.f. and moments of the limiting distribution of $\hat{\alpha}$ substituting this moment–generating function in expressions (2.5) through (2.7). To get an idea of the type of distribution involved, Figure 1 graphs the mean and standard deviation of the limiting distribution of $\hat{\alpha}$ as a function of δ, for the three cases c = 0.0, – 5.0 and 2.0. As expected for δ close to 0 the mean is 0, and as δ increases the mean approaches 1. The standard deviation of the process is close to 0 when either δ is very small or very large. As δ moves away from 0, both the mean and standard deviation increase more rapidly with the parameter c. From these considerations, we would expect : a) the approximation of the finite sample distribution to worsen as either δ gets large or close to zero (due to the implied zero variance) , and b) the approximation of the mean to be more adequate for small values of c (due to a less rapid approach of the mean of the asymptotic distribution towards 1) ; and c) the approximation of the variance to be more adequate for large values of c (due to a less flat asymptotic function).

To use the asymptotic distribution of Theorem 3 as an approximation to the exact distribution of $T(\hat{\alpha} - \alpha)$, we specify the correspondence $\delta = T^{1/2}(1 + \theta)$. From the comments above one would expect a better approximation for combinations of T and θ such that δ is neither too small nor too large. Table IV presents the percentage points of the distribution of $T(\hat{\alpha} - \alpha)$ calculated using this nearly white noise – nearly integrated asymptotic distribution. The cases considered are c = 0.0 ; $\theta = -0.9, -0.7$ and -0.5, with T = 25, 50, 100, 500, 1000 (for $\theta = -0.9$), and T = 5000 (for $\theta = -0.9$). These values are to be compared with the exact percentage points given in Table I.A. For $\theta = -0.9$, the approximation is excellent with $T \geq 500$, especially in the left tail. When T = 100 the approximation is still respectable but deteriorates as T reaches 50 or 25, especially in the right tail. Nevertheless, in all cases the approximation is much better than the standard asymptotic distribution considered in Section 3. When $\theta = -0.7$, the approximation is best when T = 25 or 50 and deteriorates as T gets larger. Again the left tail is much better approximated than the right tail. When $\theta = -0.5$, the extreme right tail of the distribution is badly approximated due to the implied negativity of the asymptotic distribution of $T(\hat{\alpha} - \alpha)$ provided by the nearly white noise local framework.

Table V presents the approximation to the mean and variance of $T(\hat{\alpha} - \alpha)$ provided by the nearly white noise asymptotic framework for the three values of c considered (c = 0.0, – 5.0, 2.0). These results are to be compared to those of Tables I (A,B, and C). With c = 0.0, the approximation is excellent for all sample sizes when $\theta = -0.9$. With $\theta = -0.7$, the approximation is adequate for samples of size less than 500. When θ is -0.50 the approximation is not as adequate for any sample sizes, though it is highly superior to the standard asymptotic approximation when $T \leq 500$. The same qualitative features hold when $c = 0.0$ or -5.0 but with a better approximation with $c = -5.0$ and less adequate with $c = 2.0$.

Consider now the variance of the distribution of $T(\hat{\alpha} - \alpha)$. When $\theta = -0.9$, the variance is badly approximated unless $T \geq 500$. When θ is -0.7 the approximation is reasonable for $T = 100$ and 500. When θ is -0.5 it is reasonable for smaller sample sizes. For the cases $c = -5.0$ and 2.0, the results show the same qualitative features but now, interestingly, the approximation is better when $c = 2.0$ and worse when $c = -5.0$ (unlike what was found for the mean of the distribution). A feature of particular interest is the behavior of the variance as θ approaches one with a given sample size. As remarked in Section 3, the exact results shows a non–monotonic behavior. This feature is well explained by this local asymptotic theory. Indeed, this non–monotonic behavior is present in several of the cases presented in Table V.B. The rational for this behavior can be obtained by looking at Figure 1 where it is shown that the standard deviation of the local asymptotic distribution of $T(\hat{\alpha} - \alpha)$ is zero when $\delta = 0$ and eventually approaches zero again as δ increases. Given that $\delta = T^{1/2}(1 + \theta)$, a decrease in θ for a fixed T implies that δ approaches 0. The non–monotonic behavior occurs when the change in θ is such as to move δ over the hump in the standard deviation function presented in Figure 1.

The results of our experiments show the nearly white noise – nearly integrated asymptotic distribution to be a far better approximation to the finite sample distribution of $T(\hat{\alpha} - \alpha)$ when θ is close to -1. However, the approximation still lacks some accuracy in an important range of cases. First when θ is somewhat away from -1 and T is large. This case, however, is not of much consequences since in this region the usual asymptotic theory is adequate. Of more consequence is the fact that the approximation is inadequate when T is small and θ is close to -1 (i.e., when δ is close to 0). Here none of the asymptotic distributions considered provide a satisfactory approximation.

6. A NEARLY TWICE INTEGRATED MODEL

The aim, in this Section, is to provide a local asymptotic framework that could explain the behavior of the distribution of $T(\hat{\alpha} - \alpha)$ when the errors have an AR(1) structure with (large) positive correlation. Our intention is also to assess whether this alternative asymptotic distribution provides a better approximation to the finite sample distribution, and to investigate to what extent it can complement the increased accuracy provided, in some cases, by the asymptotic expansion of order $O(T^{-1})$. We start with the following parameterization of the process of interest :

(6.1) $$y_t = \exp(c/T)y_{t-1} + u_t ,$$

(6.2) $$u_t = \exp(\phi/T)u_{t-1} + e_t ,$$

where , for simplicity, we specify $e_t \sim$ i.i.d. $N(0,\sigma_e^2)$. We can write (6.1) and (6.2) as :

(6.3) $$y_t = [\exp(c/T) + \exp(\phi/T)]y_{t-1} - \exp((c + \phi)/T)y_{t-2} + e_t .$$

As T converges to infinity y_t becomes :

$$y_t = 2y_{t-1} - y_{t-2} + e_t .$$

Therefore, as T increases, $\{y_t\}$ converges to a process with two real valued roots on the unit circle, hence the expression "nearly twice integrated". Our aim is to study the asymptotic behavior of $T(\hat{\alpha} - \alpha)$ under this specification.

We first need to define some new notation . Consider the following transformation of the random process $J_\phi(r)$:

(6.4) $$Q_c(J_\phi(r)) \equiv \int_0^r \exp((r - v)c)J_\phi(v)dv$$

where, as before, $J_\phi(v) = \int_0^v \exp((v - s)\phi)dW(s)$. Hence, $Q_c(J_\phi(r))$ is a weighted integral of the process $J_\phi(v)$ where the weight function depends upon the parameter c. If $c = 0$, we have $Q_0(J_\phi(r)) = \int_0^r J_\phi(v)dv$ and if $c = \phi = 0$ $Q_0(J_0(r)) = \int_0^r W(v)dv$. Using this notation, we characterize in the next Theorem the asymptotic distribution of $T(\hat{\alpha} - \alpha)$.

THEOREM 5 : *Let $\{y_t\}$ be a stochastic process generated by (6.1) and (6.2) with $\alpha = exp(c/T)$, and let the function $Q_c(J_\phi(r))$ be as defined in (6.4), then as $T \to \infty$:*

$$T(\hat{\alpha} - \alpha) \Rightarrow (1/2)\, Q_c(J_\phi(1))^2 \left\{\int_0^1 Q_c(J_\phi(r))^2 dr\right\}^{-1} - c\,.$$

There are several interesting features to note about Theorem 5. First, neither c nor ϕ is restricted to be negative; these variables can take any real value. Hence the result applies to many cases of interest besides those specified in the experiment of Section 3. In particular it can encompass a stationary process ($c, \phi < 0$), a process with an explosive and a stationary root (either c or γ is negative and the other is positive) , a process with two explosive roots ($c, \phi > 0$); or a process with two unit roots ($c = \phi = 0$). The latter case is of particular interest since it gives the properties of $T(\hat{\alpha} - 1)$ when the true process contains two unit roots. In that special case we have :

(6.5) $$T(\hat{\alpha} - 1) \Rightarrow (1/2)\left\{\int_0^1 W(r)dr\right\}^2 \left\{\int_0^1 \int_0^r W(s)dsdr\right\}^{-1}.$$

This latter result is interesting in view of the simulation experiment reported in Dickey and Pantula (1987). They showed that with a sample of length 50, the standard Dickey–Fuller (1979) test rejects the null hypothesis of a single unit root in favor of a stationary process slightly more than 5% of the time when the series actually has two unit roots. Given our result in (6.5) this feature is due to the small sample size used in the simulations. Indeed, in large samples, the Dickey–Fuller criterion would never reject a unit root in favor of a stationary process when two unit roots are present as the limiting distribution in (6.5) has a positive support.

It is worth emphasizing about Theorem 5 that, contrary to the case analyzed in the previous section with an MA(1) root local to – 1, $T(\hat{\alpha} - \alpha)$ has a non–degenerate asymptotic distribution. This explains the relatively small discrepancies between the usual asymptotic approximation and the exact distribution when ρ is close to one (as opposed to the large ones when the MA root is close to – 1). For ρ close to one, the fact that the exact distribution approaches its asymptotic counterpart quite slowly is explained by the difference between the local asymptotic distribution described above and the asymptotic distribution described by (1.4).

Theorem 5 can then be used to compute, for a given pair of values of c and ϕ, an alternative approximation to the exact distribution of $T(\hat{\alpha} - \alpha)$ when the errors have an autoregressive root close to one. We have not been able to derive, as in previous sections, a closed form solution for the limiting joint–moment generating function of $\{T^{-3}\Sigma_1^T y_{t-1}u_t, T^{-4}\Sigma_1^T y_{t-1}^2\}$ in this nearly twice integrated setting. We henceforth resort to simulations to tabulate the asymptotic distribution. The simulations are performed using the fact that $n^{-3/2}\Sigma_{k=1}^n \exp(c(1-k)/n)\, \Sigma_{j=0}^k \exp(\phi(k-j)/n)e_j \Rightarrow Q_c(J_\phi(1))$ as $n \to \infty$. Similarly $n^{-4}\Sigma_{t=1}^n\left\{\Sigma_{k=0}^t \exp(c(t-k)/n)\Sigma_{j=0}^k \exp(\phi(k-j)/n)e_j\right\}^2 \Rightarrow \int_0^1 Q_cJ_\phi(r)^2dr$ as $n \to \infty$. We specify $\{e_j\}$ as a sequence of i.i.d. N(0,1) variates and use n = 1000. Under these specifications we can expect the approximation to be quite accurate. We performed 5000 replications of the limiting distributions of $T(\hat{\alpha} - \alpha)$ and obtained the critical values from the sorted vector. See Chan (1988) for some evidence about the adequacy of such a method to calculate limiting distributions involving functionals of Wiener processes.

The results are presented in Tables VI, VII and VIII for the cases c = 0.0, – 5.0 and 2.0 respectively. We also include in these tables the corresponding results obtained by computation of the asymptotic expansion derived in Theorem 1. The results are quite interesting and merit some discussion. Consider first the behavior of the $O(T^{-1})$ asymptotic expansion. As mentioned previously, it is only in the case of AR(1) errors with positive correlation that it is likely to provide any improvement over the usual asymptotic distribution. This is indeed the case.

Consider first the case of a unit root process, c = 0.0 presented in Table VI. These results are to be compared to those in Table II.A. For $\rho = 0.5$, the asymptotic expansion provides a substantial improvement in the left tail of the distribution. Consider, for example, the first percentage point with T = 50. The exact value is – 4.47 ; the O(1) asymptotic value is –4.23 while the $O(T^{-1})$ value is –4.53. When ρ is closer to 1 , the asymptotic expansion still provides an improvement in the left tail of the distribution but not as significant. This is essentially due to the fact that the $O(T^{-1})$ expansion provides no adjustment to the mean of the distribution and as ρ approaches 1 the mean of the distribution decreases substantially as T increases. These observations are consistent with the fact that the $O(T^{-1})$ expansion also provides an adequate adjustment to the variance of the distribution when $\rho = 0.5$, but the adjustment deteriorates rapidly as ρ approaches 1. The results concerning the right tail of the distribution are quite different. Here, the $O(T^{-1})$ expansion provides no significant improvement over the usual O(1) asymptotic distribution for any value of ρ considered.

The "nearly twice integrated model" provides an interesting contrasting result. Here the left tail of the distribution is not well approximated for any value of ρ. This is due to the fact that when c = 0.0 , the local asymptotic distribution considered implies a non–negative variable in the limit. However, the left tail of the finite sample distribution is in the negative part. On the other hand, the right tail of the distribution is much better approximated by the local asymptotic distribution than by the usual O(1) distribution for a ρ value of 0.95 and to some extent 0.9. It provides, however, no improvement when ρ is 0.5. These facts are corroborated by the behavior of the mean and variance of the distribution. Both are better approximated by the local asymptotic distribution when $\rho = 0.95$ and to some extent when $\rho = 0.90$, but not when $\rho = 0.50$. Note that, while the O(1) asymptotic distribution understates the exact values of both the mean and the variance, the local asymptotic distribution overstates them.

Consider now the case where c = – 0.5. The results presented in Table VII are to be compared to the results in Table II.B. The case c = – 5.0 is quite different from the case where c = 0.0. Here the "nearly doubly integrated" local asymptotic distribution seems to provide a worse approximation than the usual O(1) asymptotic distribution for all values of ρ considered (even though the variance is better approximated). The results concerning the $O(T^{-1})$ expansion are similar to those when c = 0.0, namely an improvement in the left tail of the distribution especially when $\rho = 0.5$. The approximation for the variance is better when $\rho = 0.5$ but the improvement diminishes as ρ gets closer to one.

The case with c = 2.0 is presented in Table VIII (to be compared with Table II.C). The picture is again different. For $\rho = 0.5$, the $O(T^{-1})$ expansion provides a slight improvement for the 1% and 2.5 % points, but a worse approximation for the 5% and 10% points. The improvement in the right tail of the distribution and in the variance is marginal. When $\rho = 0.9$ or 0.95, there is very little improvement ; indeed there is little change in the $O(T^{-1})$ distribution as T changes. Consider now the "nearly twice integrated" asymptotic distribution. For $\rho = 0.50$, the approximation is again worse than the usual O(1) distribution. On the other hand, when $\rho = 0.90$ or 0.95, the improvement is substantial. The median and the right tail of the distribution are very well approximated. For example, when $\rho = 0.9$ and T = 25, the values for the 97.5% point are 2.42 (exact) and 2.41 (local asymptotic). Unlike the case with c = 0.0 or – 5.0 the left tail of the distribution is not as badly approximated especially for the 10% point.

In summary, the different types of asymptotic distributions considered for the case with positively correlated AR(1) errors appear to be complementary. None of them provides an approximation to the finite sample distribution that is satisfactory for all values of ρ and all percentage points. However, for a wide range of parameter configurations there is a particular asymptotic framework that seems appropriate. When ρ is small, say less than 0.5, the usual O(1) asymptotic performs quite well. When ρ is close to 0.5, the $O(T^{-1})$ expansion provides a substantial improvement in the left tail of the distribution (unless c = 2.0) but not in the right tail. This feature is not too troublesome, given that there are much less variations in the right tail of the distribution as T changes. When ρ approaches 1, the $O(T^{-1})$ expansion fails to provide much of an improvement. On the other hand, the "nearly twice integrated" model proposed here seems to provide a marked improvement in approximating the percentage points in the right tail of the distribution when ρ is close to one (especially when c = 0.0 or c = 2.0). The region where none of the asymptotic frameworks considered provide an adequate approximation for sample sizes in the range from 25 to 100 is in the left tail of the distribution when ρ is close to one.

7. A NEARLY INTEGRATED SEASONAL MODEL

In this Section, our aim is to provide a local asymptotic framework for the case where the errors have an autoregressive root near minus unity. We consider, on the one hand, the adequacy of such an asymptotic approximation, and we also investigate how the theoretical results can shed light on the differing behavior of $T(\hat{\alpha} - \alpha)$ when the errors are negatively correlated with either AR(1) or MA(1) structures. Consider first the following parameterization of the process under study :

$$y_t = \exp(c/T)y_{t-1} + u_t \, , \tag{7.1}$$

$$u_t = - \exp(\phi/T)u_{t-1} + e_t \, ; \tag{7.2}$$

where we again specify $e_t \sim$ i.i.d. $N(0,\sigma_e^2)$. The model (7.1) and (7.2) can be written as :

$$y_t = [\exp(c/T) - \exp(\phi/T)]y_{t-1} + \exp((c + \phi)/T)y_{t-2} + e_t \, . \tag{7.3}$$

As T increases to infinity $\{y_t\}$ approaches the process :

$$y_t = y_{t-2} + e_t \, . \tag{7.4}$$

The equation (7.4) characterizes a seasonal model of period 2 with a root on the unit circle. We therefore label the process (7.1) and (7.2) as a "nearly integrated seasonal model". To get some insights into the result presented below, consider a special case where $c = \phi$. Then (7.3) reduces to :

$$y_t = \exp(2c/T)y_{t-2} + e_t \, . \tag{7.5}$$

This is a special case of a class of nearly integrated seasonal models that have recently been studied by Chan (1989) and Perron (1990b). Chan (1989) derives the asymptotic distribution of $T(\hat{\alpha}_d - \alpha)$ where $\hat{\alpha}_d$ is the least–squares estimator of the coefficient on y_{t-2} in equation (7.5). Perron (1990b) tabulates the percentage points of this asymptotic distribution. The difference in focus here is that we wish to study the asymptotic distribution of the *first*–order autocorrelation coefficient when the process is a nearly integrated seasonal model of period 2.

Recall that $\hat{\alpha} = T^{-2}\Sigma_1^T y_t y_{t-1}/T^{-2}\Sigma_1^T y_{t-1}^2$. Under (7.5), it is easy to deduce from Chan (1989,Lemma 2.i) that :

$$(7.6) \qquad T^{-2}\Sigma_1^T y_{t-1}^2 \Rightarrow (\sigma_e^2/4)\Sigma_{i=1}^2 \int_0^1 [J_{c,i}(r)]^2 \, dr \, ;$$

where $J_{c,i}(r) = \int_0^r \exp((r-s)c)dW_i(s)$, i= 1,2 ; and $W_1(r)$ and $W_2(r)$ and independent Wiener processes. Consider now the numerator of $\hat{\alpha}$. First note that we can write :

$$(7.7) \qquad y_t = \Sigma_{j=0}^{[t/2]} \exp(2cj/T)e_{t-2j} \quad ;$$

and

$$(7.8) \qquad y_{t-1} = \Sigma_{j=0}^{[(t-1)/2]} \exp(2cj/T)e_{t-2j-1} \quad ;$$

where [ ] denotes the integer part of the number. Given that the errors $\{e_t\}$ are i.i.d. , y_t and y_{t-1} are independent processes as they are functions of different subsets of the sequence $\{e_t\}$. Hence $y_t y_{t-1}$ is the product of two independent nearly integrated random processes having [t/2] and [(t–1)/2] elements respectively. Following the results on the sums of products of two independent random walks, it is straightforward to show that :

$$(7.9) \qquad T^{-2}\Sigma_1^T y_t y_{t-1} \Rightarrow (\sigma_e^2/2) \int_0^1 J_{c,1}(r)J_{c,2}(r)dr \, .$$

Hence we have the following asymptotic result when $c = \phi$:

$$(7.10) \qquad \hat{\alpha} \Rightarrow \left\{2\int_0^1 J_{c,1}(r)J_{c,2}(r)dr\right\}\left\{\Sigma_{i=1}^2 \int_0^1 [J_{c,i}(r)]^2 \, dr\right\}^{-1} .$$

Note that $\hat{\alpha}$ has a degenerate asymptotic distribution , in the sense that it converges to a random variable instead of a fixed constant as was the case with MA(1) errors with a root approaching – 1. Our result is consistent with that of Yajima (1985) who showed, among other things, that in seasonally integrated models of period k, the sample autocorrelations of order other than kn (for any integer n) have a degenerate asymptotic distribution in the sense that they converge to random variables instead of fixed constants.

Given (7.10), $T(\hat{\alpha} - \alpha)$ is unbounded as T increases, which explains the large discrepancies between the exact and asymptotic distributions reported in Section 3. Note, however, the different rate at which the root is permitted to approach – 1 as T increases to infinity. In the AR(1) case it does so at rate T, while in the MA(1) case the rate is $T^{1/2}$. This feature explains well the differences in the discrepancies between the finite sample and asymptotic distributions reported in Section 3. As was discussed, the discrepancies are much larger in the MA(1) case for an equal value of θ and ρ. Such a feature can be theoretically interpreted by noting that the local asymptotic distribution is approached faster in the MA case than in the AR case. Indeed, in the MA case, we have $\delta = T^{1/2}(1 + \theta)$ and, in the AR case, $\phi = T(1 + \rho)$. For a given same value for θ and ρ, ϕ is further away than δ from the zero boundary because of the different normalizing power on T.

Note also that (7.10) does not presume that c is negative ; it can also accommodate locally explosive processes as well as a seasonal random walk. In the latter case we have an interesting result, namely the asymptotic behavior of the first–order autocorrelation coefficient when the true model is a seasonal random walk of period 2. In this case :

$$\hat{\alpha} \Rightarrow \left\{2\int_0^1 W_1(r)W_2(r)dr\right\}\left\{\Sigma_{i=1}^2 \int_0^1 W_i(r)^2dr\right\}^{-1}, \tag{7.11}$$

where $W_1(r)$ and $W_2(r)$ are independent Wiener processes.

The general case where $c \neq \phi$ is more complex but yields qualitatively similar results. The following Theorem, proved in the Appendix, provides the formal asymptotic distribution.

THEOREM 6 : *Let $\{y_t\}$ be a stochastic process generated by (7.1) and (7.2) with $\alpha = exp(c/T)$. Define the random functions $J_{\phi,i}(s) = \int_0^s exp((s - v)\phi)dW_i(v)$ and $Q_c(J_{\phi,i}(r)) = \int_0^r exp((r - s)c)\ J_{\phi,i}(s)ds$, with $i = 1,2$; where $W_1(r)$ and $W_2(r)$ are independent Wiener processes. Also let $J_{c,1}(s) = \int_0^s exp((s - v)c)dW_1(v)$. Then as $T \to \infty$:*

$$\hat{\alpha} \Rightarrow 1 - 2\int_0^1 B(r)^2 dr \left\{\int_0^1 \{[A(r) - B(r)]^2 + B(r)^2\} dr\right\}^{-1} ;$$

where $A(r) = (\phi - c)[Q_c(J_{\phi,1}(r)) - Q_c(J_{\phi,2}(r))] + 2J_{c,1}(r)$,

and $B(r) = J_{\phi,1}(r) - J_{\phi,2}(r)$.

Note that the result in Theorem 6 reduces to (7.10) when $c = \phi$. Note also that again , in this general context, c and ϕ can take any real value. This result shows that $\hat{\alpha}$ has a degenerate asymptotic distribution even in the general case where $c \neq \phi$. Hence $\hat{\alpha}$ is not a consistent estimator of α and $T(\hat{\alpha} - \alpha)$ is unbounded as T increases. However unlike the MA(1) case and similar to the AR(1) case with a positive root on the unit circle, the asymptotic distribution in Theorem 6 has a non–zero variance even on the boundary $\phi = 0$.

As in Section 6, we were not able to find an explicit solution for the limiting joint moment–generating function of $\{T^{-2}\Sigma_1^T y_t y_{t-1}, T^{-2}\Sigma_1^T y_{t-1}^2\}$. Therefore to study how well this type of asymptotic distribution approximates the finite sample distribution, we resort to simulation experiments [3] .

Table IX presents the distribution of $T(\hat{\alpha} - \alpha)$ based upon the local asymptotic framework described in Theorem 6 for the case where $\rho = -0.9$ [4]. The cases considered are again c = 0.0, – 5.0 and 2.0 with T = 25, 50, 100 and 500. The numerical values presented in this Table are to be compared to those in the top portion of Tables II (A, B and C). In general the approximation is satisfactory and certainly represent a major improvement over the standard O(1) asymptotic distribution. The approximation is best, and indeed very good when c = – 5.0 (most notably with T = 50 and 100). It deteriorates as c increases. Also, for a fixed value of c, the approximation is better when T is small ; it deteriorates as T increases to 500. This last feature is to be expected given that our asymptotic framework is local to the boundary $\rho = -1$; when T increases the noncentrality parameter ϕ is correspondingly higher for a fixed value of c. Finally, it is to be noted that the approximation is better in the left tail of the distribution.

CONCLUDING COMMENTS

Our paper first characterized and tabulated the asymptotic distribution of the normalized least–squares estimator in a nearly integrated first–order autoregressive process allowing for dependence in the error structure with special emphasis on the MA(1) and AR(1) cases. These two cases were sufficient to provide a rich array of cases with interestingly different patterns. Special attention was given to analyzing the adequacy of this standard asymptotic distribution as an approximation to the finite sample distribution of the statistic. Our results showed that, in a substantial part of the parameter space, the approximation is seriously inadequate. This result points to the inherent danger associated with an asymptotic framework that allows very general conditions in the underlying data generating mechanism, in particular with respect to the amount and type of dependence permitted. There indeed appears to be a substantial tradeoff between the generality in the conditions allowed and the amount of data needed for the asymptotic distribution to provide a reasonable guide to the finite sample distribution.

While this conclusion is issued from an admitably simple model, it should carry over to more general ones. Indeed the same analysis could be performed with additional regressors added such as a constant or a time trend. The analytical derivations concerning the limiting distribution would be different and more complex but the same qualitative results should hold. They are also expected to hold in more general models of the type analyzed by White (1984), for example. The same tradeoff would result between generality and the adequacy of the approximation for sample sizes usually available in economics.

An important sense in which our analysis is a preliminary step in a more complete analysis of the adequacy of asymptotic distributions in time series models allowing general dependence and heterogeneity is that, in practice, the statistics are used along with a correction factor that asymptotically eliminates the dependence of the asymptotic distribution upon nuisance parameters. Examples of the use of such corrections include the class of unit root tests proposed by Phillips (1987) and Phillips and Perron (1988) as well as those involving the Newey–West (1987) covariance matrix estimator in more general structural models. Nevertheless, suppose that the finite sample distribution of such a correction factor was an adequate approximation to the required asymptotic correction necessary to eliminate the dependence upon the nuisance parameters. Our results would still imply an inadequate corrected statistic as the asymptotic distribution of the uncorrected part is far from the finite sample distribution in an important range of the parameter space. In a sense, the finite sample distribution would, if at all, bear an adequate correspondence to the asymptotic distribution by fortuitous cancellation of approximation errors for the distribution of the original statistic and the correction factor. Such a situation is unlikely to occur as demonstrated for the case of unit root tests in the simulation study of Schwert (1989).

The second contribution of this study was to present alternative frameworks that could provide a better approximation to the finite sample distribution of the statistic of interest. Here our results are encouraging in that our local asymptotic distributions provide a substantial improvement in approximating the finite sample distribution in the region of the parameter space where the traditional asymptotic framework provides severe inaccuracies. These local asymptotic distributions still depend upon nuisance parameters, namely those indexing the extent of correlation in the residuals. In practice one would need to have an estimate of these parameters in order to use our distributional results. These could be obtained by a preliminary investigation of the nature of the correlation structure of the residuals. Consider, for example, the case of testing for a unit root. A preliminary estimate of the correlation structure under the null hypothesis can be obtained by analyzing the sample correlation of the first–difference of the data. Suppose, for illustration, that a large negative MA(1) component is estimated. The test can then be carried using the local asymptotic distribution described in Section 5 with δ chosen according to the estimated value of the MA parameter θ. Of course, similar procedures can be followed in the case where the residuals have an autoregressive structure.

On a theoretical side, our study shows how different asymptotic frameworks can be complementary in several respects. First, each framework provides a better approximation to the finite sample distribution where the other shows great inaccuracies. Secondly, the asymptotic results in the local asymptotic frameworks were shown to be useful in explaining why and when the usual asymptotic theory may fail. Nevertheless, our results also show the need for a unified asymptotic theory that could provide a sensible guide to the finite sample distribution over most of the relevant parameter space, albeit with possibly the need to estimate nuisance parameters. Such a topic is of interest for future research.

FOOTNOTES

1 A much wider range of experiments were performed. For the sake of brevity we report only a subset of the results. Some comments made in the text pertain to the full set of results, however. These are available upon request.

2 Some slight exceptions to this rule occur for large values of T because of sampling variability induced by the simulations.

3 The procedure for simulating the asymptotic distribution stated in Theorem 6 is similar to that outlined for the limiting distribution analyzed in Section 6.

4 We also performed a similar comparison with $\rho = -0.95$. The qualitative features being the same, these results are not reported but are available upon request. They showed a marginally better approximation.

MATHEMATICAL APPENDIX

Proof of Theorem 1 : The proof of Theorem 1 relies on Theorem 3.1 of Phillips (1987c) which shows that under the conditions of Assumptions 1 and 2 we have the following expansion for the partial sums $X_T(r)$:

$$X_T(r) \overset{d}{=} W(r) + O_p(T^{-1}) . \tag{A.1}$$

where $\overset{d}{=}$ signifies equality in distribution. Using (A.1) we can prove the following Lemma related to the sample moments of $\{y_t\}$.

LEMMA A.1 : *Let $\{y_t\}$ be generated by (1.1) and (1.3) and let the innovation sequence $\{u_t\}$ satisfy the conditions of Assumptions 1 and 2, then :*

a) $T^{-1/2}y_{[Tr]} \overset{d}{=} \sigma J_c(r) + T^{-1/2}exp(cr)y_0 + O_p(T^{-1})$;

b) $T^{-3/2}\Sigma_{t=1}^{T} y_t \overset{d}{=} \sigma \int_0^1 J_c(r)dr + T^{-1/2}y_0(exp(c) - 1)/c + O_p(T^{-1})$;

c) $T^{-2}\Sigma_{t=1}^{T} y_t^2 \overset{d}{=} \sigma^2 \int_0^1 J_c(r)^2 dr + 2T^{-1/2}\sigma y_0 \int_0^1 exp(cr)J_c(r)dr + O_p(T^{-1})$;

d) $T^{-1}\Sigma_{t=1}^{T} y_{t-1}u_t \overset{d}{=} \sigma^2 \int_0^1 J_c(r)dW(r) + (\sigma^2 - \sigma_u^2) + T^{-1/2}\sigma y_0 \int_0^1 exp(cr)dW(r)$

$$- T^{-1/2}(\nu/2\sigma^2)\eta + O_p(T^{-1}) .$$

Proof : The proof of Lemma (1) follows closely the proof of Lemma 4.2 of Phillips (1987c). Using (1.1) and (1.3) we can write :

$$T^{-1/2}y_{[Tr]} = T^{-1/2}\Sigma_{j=1}^{[Tr]} \exp(([Tr] - j)c)/T)\, u_j + T^{-1/2}\exp([Tr]c/T)y_0$$

$$= \sigma\left\{X_T(r) + c\int_0^r \exp((r-s)c)X_T(s)ds\right\} + T^{-1/2}\exp([Tr]c/T)y_0.$$

Using (A.1), we deduce that :

$$T^{-1/2}y_{[Tr]} \overset{d}{=} \sigma\left\{W(r) + c\int_0^r \exp((r-s)c)W(s)ds\right\} + T^{-1/2}\exp(cr)y_0 + O_p(T^{-1})$$

$$= \sigma J_c(r) + T^{-1/2}\exp(cr)y_0 + O_p(T^{-1})\ ;$$

using the fact that $J_c(r) \equiv W(r) + c\int_0^r \exp((r-s)c)W(s)ds$. The proof of parts (b) and (c) are analogous and omitted. To prove part (d) note that squaring (1.1), summing over t and rearranging we obtain :

$$T^{-1}\Sigma_{t=1}^T y_{t-1}u_t =$$

$$(1/2)\exp(-2c/T)\left\{T^{-1}y_T^2 - T^{-1}y_0^2 - T(\exp(2c/T)-1)T^{-2}\Sigma_{t=1}^T y_{t-1}^2 - T^{-1}\Sigma_{t=1}^T u_t^2\right\}$$

Note that :

$$T^{-1}\Sigma_{t=1}^T u_t^2 = T^{-1/2}[T^{-1/2}\Sigma_{t=1}^T(u_t^2 - \sigma_u^2)] + \sigma_u^2 \overset{d}{=} T^{-1/2}\zeta + \sigma_u^2 + O_p(T^{-1/2})\ ;$$

where $\zeta \sim N(0,\nu^2)$ (see Phillips (1987c), Lemma 4.2). Hence using parts (a), (b) and (c) we have :

$$T^{-1}\Sigma_{t=1}^T y_{t-1}u_t \overset{d}{=} (1/2)\left\{[\sigma J_c(1) + T^{-1/2}\exp(c)y_0]^2\right.$$

$$- 2c\ \{\ \sigma^2\int_0^1 J_c(r)^2dr + 2T^{-1/2}\sigma y_0\int_0^1 \exp(cr)J_c(r)dr\}$$

$$\left.- \sigma_u^2 - T^{-1/2}\zeta\right\} + O_p(T^{-1})\ ;$$

$$\overset{d}{=} (\sigma^2/2)\left\{[J_c(1)^2 - 2c\int_0^1 J_c(r)^2dr - 1] + (\sigma^2 - \sigma_u^2)/2\right.$$

$$\left.+ T^{-1/2}[\sigma\exp(c)y_0J_c(1) + 2\sigma cy_0\int_0^1 \exp(cr)J_c(r)dr - \zeta/2]\right\} + O_p(T^{-1})\ .$$

The result follows by noting that $(1/2)[J_c(1)^2 - 2c\int_0^1 J_c(r)^2dr - 1] \overset{d}{=} \int_0^1 J_c(r)dW(r)$ and that $\exp(c)J_c(1) + 2c\int_0^1 \exp(cr)J_c(r)dr \overset{d}{=} \int_0^1 \exp(cr)dW(r)$ (see Perron (1988a)); and using the fact that $\nu\eta \overset{d}{=} \zeta$. The proof of Theorem 1 follows using (1.2) and parts (c) and (d) of Lemma A.1. □

Proof of Theorem 3 : Assuming , for simplicity that $y_0 = e_0 = 0$, the process y_t can be written as $y_t = \Sigma_{j=1}^{t} \exp((t - j)c/T)\, u_j$. Given that $u_t = e_t - e_{t-1} + \gamma T^{-1/2} e_{t-1}$, simple manipulations show that :

$$y_t = (1 - \gamma T^{-1/2} \exp(-c/T))\, e_t + (1 - \exp(-c/T)(1 + \gamma T^{-1/2}))\, X_t \ ,$$

where $X_t = \Sigma_{j=1}^{t} \exp((t-j)c/T)\, e_j$ is a near–integrated process given by

$$X_t = \exp(c/T)\, X_{t-1} + e_t \ ;$$

with $e_t \sim$ i.i.d. $(0,\sigma_e^2)$ and $X_0 = 0$. Let :

$$a_T = 1 - \gamma T^{-1/2} \exp(-c/T) \ ; \tag{A.2}$$

$$b_T = 1 - \exp(-c/T)(1 - \gamma T^{-1/2}) \ ; \tag{A.3}$$

and note that $a_T \to 1$ and $T^{1/2} b_T \to \gamma$ as $T \to \infty$. Consider first the second sample moment of y_t. We have :

$$T^{-1} \Sigma_1^T y_t^2 = T^{-1} \Sigma_1^T (a_T e_t + b_T X_t)^2$$

$$= a_T^2 T^{-1} \Sigma_1^T e_t^2 + T b_T^2 T^{-2} \Sigma_1^T X_t^2 + 2 a_T T^{1/2} b_T T^{-3/2} \Sigma_1^T X_t e_t \ .$$

We have $T^{-1} \Sigma_1^T e_t^2 \to \sigma_e^2$ (in probability) and $T^{-2} \Sigma_1^T X_t^2 \Rightarrow \sigma_e^2 \int_0^1 J_c(r)^2 dr$ as $T \to \infty$. Furthermore, in a manner similar to Theorem 2.4 of Chan and Wei (1988), it can be shown that $\Sigma_1^T X_t e_t = O_p(T)$. Hence :

$$T^{-1} \Sigma_1^T y_t^2 \Rightarrow \sigma_e^2 + \sigma_e^2\, \gamma^2 \int_0^1 J_c(r)^2 dr \ . \tag{A.4}$$

Consider now the sum $T^{-1}\Sigma_1^T y_{t-1}u_t$. Using (5.1) and (5.2) we can write :

$$T^{-1}\Sigma_1^T y_{t-1}u_t = a_T T^{-1}\Sigma_1^T e_t e_{t-1} - a_T(1-\gamma T^{-1/2})\, T^{-1}\Sigma_1^T e_{t-1}^2$$

$$+ b_T T^{-1}\Sigma_1^T X_{t-1}e_t - (1-\gamma T^{-1/2})b_T\, T^{-1}\Sigma_1^T X_{t-1}e_{t-1}\,,$$

where a_T and b_T are defined in (A.2) and (A.3). We can show that $\Sigma_1^T e_t e_{t-1}$ is $O_p(T^{1/2})$, $\Sigma_1^T X_{t-1}e_t$ and $\Sigma_1^T X_{t-1}e_{t-1}$ are $O_p(T)$ and that $T^{-1}\Sigma_1^T e_t^2 \to \sigma_e^2$ (in probability) as $T \to \infty$. Using these results and the fact that $a_T \to 1$ and $b_T \to 0$ as $T \to \infty$, we obtain :

$$\text{(A.5)} \qquad T^{-1}\Sigma_1^T y_{t-1}u_t \to \sigma_e^2\,.$$

This proves Theorem 3 using $\hat{\alpha} - \alpha = T^{-1}\Sigma_1^T y_{t-1}u_t / T^{-1}\Sigma_1^T y_{t-1}^2$ with (A.4) and (A.5) and the fact that $\alpha = \exp(c/T) \to 1$ as $T \to \infty$. □

Proof of Theorem 5 : We first prove the following Lemma concerning the sample moments of $\{y_t\}$ under the nearly twice integrated framework of Section 6.

LEMMA A.2 : *Suppose that $\{y_t\}$ is a sequence of random variables generated according to (6.1) and (6.2) , then as $T \to \infty$:*

a) $\quad T^{-3/2}y_{[Tr]} \Rightarrow \sigma_e\, Q_c(J_\phi(r))\,;$

b) $\quad T^{-4}\Sigma_1^T y_t^2 \Rightarrow \sigma_e^2 \int_0^1 Q_c(J_\phi(r))^2 dr\,;$

c) $\quad T^{-3}\Sigma_1^T y_{t-1}u_t \Rightarrow (\sigma_e^2/2)\left\{Q_c(J_\phi(1))^2 - 2c\int_0^1 Q_c(J_\phi(r))^2 dr\right\}.$

Proof : We first define the random process $X_T^*(r)$ as :

$$X_T^*(r) = \sigma_e^{-1}T^{-1/2}S_{[Tr]}^* = \sigma_e^{-1}T^{-1/2}S_{j-1}^* \qquad (j-1)/T \le r < j/T$$

$$(j = 1,\dots,T)$$

$$X_T^*(1) = \sigma_e^{-1} T^{-1/2} S_T^* ;$$

where $S_j^* = \Sigma_{t=1}^{j} e_t$. Since $e_t \sim$ i.i.d. $(0,\sigma_e^2)$ we have $X_T^*(r) \Rightarrow W(r)$, the unit Wiener process. To prove part (a), we assume, for simplicity, that $y_0 = e_0 = 0$, and using (6.1) and (6.2) we have $(1 - \exp(\phi/T)L)(1 - \exp(c/T)L)y_t = e_t$, where L is the lag operator. We can therefore write y_t as :

$$y_t = \Sigma_{k=0}^{t} \exp(c(t-k)/T) \Sigma_{j=0}^{k} \exp(\phi(k-j)/T)\, e_j .$$

Then :

$$T^{-3/2} y_{[Tr]} = T^{-3/2} \Sigma_{k=0}^{[Tr]} \exp(c([Tr]-k)/T) \Sigma_{j=0}^{k} \exp(\phi(k-j)/T)\, e_j$$

$$= \Sigma_{k=0}^{[Tr]} \int_{(k-1)/T}^{k/T} \exp(c([Tr]-k)/T) \left\{ \sigma_e \Sigma_{j=0}^{k} \int_{(j-1)/T}^{j/T} \exp(\phi(k-j)/T) dX_T^*(s) \right\} dv$$

$$= \sigma_e \int_0^r \exp(c(r-v)) \int_0^v \exp(\phi(v-s)) dX_T^*(s) dv$$

$$= \sigma_e \int_0^r \exp(c(r-v)) \left\{ X_T^*(v) + \phi \int_0^v \exp(\phi(v-s)) X_T^*(s) ds \right\} dv$$

$$\Rightarrow \sigma_e \int_0^r \exp(c(r-v)) \left\{ W(v) + \phi \int_0^v \exp(\phi(v-s)) W(s) ds \right\} dv$$

$$\equiv \sigma_e \int_0^r \exp(c(r-v)) J_\phi(v) dv \equiv Q_c(J_\phi(r)).$$

This proves part (a). To prove part (b), we have :

$$T^{-4} \Sigma_1^T y_t^2 = T^{-4} \Sigma_{t=1}^{T} \left\{ \Sigma_{k=0}^{t} \exp(c(t-k)/T) \Sigma_{j=0}^{k} \exp(\phi(k-j)/T)\, e_j \right\}^2$$

$$= T^{-1} \Sigma_{t=1}^{T} \left\{ \sigma_e \int_0^{t/T} \exp(c(t/T - v)) \int_0^v \exp(\phi(v-s)) dX_T^*(s) dv \right\}^2$$

$$= \sigma_e^2 \int_0^1 \left\{ \int_0^r \exp(c(r-v)) \int_0^v \exp(\phi(v-s)) dX_T^*(s) dv \right\}^2 dr$$

$$\Rightarrow \sigma_e^2 \int_0^1 Q_c(J_\phi(r))^2 dr \; ;$$

using arguments similar to those of part (a). To prove part (c), note that squaring (6.1), summing over t and rearranging, we have :

(A.6) $$T^{-3}\Sigma_1^T y_{t-1}u_t =$$

$$(1/2)\exp(-2c/T)\left\{T^{-3}y_T^2 - T(\exp(2c/T)-1)T^{-4}\Sigma_1^T y_{t-1}^2 - T^{-3}\Sigma_1^T u_t^2\right\}.$$

Note that, from parts (a) and (b), $T^{-3}y_T^2 \Rightarrow \sigma_e^2\; Q_c(J_\phi(1))^2$ and $T^{-4}\Sigma_1^T\; y_{t-1}^2 \Rightarrow \sigma_e^2 \int_0^1 Q_c(J_\phi(r))^2 dr$. We also have $T^{-2}\Sigma_1^T\; u_t^2 \Rightarrow \sigma_e^2 \int_0^1 J_\phi(r)^2 dr$ given that $\{u_t\}$ is a nearly integrated process with non–centrality parameter ϕ, hence $T^{-3}\Sigma_1^T\; u_t^2 \to 0$ (in probability). Taking the limit of (A.6) and noting that $T(\exp(2c/T) - 1) \to 2c$, we have :

$$T^{-3}\Sigma_1^T\; y_{t-1}u_t \Rightarrow (\sigma_e^2/2)\left\{Q_c(J_\phi(1))^2 - 2c\int_0^1 Q_c(J_\phi(r))^2 dr\right\}.$$

To prove Theorem 5, simply note that $T(\hat{\alpha} - \alpha) = T^{-3}\Sigma_1^T\; y_{t-1}u_t/T^{-4}\Sigma_1^T\; y_{t-1}^2$, hence :

$$T(\hat{\alpha} - \alpha) \Rightarrow (1/2)\; Q_c(J_\phi(1))^2\left\{\int_0^1 Q_c(J_\phi(r))^2 dr\right\}^{-1} - c\;. \;\Box$$

Proof of Theorem 6 : To prove Theorem 6, we proceed with a series of Lemmas concerning various sample moments of the data. For ease of notation, assume without loss of generality, that the sample size T is an even number and let $m = T/2$. Also let $\alpha = \exp(c/T)$ in (7.1) and $\rho = \exp(\phi/T)$ in (7.2). The first Lemma is concerned with the asymptotic distribution of sample moments involving different subsets of the data, i.e. separating the sequence $\{y_t\}$ and $\{u_t\}$ into two subsets corresponding to whether the time index t is even or odd.

LEMMA A.3 : *Let the functions $A(r)$ and $B(r)$ be as defined in Theorem 6 and consider a sequence of random variables $\{y_t\}$ defined by (7.1) and (7.2). Then as $T \to \infty$:*

a) For $[Tr]$ an even number : $T^{-1/2}y_{[Tr]} \Rightarrow 2^{-3/2}A(r)$;

b) $T^{-2}\Sigma_{k=1}^{m} u_{2k-1}^2 \Rightarrow (\sigma_e^2/4) \int_0^1 B(r)^2 dr$;

c) $T^{-2}\Sigma_{t=1}^{T} u_t^2 \Rightarrow (\sigma_e^2/2) \int_0^1 B(r)^2 dr$;

d) $T^{-2}\Sigma_{k=1}^{m} y_{2k-2}^2 \Rightarrow (\sigma_e^2/16) \int_0^1 A(r)^2 dr$;

e) $T^{-2}\Sigma_{k=1}^{m} y_{2k-2} u_{2k-1} \Rightarrow -(\sigma_e^2/8) \int_0^1 A(r)B(r) dr$.

Proof : To prove part (a), note that from (7.1) : $y_t = \Sigma_{j=1}^{t} \alpha^{t-j} u_j$. Hence, for t an even number, we have :

$$y_{2k} = \Sigma_{j=1}^{2k} \alpha^{2k-j} u_j \qquad (k = 1,\ldots,m).$$

Separating the sequence $\{u_j\}$ according to whether j is even or odd we have :

(A.7) $$y_{2k} = \Sigma_{j=1}^{k} \alpha^{2k-2j} u_{2j} + \Sigma_{j=1}^{k} \alpha^{2k-2j+1} u_{2j-1} .$$

Now define the following variables :

(A.8) $$X_{1,k} = \Sigma_{j=1}^{k} (\rho^2)^{k-j} e_{2j} ,$$

and

(A.9) $$X_{2,k} = \Sigma_{j=1}^{k} (\rho^2)^{k-j} e_{2j-1} .$$

Note that $X_{1,k}$ and $X_{2,k}$ are independent nearly integrated random processes with noncentrality parameter ϕ given that $\rho^2 = \exp(2\phi/T) = \exp(\phi/m)$ and that the random sequences $\{e_{2j}\}_{j=1}^{m}$ and $\{e_{2j-1}\}_{j=1}^{m}$ are independent by assumption (since the innovation sequence $\{e_t\}_{t=1}^{T}$ is i.i.d.). It is straightforward to show that :

(A.10) $$u_{2k} = X_{1,k} - \rho X_{2,k} ,$$

and

(A.11) $$u_{2k-1} = X_{2,k} - (1/\rho)X_{1,k} + (1/\rho)e_{2k} .$$

Using (A.7) through (A.11) we deduce that :

$$T^{-1/2}y_{2k} = T(1-\alpha/\rho)T^{-3/2}\Sigma_{j=1}^{k}(\alpha^2)^{k-j}X_{1,j}$$

(A.12) $$+ T(\alpha-\rho)T^{-3/2}\Sigma_{j=1}^{k}(\alpha^2)^{k-j}X_{2,j} + (\alpha/\rho)T^{-1/2}\Sigma_{j=1}^{k}(\alpha^2)^{k-j}e_{2j} .$$

Noting that $\alpha^2 = \exp(2c/T) = \exp(c/m)$ and $\rho^2 = \exp(2\phi/T) = \exp(\phi/m)$, using standard limiting arguments, we have (see the proof of Lemma A.2 (a)) :

$$m^{-3/2}\Sigma_{j=1}^{k}(\alpha^2)^{k-j}X_{1,j} \Rightarrow \sigma_e\int_0^r \exp(c(r-s))\int_0^s \exp(\phi(s-v))dW_1(v)ds \equiv Q_c(J_{\phi,1}(r))$$

and, similarly,

$$m^{-3/2}\Sigma_{j=1}^{k}(\alpha^2)^{k-j}X_{2,j} \Rightarrow \sigma_e\int_0^r \exp(c(r-s))\int_0^s \exp(\phi(s-v))dW_2(v)ds \equiv Q_c(J_{\phi,2}(r)) ,$$

with $W_1(v)$ and $W_2(v)$ independent Wiener processes. Given that $m^{-1/2}\ \Sigma_{j=1}^{k}(\alpha^2)^{k-j}e_{2j} \Rightarrow \int_0^r \exp(\phi(r-v))dW_1(v) \equiv J_{c,1}(r)$, $T(1-\alpha/\rho) \to (\phi - c)$, $T(\alpha-\rho) \to (c-\phi)$, $\alpha \to 1$ and $\rho \to 1$ as $T \to \infty$, we obtain from (A.12) :

$$T^{-1/2}y_{2k} \Rightarrow 2^{-3/2}(\phi-c)Q_c(J_{\phi,1}(r)) + 2^{-3/2}(c-\phi)Q_c(J_{\phi,2}(r)) + 2^{-1/2}J_{c,1}(r)$$

$$\equiv 2^{-3/2}A(r) .$$

To prove part (b), first note that from (7.2) :

$$u_{2k-1} = \Sigma_{j=1}^{2k-1}\ (-\rho)^{2k-j-1}e_j .$$

Separating this sum into ones that involve even and odd values of j we have :

$$u_{2k-1} = \Sigma_{j=1}^{k}(\rho^2)^{k-j}e_{2j-1} - (1/\rho)\Sigma_{j=1}^{k-1}(\rho^2)^{k-j}e_{2j}$$

(A.13) $$= X_{2,k} - (1/\rho)X_{1,k} + (1/\rho)e_{2k}$$

$$= X_{2,k} - X_{1,k} + o_p(T^{1/2}) ,$$

given that $\rho \to 1$ as $T \to \infty$. Hence, we have :

$$T^{-2}\Sigma_{k=1}^{m}u_{2k-1}^2 = (1/4)m^{-2}\Sigma_{k=1}^{m}u_{2k-1}^2 = (1/4)m^{-2}\Sigma_{k=1}^{m}(X_{2,k} - X_{1,k} + o_p(T^{1/2}))^2$$

$$\Rightarrow (\sigma_e^2/4)\int_0^1[J_{\phi,2}(r) - J_{\phi,1}(r)]^2dr \equiv (\sigma_e^2/4)\int_0^1 B(r)^2dr .$$

To prove part (c), note that

$$\Sigma_{t=1}^{T}u_t^2 = \Sigma_{k=1}^{m}u_{2k}^2 + \Sigma_{k=1}^{m}u_{2k-1}^2 .$$

In a manner similar to part (b), it is easy to show that $u_{2k} = X_{1,k} - \rho X_{2,k}$. Hence :

$$\Sigma_{k=1}^{m}u_{2k}^2 = \Sigma_{k=1}^{m}[X_{1,k} - \rho X_{2,k}]^2.$$

Using (A.13) we have :

(A.14) $$\Sigma_{t=1}^{T}u_t^2 = (1 + 1/\rho^2)\Sigma_{k=1}^{m}X_{1,k}^2 + (\rho^2 + 1)\Sigma_{k=1}^{m}X_{2,k}^2 - 2(\rho + 1/\rho)\Sigma_{k=1}^{m}X_{1,k}X_{2,k}$$

$$+ (1/\rho^2)\Sigma_{k=1}^{m}e_{2k}^2 + (2/\rho)\Sigma_{k=1}^{m}X_{2,k}e_{2k} - (2/\rho^2)\Sigma_{k=1}^{m}X_{1,k}e_{2k} .$$

It is easy to verify that the last three terms in (A.14) are $O_p(T)$. Hence using standard convergence arguments and the fact that $\rho \to 1$ as $T \to \infty$:

$$T^{-2}\Sigma_{t=1}^{T}u_t^2 = (1/2)m^{-2}\Sigma_{k=1}^{m}[X_{1,k} - X_{2,k}]^2 + o_p(1)$$

$$\Rightarrow (\sigma_e^2/2) \int_0^1 [J_{\phi,1}(r) - J_{\phi,2}(r)]^2 dr \equiv (\sigma_e^2/2) \int_0^1 B(r)^2 dr .$$

The proof of part (d) follows straightforwardly from part (a) using the fact that $T^{-2}\Sigma_{k=1}^m y_{2k-2}^2 = (1/4)m^{-2}\Sigma_{k=1}^m y_{2k-2}^2$ and that $m^{-1/2}y_{2k} \Rightarrow (1/2)A(r)$. To prove part (e) note that (using $y_0 = 0$) :

$$T^{-2}\Sigma_{k=1}^m y_{2k-2}u_{2k-1} = T^{-2}\Sigma_{k=1}^{m-1} y_{2k}u_{2k+1} = T^{-2}\Sigma_{k=1}^m y_{2k}u_{2k+1} - T^{-2}y_T u_{T+1}$$

$$= T^{-2}\Sigma_{k=1}^m y_{2k}(-\rho u_{2k} + e_{2k+1}) - T^{-2}y_T u_{T+1}$$

$$\text{(A.15)} \qquad = -T^{-2}\Sigma_{k=1}^m y_{2k}u_{2k} + o_p(1) ,$$

given that $\rho \to 1$ as $T \to \infty$ and that both $T^{-2}\Sigma_{k=1}^m y_{2k}e_{2k+1}$ and $T^{-2}y_T u_{T+1}$ are $o_p(1)$. Now using (A.12) and the fact that $u_{2k} = X_{1,k} - \rho X_{2,k}$ we have :

$$T^{-2}\Sigma_{k=1}^m y_{2k}u_{2k} = (1/4)\, m^{-2}\Sigma_{k=1}^m [X_{1,k} - \rho X_{2,k}][(1 - \alpha/\rho)\, \Sigma_{j=1}^k (\alpha^2)^{k-j} X_{1,j}$$

$$+ (\alpha - \rho)\Sigma_{j=1}^k (\alpha^2)^{k-j} X_{2,j} + (\alpha/\rho)\Sigma_{j=1}^k (\alpha^2)^{k-j} e_{2j}]$$

$$\Rightarrow (\sigma_e^2/8) \int_0^1 \Big\{(\phi - c)[Q_c(J_{\phi,1}(r)) - Q_c(J_{\phi,2}(r))] + 2J_{c,1}(r)\Big\}$$

$$\Big\{J_{\phi,1}(r) - J_{\phi,2}(r)\Big\} dr \equiv (\sigma^2/8)\int_0^1 A(r)B(r)dr.$$

This proves part (e) using (A.15).

The next Lemma characterizes the limiting distribution of the numerator and denominator of $(\hat{\alpha} - \alpha)$, namely $T^{-2}\Sigma_{t=1}^T y_{t-1}u_t$ and $T^{-2}\Sigma_{t=1}^T y_{t-1}^2$.

LEMMA A.4 : *Let the functions $A(r)$ and $B(r)$ be as defined in Theorem 6 and consider a sequence of random variables $\{y_t\}$ defined by (7.1) and (7.2). Then as $T \to \infty$:*

a) $T^{-2}\Sigma_{t=1}^{T} y_{t-1}^2 \Rightarrow (\sigma_e^2/8)\int_0^1\left\{[A(r) - B(r)]^2 + B(r)^2\right\}dr$;

b) $T^{-2}\Sigma_{t=1}^{T} y_{t-1}u_t \Rightarrow -(\sigma_e^2/4)\int_0^1 B(r)^2 dr$.

To prove part (a), first note that $\Sigma_1^T y_{t-1}^2 = \Sigma_{k=1}^m y_{2k-1}^2 + \Sigma_{k=1}^m y_{2k-2}^2$. Using the fact that $y_{2k-1} = \alpha y_{2k-2} + u_{2k-1}$, we deduce that :

$$\Sigma_1^T y_{t-1}^2 = (\alpha^2 + 1)\Sigma_{k=1}^m y_{2k-2}^2 + 2\alpha\Sigma_{k=1}^m y_{2k-2}u_{2k-1} + \Sigma_{k=1}^m u_{2k-1}^2.$$

Using Lemma A.3 (b,d and e), we deduce that :

$$T^{-2}\Sigma_1^T y_{t-1}^2 \Rightarrow (\sigma_e^2/8)\int_0^1 A(r)^2 dr - (\sigma_e^2/4)\int_0^1 A(r)B(r)dr + (\sigma_e^2/4)\int_0^1 B(r)^2 dr$$

$$= (\sigma_e^2/8)\int_0^1\left\{[A(r) - B(r)]^2 + B(r)^2\right\} dr ,$$

as required. To prove part (b), note that using derivations similar to those used to obtain (A.6), we have :

$$T^{-2}\Sigma_1^T y_{t-1}u_t = (1/2\alpha)[T^{-2}y_T^2 - T^{-2}(\alpha^2 - 1)\Sigma_1^T y_{t-1}^2 - T^{-2}\Sigma_1^T u_t^2] .$$

Note that $\alpha \to 1$ and $T(\alpha^2 - 1) \to 2c$ as $T \to \infty$, $y_T^2 = O_p(T)$ using Lemma A.3 (a) and $\Sigma_1^T y_{t-1}^2 = O_p(T^2)$ using part (a). Hence :

$$T^{-2}\Sigma_1^T y_{t-1}u_t = -(1/2)T^{-2}\Sigma_1^T u_t^2 + o_p(1) \Rightarrow -(\sigma_e^2/4)\int_0^1 B(r)^2 dr ,$$

using Lemma A.3 (c). □

The proof of Theorem 6 follows using the fact that $\hat{\alpha} = \alpha + T^{-2}\Sigma_1^T y_{t-1}u_t / T^{-2}\Sigma_1^T y_{t-1}^2$ with Lemma A.4 and noting that $\alpha \to 1$ as $T \to \infty$. □

REFERENCES

Andrews, D.W. and R. Fair (1988): "Inference in Nonlinear Econometric Model with Structural Change," *Review of Economic Studies*, 55, 615–640.

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

Chan, N. H. (1988): "The Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association*, 83, 857–862.

Chan, N.H. (1989): "On the Nearly Nonstationary Seasonal Time Series," *The Canadian Journal of Statistics*, 17, 279–284.

Chan, N.H., and C.Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics*, 15, 1050–1063.

Dickey, D.A. (1976): "Estimation and Hypothesis Testing for Nonstationary Time Series," Unpublished Ph.D. Dissertation, Iowa State University, Ames.

Dickey, D.A., and W.A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Dickey, D.A., and S.G. Pantula (1987): "Determining the Order of Differencing in Autoregressive Processes," *Journal of Business and Economic Statistics*, 5, 455–461.

Evans, G.B.A., and N.E. Savin (1981a): "The Calculation of the Limiting Distribution of the Least Squares Estimator of the Parameter in a Random Walk Model," *Annals of Statistics*, 9, 1114–1118.

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

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

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

Herrndorf, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *Annals of Probability*, 12, 141–153.

Liptser, R.S., and A.N. Shiryayev (1978): *Statistics of Random Processes I: General Theory*, Springer–Verlag: New York.

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): "A General Approach to the Limiting Distribution for Estimators in Time Series Regression with Nonstable Autoregressive Errors," *Econometrica*, 58, 145–163.

Newey, W. and K. West (1987): "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pantula, S.G. (1988): "Asymptotic Distributions of the Unit Root Tests when the Process is Nearly Stationary," mimeo, Department of Statistics, North Carolina State University.

Perron, P. (1988a): "A Continuous Time Approximation to the Unstable First–Order Autoregressive Process : The Case without an Intercept," forthcoming in *Econometrica*.

Perron, P. (1988b): "A Continuous–Time Approximation to the Stationary First–order Autoregressive Model," Econometric Research Program Memorandum No. 337, Princeton University.

Perron, P. (1989a): "The Calculation of the Limiting Distribution of the Least Squares Estimator in a Near–Integrated Model," *Econometric Theory*, 5, 241–255.

Perron, P. (1989b): "Test Consistency with Varying Sampling Frequency," Econometric Research Program Memorandum No. 345, Princeton University.

Perron, P. (1990a): "Tests of Joint Hypotheses for Time Series Regression with a Unit Root," in T.B. Fomby and G.F. Rhodes (eds), *Advances in Econometrics, 8: Cointegration, Spurious Regression, and Unit Roots.* Greenwich, CT: JAI Press, 135–159.

Perron, P. (1990b): "The Limiting Distribution of Nearly Integrated Seasonal Models," Princeton University, in preparation.

Phillips, P.C.B. (1987a): "Time Series Regression with Unit Roots," *Econometrica*, 55, 277–302.

Phillips, P.C.B. (1987b): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–47.

Phillips, P.C.B. (1987c): "Asymptotic Expansions in Nonstationary Vector Autoregressions," *Econometric Theory*, 3, 45–68.

Phillips, P.C.B. and S. Ouliaris (1987): "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165–193.

Phillips, P.C.B., and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

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

Schwert, G.W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–160.

White, H. (1984): *Asymptotic Theory for Econometrician*, Academic Press.

White, J.S. (1958): "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics*, 29, 1188–1197.

Yajima, Y. (1985): "Asymptotic Properties of the Sample Autocorrelations and Partial Autocorrelations of a Multiplicative Process," *Journal of Time Series Analysis*, 6, 187–201.

TABLE I.A : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha = \exp(c/T)$, $c = 0.0$; MA Errors , $u_t = e_t + \theta e_{t-1}$.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\theta = -0.90$											
T=25	−34.69	−32.80	−31.05	−28.95	−21.23	−13.05	−10.90	−9.14	−7.42	−21.13	37.34
T=50	−61.57	−58.22	−55.76	−52.61	−39.95	−23.90	−19.87	−16.52	−13.64	−39.07	119.64
T=100	−110.28	−106.09	−101.79	−96.30	−71.38	−37.56	−29.85	−24.68	−19.33	−69.00	482.88
T=500	−427.96	−407.79	−385.31	−355.24	−192.30	−63.81	−45.28	−33.94	−24.94	−201.45	11573.26
T=1000	−733.45	−675.98	−625.27	−553.72	−234.38	−67.17	−49.53	−38.65	−29.70	−275.74	33716.06
T=5000	−1867.76	−1570.88	−1275.06	−999.41	−315.94	−92.19	−71.18	−57.97	−43.94	−447.14	158516.00
T=∞	−2625.73	−2036.37	−1601.94	−1181.47	−310.56	−74.68	−53.65	−41.39	−31.46	−502.44	302105.93
$\theta = -0.70$											
T=25	−30.34	−27.72	−25.47	−22.73	−11.72	−3.43	−2.24	−1.47	−0.78	−12.43	52.61
T=50	−49.25	−44.76	−40.64	−35.63	−16.52	−4.41	−3.05	−2.10	−1.34	−18.49	142.37
T=100	−77.81	−70.13	−62.69	−53.22	−20.46	−5.23	−3.45	−2.47	−1.59	−25.32	352.78
T=500	−162.58	−135.44	−115.70	−92.08	−27.12	−5.86	−3.80	−2.45	−1.46	−39.19	1347.55
T=1000	−197.33	−160.71	−133.73	−102.39	−26.75	−5.72	−3.92	−2.82	−1.91	−42.33	1882.63
T=∞	−239.25	−185.36	−145.63	−107.17	−27.50	−5.90	−3.96	−2.83	−1.90	−45.05	2528.21
$\theta = -0.50$											
T=25	−24.90	−21.54	−18.88	−15.64	−5.47	−0.63	−0.07	0.36	0.77	−6.92	36.21
T=50	−34.69	−30.10	−25.55	−20.45	−6.36	−0.83	−0.21	0.17	0.56	−8.77	66.38
T=100	−47.17	−39.19	−32.54	−25.19	−6.86	−0.95	−0.32	0.07	0.43	−10.39	111.29
T=500	−62.77	−49.76	−41.22	−31.25	−7.88	−1.00	−0.35	0.07	0.45	−12.65	192.09
T=1000	−76.00	−58.10	−45.59	−34.38	−7.97	−0.94	−0.36	0.08	0.45	−13.77	258.00
T=∞	−71.60	−55.32	−43.32	−31.71	−7.63	−1.02	−0.39	0.00	0.36	−12.91	231.35
$\theta = -0.30$											
T=25	−19.26	−15.71	−12.94	−9.97	−2.50	0.37	0.78	1.12	1.45	−3.82	20.16
T=50	−22.93	−18.54	−15.17	−11.34	−2.67	0.28	0.69	0.99	1.36	−4.34	27.03
T=100	−27.44	−21.43	−17.22	−12.61	−2.76	0.26	0.65	0.93	1.26	−4.77	35.98
T=500	−29.43	−22.96	−18.81	−13.98	−3.08	0.23	0.63	0.93	1.19	−5.28	43.82
T=∞	−31.36	−24.13	−18.79	−13.62	−2.89	0.23	0.61	0.89	1.21	−5.19	46.49
$\theta = 0.50$											
T=25	−6.98	−5.32	−4.02	−2.78	−0.12	1.46	1.93	2.36	2.97	−0.45	3.67
T=50	−6.83	−5.35	−4.06	−2.84	−0.13	1.35	1.77	2.18	2.77	−0.49	3.44
T=100	−7.58	−5.67	−4.23	−2.91	−0.16	1.33	1.73	2.16	2.71	−0.54	3.79
T=500	−7.19	−5.47	−4.33	−3.07	−0.22	1.30	1.71	2.13	2.63	−0.60	3.75
T=∞	−7.35	−5.54	−4.21	−2.91	−0.16	1.31	1.71	2.11	2.64	−0.55	3.73

TABLE I.B : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha = \exp(c/T)$, $c = -5.0$; MA Errors , $u_t = e_t + \theta e_{t-1}$.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\theta = -0.90$											
T=25	−32.74	−31.24	−29.74	−28.11	−22.10	−16.00	−14.30	−12.87	−11.22	−22.07	22.11
T=50	−61.28	−58.81	−56.62	−53.96	−45.08	−36.25	−33.58	−31.27	−28.56	−45.08	48.76
T=100	−113.30	−109.27	−106.40	−102.57	−88.57	−74.44	−70.10	−66.76	−62.42	−88.46	119.51
T=500	−460.83	−446.11	−432.60	−417.19	−347.40	−255.50	−226.33	−201.48	−170.77	−341.04	3965.71
T=1000	−815.86	−787.02	−755.75	−716.84	−546.15	−346.86	−300.30	−264.37	−233.80	−539.09	19248.78
T=5000	−2393.18	−2160.39	−1951.00	−1728.98	−908.87	−470.62	−398.75	−347.74	−307.46	−1018.31	235949.10
T=∞	−4056.61	−3408.83	−2915.65	−2416.93	−1182.91	−560.18	−455.70	−382.72	−314.411	−1368.72	637403.56
$\theta = -0.70$											
T=25	−29.91	−28.00	−26.19	−24.24	−17.18	−10.35	−8.67	−7.18	−5.45	−16.25	28.73
T=50	−51.85	−48.54	−45.63	−42.37	−30.39	−18.75	−15.81	−13.49	−10.96	−30.63	82.83
T=100	−86.91	−81.15	−76.24	−70.42	−48.00	−28.07	−23.59	−20.03	−16.53	−48.70	257.62
T=500	−211.68	−192.05	−171.53	−152.07	−85.60	−40.89	−31.74	−25.45	−18.62	−91.69	1882.48
T=1000	−276.30	−244.96	−216.71	−186.56	−96.43	−42.29	−33.76	−27.50	−22.62	−106.82	3257.43
T=∞	−365.75	−306.52	−261.44	−215.85	−103.09	−46.26	−36.74	−30.09	−23.87	−120.09	5319.65
$\theta = -0.50$											
T=25	−25.71	−23.32	−21.22	−18.74	−10.72	−4.03	−2.63	−1.45	−0.30	−11.13	32.35
T=50	−39.58	−35.64	−32.40	−28.38	−15.75	−6.33	−4.30	−2.92	−1.53	−16.76	74.40
T=100	−57.95	−51.07	−44.98	−39.28	−20.19	−8.07	−5.76	−4.15	−2.57	−22.25	150.83
T=500	−88.52	−76.07	−65.14	−54.97	−26.08	−9.52	−6.44	−4.36	−2.13	−29.71	349.85
T=1000	−99.49	−85.98	−73.53	−61.24	−27.71	−9.41	−6.69	−4.74	−3.25	−32.20	450.90
T=∞	−106.42	−88.55	−74.94	−61.19	−27.21	−10.13	−7.27	−5.28	−3.41	−32.35	482.27
$\theta = -0.30$											
T=25	−20.45	−17.92	−15.47	−12.98	−5.36	−0.21	0.78	1.51	2.26	−6.08	25.65
T=50	−27.35	−23.63	−20.45	−16.87	−6.81	−0.77	0.44	1.20	1.93	−8.02	42.24
T=100	−35.15	−29.32	−24.44	−20.25	−7.87	−1.13	0.04	0.93	1.71	−9.54	61.95
T=500	−41.05	−33.95	−28.24	−23.18	−9.06	−1.30	0.11	1.09	2.09	−10.94	84.61
T=∞	−44.16	−36.22	−30.17	−24.06	−8.99	−1.43	−0.16	0.73	1.58	−11.27	94.69
$\theta = 0.50$											
T=25	−6.06	−4.28	−3.05	−1.72	1.73	3.75	4.21	4.68	5.32	1.32	5.39
T=50	−6.18	−4.47	−3.13	−1.79	1.81	3.82	4.28	4.76	5.31	1.35	5.57
T=100	−7.01	−5.07	−3.40	−1.91	1.83	3.86	4.29	4.76	5.44	1.32	6.27
T=500	−6.92	−4.82	−3.26	−1.80	1.85	3.93	4.45	4.84	5.36	1.37	6.08
T=∞	−6.84	−4.86	−3.36	−1.84	1.92	3.91	4.36	4.96	5.39	1.40	5.78

TABLE I.C : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha = \exp(c/T)$, $c = 2.0$; MA Errors , $u_t = e_t + \theta e_{t-1}$.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\theta = -0.90$											
T=25	–34.87	–32.28	–29.98	–27.00	–12.26	–3.07	–2.08	–1.51	–0.97	–13.82	82.30
T=50	–59.92	–56.57	–53.11	–47.87	–20.75	–5.06	–3.65	–2.64	–2.01	–23.96	263.61
T=100	–107.70	–101.89	–95.45	–86.76	–30.61	–7.05	–5.12	–3.80	–2.86	–36.64	904.97
T=500	–409.85	–378.59	–337.85	–274.69	–53.64	–10.14	–6.70	–4.93	–3.45	–101.14	11469.52
T=1000	–680.46	–611.17	–517.97	–396.92	–57.66	–10.33	–7.31	–5.44	–4.12	–133.57	27448.45
T=5000	–1515.21	–1157.27	–827.82	–509.39	–58.66	–11.16	–8.31	–6.61	–4.80	–183.44	92427.60
T=∞	–1989.44	–1420.93	–1011.29	–630.96	–61.74	–10.81	–7.58	–5.76	–4.31	–221.30	163176.94
$\theta = -0.70$											
T=25	–30.00	–26.42	–23.08	–18.77	–3.69	–0.45	–0.13	0.09	0.32	–6.89	56.21
T=50	–45.72	–40.67	–34.72	–27.12	–4.48	–0.60	–0.29	–0.09	0.13	–9.41	126.45
T=100	–71.86	–61.79	–51.50	–37.67	–4.84	–0.70	–0.37	–0.17	0.01	–12.45	279.97
T=500	–138.78	–111.47	–80.94	–52.35	–5.70	–0.78	–0.42	–0.21	–0.05	–18.06	843.09
T=1000	–164.70	–127.00	–91.14	–58.43	–5.55	–0.78	–0.45	–0.25	–0.05	–19.61	1129.94
T=∞	–182.64	–130.64	–93.14	–58.22	–5.55	–0.78	–0.46	–0.26	–0.09	–20.27	1382.53
$\theta = -0.50$											
T=25	–24.01	–19.92	–15.93	–11.75	–1.47	0.11	0.33	0.52	0.76	–3.80	29.59
T=50	–30.91	–25.37	–20.01	–13.67	–1.55	0.06	0.25	0.42	0.62	–4.46	46.39
T=100	–41.42	–31.89	–24.59	–16.14	–1.55	0.03	0.21	0.35	0.52	–5.23	75.04
T=500	–52.35	–40.01	–28.23	–17.59	–1.70	0.00	0.18	0.32	0.46	–6.09	116.79
T=1000	–57.02	–42.79	–30.16	–18.88	–1.64	0.00	0.19	0.36	0.52	–6.31	133.04
T=∞	–55.80	–40.06	–28.67	–18.03	–1.63	0.01	0.18	0.31	0.47	–6.14	130.55
$\theta = -0.30$											
T=25	–18.06	–14.14	–10.59	–7.31	–0.72	0.37	0.61	0.84	1.13	–2.25	15.21
T=50	–19.89	–15.57	–11.64	–7.67	–0.71	0.33	0.53	0.74	1.00	–2.39	17.91
T=100	–23.70	–17.38	–13.07	–8.37	–0.71	0.32	0.50	0.68	0.90	–2.64	23.94
T=500	–25.39	–19.08	–13.37	–8.49	–0.76	0.27	0.47	0.65	0.85	–2.81	27.85
T=∞	–25.38	–18.35	–13.26	–8.45	–0.71	0.30	0.48	0.65	0.87	–2.75	27.66
$\theta = 0.50$											
T=25	–7.57	–5.66	–4.28	–3.00	–0.15	0.78	1.12	1.54	2.07	–0.67	3.31
T=50	–6.97	–5.38	–4.22	–2.92	–0.15	0.70	1.03	1.34	1.81	–0.67	2.87
T=100	–7.48	–5.70	–4.35	–3.00	–0.16	0.68	0.95	1.29	1.81	–0.72	3.20
T=500	–7.40	–5.70	–4.38	–2.96	–0.18	0.66	0.96	1.26	1.70	–0.74	3.17
T=∞	–7.30	–5.53	–4.22	–2.95	–0.16	0.66	0.95	1.26	1.71	–0.71	2.71

TABLE II.A : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha=\exp(c/T)$, $c=0.0$; AR Errors , $u_t=\rho u_{t-1}+e_t$.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho=-0.90$											
T=25	−44.24	−41.22	−38.31	−33.76	−13.89	−2.70	−1.39	−0.65	−0.04	−16.24	134.04
T=50	−76.33	−69.41	−62.43	−53.52	−19.36	−4.16	−2.42	−1.46	−0.71	−24.35	359.69
T=100	−122.62	−108.94	−93.53	−75.20	−23.93	−5.36	−3.34	−2.21	−1.34	−33.10	826.35
T=500	−214.48	−178.94	−146.79	−112.24	−31.24	−6.58	−4.09	−2.83	−1.71	−47.75	2284.21
T=1000	−273.27	−216.22	−175.23	−130.44	−32.63	−6.81	−4.65	−3.42	−2.35	−53.65	3359.83
T=∞	−274.69	−212.87	−167.27	−123.14	−31.71	−6.93	−4.71	−3.41	−2.35	−51.85	3329.55
$\rho=-0.50$											
T=25	−24.84	−20.50	−17.28	−13.48	−3.54	0.08	0.52	0.85	1.24	−5.31	33.31
T=50	−30.92	−25.19	−20.46	−15.46	−3.86	−0.05	0.42	0.74	1.09	−6.11	47.49
T=100	−36.92	−29.36	−23.45	−17.22	−3.99	−0.11	0.36	0.66	0.98	−6.74	64.12
T=500	−40.14	−31.21	−25.53	−19.16	−4.42	−0.13	0.34	0.64	0.94	−7.46	79.20
T=∞	−42.61	−32.84	−25.64	−18.67	−4.21	−0.14	0.30	0.60	0.91	−7.34	83.94
$\rho=0.50$											
T=25	−5.28	−3.60	−2.67	−1.73	0.27	1.76	2.31	2.82	3.46	0.10	2.55
T=50	−4.47	−3.38	−2.47	−1.61	0.24	1.65	2.11	2.62	3.26	0.09	2.08
T=100	−4.67	−3.42	−2.44	−1.63	0.20	1.62	2.06	2.53	3.26	0.07	2.08
T=500	−4.32	−3.15	−2.47	−1.67	0.14	1.57	2.05	2.50	3.14	0.03	1.96
T=∞	−4.23	−3.15	−2.35	−1.57	0.19	1.57	2.03	2.49	3.11	0.07	1.92
$\rho=0.90$											
T=25	−1.66	−0.95	−0.57	−0.28	1.25	2.83	3.44	4.06	4.86	1.27	1.67
T=50	−1.04	−0.64	−0.38	−0.17	1.04	2.61	3.25	3.91	4.63	1.14	1.36
T=100	−0.73	−0.45	−0.28	−0.12	0.86	2.33	3.00	3.54	4.47	1.01	1.14
T=500	−0.58	−0.40	−0.26	−0.12	0.69	2.15	2.73	3.34	4.07	0.87	0.98
T=∞	−0.53	−0.36	−0.23	−0.12	0.67	2.08	2.65	3.23	4.03	0.85	0.93
$\rho=0.95$											
T=25	−1.35	−0.70	−0.40	−0.11	1.52	3.07	3.73	4.33	5.06	1.53	1.68
T=50	−0.69	−0.37	−0.21	−0.05	1.33	2.89	3.56	4.24	5.02	1.42	1.42
T=100	−0.43	−0.25	−0.13	−0.02	1.11	2.63	3.29	3.87	4.76	1.24	1.22
T=500	−0.28	−0.17	−0.10	−0.03	0.79	2.29	2.89	3.51	4.43	0.99	1.01
T=1000	−0.28	−0.19	−0.11	−0.04	0.71	2.24	2.78	3.42	4.28	0.94	0.97
T=∞	−0.22	−0.14	−0.08	−0.01	0.73	2.15	2.73	3.33	4.14	0.93	0.90

TABLE II.B : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha=\exp(c/T)$, $c=-5.0$; AR Errors , $u_t=\rho u_{t-1}+e_t$.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho=-0.90$											
T=25	–42.21	–41.04	–39.69	–37.58	–25.90	–11.74	–8.71	–6.27	–3.60	–25.29	91.39
T=50	–81.36	–76.85	–73.48	–68.17	–44.22	–20.83	–15.84	–12.03	–8.38	–44.42	308.97
T=100	–142.18	–131.84	–123.01	–110.87	–66.12	–31.37	–24.38	–19.51	–15.03	–68.69	899.28
T=500	–302.56	–262.73	–233.83	–199.97	–104.99	–46.22	–35.38	–28.33	–21.05	–115.80	3805.87
T=1000	–355.45	–315.41	–271.69	–231.61	–113.44	–49.51	–38.99	–31.82	–26.04	–129.27	5475.29
T=∞	–420.62	–352.65	–300.89	–248.56	–119.14	–53.90	–42.97	–35.33	–28.19	–138.65	7009.04
$\rho=-0.50$											
T=25	–26.65	–24.10	–21.40	–18.11	–8.11	–1.64	–0.42	0.53	1.49	–9.29	41.44
T=50	–37.40	–32.91	–28.63	–23.96	–10.68	–2.54	–1.04	0.01	0.90	–12.29	74.12
T=100	–48.54	–40.85	–34.73	–29.21	–12.33	–3.26	–1.62	–0.64	0.49	–14.59	111.84
T=500	–56.94	–48.26	–40.07	–33.25	–14.13	–3.62	–1.74	–0.44	0.98	–16.67	154.81
T=1000	–61.29	–52.30	–44.02	–35.89	–14.72	–3.54	–1.84	–0.64	0.22	–17.70	180.81
T=∞	–61.55	–50.83	–42.68	–34.43	–14.08	–3.87	–2.16	–0.96	0.17	–17.16	172.84
$\rho=0.50$											
T=25	–3.73	–2.17	–1.07	0.10	2.83	4.45	4.98	5.48	6.20	2.50	3.74
T=50	–2.85	–1.42	–0.44	0.56	3.05	4.57	5.08	5.60	6.21	2.78	3.02
T=100	–2.77	–1.39	–0.33	0.69	3.17	4.63	5.13	5.69	6.43	2.89	3.03
T=500	–1.92	–0.83	0.14	1.02	3.26	4.72	5.26	5.70	6.30	3.04	2.59
T=∞	–1.93	–0.75	0.16	1.07	3.33	4.72	5.23	5.73	6.41	3.09	2.10
$\rho=0.90$											
T=25	1.16	2.22	2.82	3.37	4.59	6.30	6.92	7.54	8.33	4.71	1.69
T=50	2.77	3.30	3.63	3.96	4.79	6.39	7.10	7.71	8.47	5.01	1.18
T=100	3.33	3.74	4.02	4.25	4.87	6.37	7.01	7.63	8.42	5.12	0.95
T=500	3.92	4.13	4.31	4.48	4.94	6.32	6.92	7.60	8.34	5.21	0.75
T=∞	4.05	4.24	4.38	4.53	5.03	6.26	6.87	7.50	8.33	5.22	0.24
$\rho=0.95$											
T=25	1.69	2.71	3.25	3.72	4.95	6.64	7.27	7.89	8.64	5.05	1.68
T=50	3.30	3.72	4.02	4.29	5.10	6.79	7.52	8.09	8.91	5.35	1.24
T=100	3.96	4.23	4.41	4.57	5.13	6.75	7.43	8.03	9.02	5.43	1.01
T=500	4.44	4.57	4.66	4.76	5.12	6.59	7.28	7.92	8.72	5.44	0.79
T=1000	4.48	4.59	4.68	4.76	5.12	6.49	7.15	7.78	8.55	5.42	0.71
T=∞	4.55	4.65	4.73	4.81	5.12	6.46	7.08	7.73	8.57	5.42	0.28

TABLE II.C : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha = \exp(c/T)$, $c = 2.0$; AR Errors , $u_t = \rho u_{t-1} + e_t$.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho = -0.90$											
T=25	−44.03	−39.32	−33.32	−25.56	−4.15	−0.34	0.00	0.22	0.47	−8.86	115.69
T=50	−70.20	−60.70	−50.24	−36.36	−5.08	−0.55	−0.22	0.01	0.20	−12.20	264.31
T=100	−109.43	−90.73	−71.88	−49.03	−5.47	−0.73	−0.39	−0.16	0.04	−16.09	573.18
T=500	−176.34	−139.82	−99.11	−63.53	−6.47	−0.89	−0.50	−0.28	−0.09	−21.86	1360.10
T=1000	−200.93	−153.88	−106.59	−67.78	−6.37	−0.92	−0.55	−0.32	−0.14	−23.08	1682.66
T=∞	−209.48	−149.84	−106.78	−66.74	−6.39	−0.94	−0.57	−0.35	−0.16	−23.26	1818.94
$\rho = -0.50$											
T=25	−22.91	−18.22	−13.60	−9.22	−0.95	0.28	0.50	0.71	0.96	−2.96	24.30
T=50	−26.15	−20.21	−15.11	−9.84	−0.96	0.24	0.43	0.60	0.83	−3.19	30.32
T=100	−31.78	−23.34	−17.23	−10.87	−0.95	0.22	0.39	0.54	0.74	−3.54	41.40
T=500	−33.97	−25.46	−17.80	−11.15	−1.01	0.17	0.36	0.52	0.72	−3.78	48.75
T=∞	−33.87	−24.40	−17.55	−11.12	−0.97	0.20	0.37	0.52	0.71	−3.70	48.67
$\rho = 0.50$											
T=25	−5.85	−4.36	−3.36	−2.43	−0.03	0.93	1.32	1.76	2.42	−0.40	2.39
T=50	−5.07	−4.09	−3.34	−2.37	−0.06	0.83	1.19	1.59	2.21	−0.40	1.93
T=100	−5.39	−4.15	−3.24	−2.39	−0.07	0.77	1.09	1.50	2.13	−0.45	1.98
T=500	−5.11	−4.02	−3.22	−2.35	−0.09	0.76	1.10	1.45	2.01	−0.46	1.91
T=∞	−4.99	−3.93	−3.13	−2.33	−0.08	0.75	1.09	1.46	1.98	−0.44	1.48
$\rho = 0.90$											
T=25	−3.11	−2.56	−2.19	−1.35	0.41	1.55	2.13	2.71	3.49	0.29	1.55
T=50	−2.63	−2.31	−2.07	−1.42	0.25	1.31	1.88	2.42	3.24	0.15	1.27
T=100	−2.74	−2.24	−2.05	−1.48	0.14	1.12	1.58	2.11	2.81	0.03	1.12
T=500	−2.40	−2.18	−2.03	−1.56	0.03	0.99	1.39	1.93	2.67	−0.08	1.03
T=∞	−2.32	−2.16	−2.00	−1.55	0.03	0.93	1.35	1.83	2.51	−0.09	0.63
$\rho = 0.95$											
T=25	−2.83	−2.40	−2.04	−1.11	0.59	1.75	2.33	2.91	3.68	0.48	1.53
T=50	−2.41	−2.18	−1.88	−1.18	0.44	1.56	2.16	2.70	3.56	0.36	1.29
T=100	−2.25	−2.11	−1.89	−1.30	0.27	1.31	1.83	2.38	3.12	0.19	1.14
T=500	−2.16	−2.06	−1.92	−1.47	0.07	1.05	1.48	2.01	2.78	−0.02	1.00
T=1000	−2.15	−2.05	−1.87	−1.41	0.04	1.00	1.45	1.96	2.57	−0.02	0.93
T=∞	−2.13	−2.04	−1.90	−1.47	0.04	0.95	1.39	1.88	2.58	−0.06	0.59

TABLE III : Percentage Error for the Variance of the Asymptotic Distribution

A : MA Errors : $u_t = e_t + \theta e_{t-1}$.

θ	T=25	T=50	T=100	T=500	T=1000	T=5000
–.95	1596.00	798.00	399.00	79.80	39.90	7.98
–.90	396.00	198.00	99.00	19.80	9.90	1.98
–.80	96.00	48.00	24.00	4.80	2.40	.48
–.70	40.44	20.22	10.11	2.02	1.01	.20
–.60	21.00	10.50	5.25	1.05	.53	.11
–.50	12.00	6.00	3.00	.60	.30	.06
–.40	7.11	3.56	1.78	.36	.18	.04
–.30	4.16	2.08	1.04	.21	.10	.02
–.20	2.25	1.12	.56	.11	.06	.01
–.10	.94	.47	.23	.05	.02	.00
.10	–.69	–.35	–.17	–.03	–.02	.00
.20	–1.22	–.61	–.31	–.06	–.03	–.01
.30	–1.63	–.82	–.41	–.08	–.04	–.01
.40	–1.96	–.98	–.49	–.10	–.05	–.01
.50	–2.22	–1.11	–.56	–.11	–.06	–.01
.60	–2.44	–1.22	–.61	–.12	–.06	–.01
.70	–2.62	–1.31	–.65	–.13	–.07	–.01
.80	–2.77	–1.38	–.69	–.14	–.07	–.01
.90	–2.89	–1.45	–.72	–.14	–.07	–.01
.95	–2.95	–1.47	–.74	–.15	–.07	–.01

B : AR Errors : $u_t = \rho u_{t-1} + e_t$.

ρ	T=25	T=50	T=100	T=500	T=1000	T=5000
–.95	39.16	20.20	10.22	2.05	1.02	.20
–.90	21.03	10.41	5.21	1.04	.52	.10
–.80	10.68	5.33	2.67	.53	.27	.05
–.70	7.14	3.57	1.78	.36	.18	.04
–.60	5.25	2.63	1.31	.26	.13	.03
–.50	4.00	2.00	1.00	.20	.10	.02
–.40	3.05	1.52	.76	.15	.08	.02
–.30	2.24	1.12	.56	.11	.06	.01
–.20	1.50	.75	.38	.07	.04	.01
–.10	.77	.38	.19	.04	.02	.00
.10	–.85	–.42	–.21	–.04	–.02	.00
.20	–1.83	–.92	–.46	–.09	–.05	–.01
.30	–3.03	–1.52	–.76	–.15	–.08	–.02
.40	–4.57	–2.29	–1.14	–.23	–.11	–.02
.50	–6.67	–3.33	–1.67	–.33	–.17	–.03
.60	–9.75	–4.88	–2.44	–.49	–.24	–.05
.70	–14.82	–7.41	–3.71	–.74	–.37	–.07
.80	–24.77	–12.44	–6.22	–1.24	–.62	–.12
.90	–49.87	–27.29	–13.74	–2.75	–1.37	–.27
.95	–75.66	–51.75	–28.52	–5.75	–2.87	–.57

TABLE IV : Percentage Points of the Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha = \exp(c/T)$, $c = 0.0$

Nearly White Noise Model ; $\theta_T = -1 + \delta/T^{1/2}$.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
$\theta = -0.90$									
T=25	–24.81	–24.73	–24.65	–24.53	–23.31	–19.25	–17.68	–16.30	–14.73
T=50	–49.17	–48.93	–48.62	–48.16	–43.66	–31.29	–27.35	–24.19	–20.89
T=100	–96.75	–95.75	–94.63	–92.89	–77.49	–45.54	–37.65	–31.91	–26.40
T=500	–427.00	–408.93	–389.68	–361.59	–203.88	–71.65	–53.89	–42.84	–33.48
T=1000	–744.25	–691.85	–638.78	–566.38	–256.12	–77.17	–56.94	–44.74	–34.58
T=5000	–1837.34	–1549.38	–1305.06	–1036.31	–322.20	–82.25	–59.69	–46.40	–35.63
$\theta = -0.70$									
T=25	–23.21	–22.72	–22.18	–21.33	–15.12	–6.77	–5.29	–4.31	–3.44
T=50	–43.36	–41.66	–39.84	–37.19	–21.67	–7.83	–5.92	–4.71	–3.69
T=100	–76.42	–71.38	–66.25	–59.21	–27.67	–8.50	–6.29	–4.95	–3.83
T=500	–196.22	–166.42	–140.88	–112.57	–35.55	–9.12	–6.63	–5.15	–3.96
T=1000	–245.11	–199.78	–164.27	–126.77	–36.84	–9.21	–6.67	–5.18	–3.97
$\theta = -0.50$									
T=25	–20.58	–19.56	–18.48	–16.92	–8.88	–2.95	–2.20	–1.74	–1.36
T=50	–35.08	–32.13	–29.31	–25.57	–10.80	–3.14	–2.30	–1.81	–1.40
T=100	–53.99	–47.33	–41.45	–34.32	–12.10	–3.24	–2.36	–1.84	–1.41
T=500	–94.55	–76.21	–62.02	–47.31	–13.40	–3.32	–2.40	–1.87	–1.43

TABLE V : Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha=\exp(c/T)$;

Nearly White Noise Model ; $\theta_T=-1+\delta/T^{1/2}$.

A : Mean .

θ	T=25	T=50	T=100	T=500	T=1000	T=5000
c = 0.0						
–0.90	–22.51	–41.50	–73.02	–210.72	–291.68	–456.60
–0.70	–14.54	–22.12	–31.01	–49.83	–54.88	
–0.50	–9.45	–12.73	–15.80	–20.37	–21.24	
c = –5.0						
–0.90	–24.45	–47.88	–91.96	–354.94	–561.33	–1103.46
–0.70	–20.98	–36.49	–58.54	–119.19	–138.92	
–0.50	–16.62	–25.48	–35.25	–52.53	–56.27	
c = 2.0						
–0.90	–17.17	–28.60	–45.37	–107.63	–141.02	–206.18
–0.70	–8.10	–11.41	–15.10	–22.56	–24.51	
–0.50	–4.73	–6.07	–7.29	–9.06	–9.39	

B : Variance

θ	T=25	T=50	T=100	T=500	T=1000	T=5000
c = 0.0						
–0.90	5.20	46.22	326.00	11380.04	34048.96	167052.12
–0.70	28.73	116.25	361.89	1928.47	2796.55	
–0.50	26.66	73.38	157.03	405.98	475.96	
c = –5.0						
–0.90	0.10	1.44	18.27	3071.33	16813.96	213117.59
–0.70	3.44	28.38	164.09	2358.53	4302.75	
–0.50	8.73	43.03	144.62	677.08	901.64	
c = 2.0						
–0.90	35.76	185.27	796.71	11931.33	28353.66	103245.87
–0.70	43.31	127.07	310.54	1205.10	1634.98	
–0.50	25.74	57.56	106.55	235.49	260.28	

TABLE VI : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha = \exp(c/T)$; $c = 0.0$; AR Errors ; $u_t = \rho u_{t-1} + e_t$.

A: $O(T^{-1})$ Expansions

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho = 0.50$											
T=25	–4.80	–3.44	–2.48	–1.60	0.22	1.60	2.07	2.54	3.19	0.07	2.17
T=50	–4.53	–3.29	–2.41	–1.58	0.20	1.59	2.05	2.52	3.15	0.07	2.05
T=100	–4.38	–3.22	–2.38	–1.57	0.20	1.58	2.04	2.51	3.13	0.07	1.98
T=500	–4.26	–3.16	–2.35	–1.57	0.19	1.57	2.03	2.50	3.12	0.07	1.93
$\rho = 0.90$											
T=25	–0.80	–0.49	–0.29	–0.11	0.68	2.09	2.66	3.25	4.05	0.85	0.96
T=50	–0.68	–0.43	–0.26	–0.11	0.68	2.08	2.66	3.24	4.04	0.85	0.94
T=100	–0.61	–0.39	–0.24	–0.11	0.68	2.08	2.65	3.24	4.03	0.85	0.93
T=500	–0.54	–0.36	–0.23	–0.11	0.68	2.08	2.65	3.24	4.03	0.85	0.93
$\rho = 0.95$											
T=25	–0.41	–0.23	–0.12	–0.02	0.73	2.15	2.74	3.34	4.15	0.93	0.92
T=50	–0.34	–0.19	–0.10	–0.02	0.73	2.15	2.74	3.34	4.14	0.93	0.91
T=100	–0.29	–0.17	–0.09	–0.02	0.73	2.15	2.73	3.33	4.14	0.93	0.90
T=500	–0.24	–0.15	–0.08	–0.02	0.73	2.15	2.73	3.33	4.14	0.93	0.90
T=1000	–0.23	–0.15	–0.08	–0.02	0.73	2.15	2.73	3.33	4.14	0.93	0.90

B : Nearly Doubly Integrated Model .

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho = 0.50$											
T=25	0.001	0.006	0.02	0.08	0.91	2.56	3.17	3.88	4.88	1.16	1.11
T=50	0.001	0.003	0.02	0.06	0.85	2.32	2.97	3.54	4.45	1.07	0.99
T=100	0.001	0.005	0.02	0.07	0.83	2.36	2.97	3.63	4.60	1.07	1.00
$\rho = 0.90$											
T=25	0.003	0.015	0.06	0.22	1.41	3.09	3.75	4.46	5.27	1.58	1.33
T=50	0.002	0.01	0.04	0.14	1.18	2.73	3.38	4.09	5.10	1.36	1.20
T=100	0.001	0.01	0.03	0.11	1.03	2.64	3.32	4.06	4.86	1.25	1.17
T=500	0.001	0.004	0.02	0.06	0.80	2.36	2.96	3.57	4.50	1.06	1.01
$\rho = 0.95$											
T=25	0.004	0.03	0.10	0.33	1.63	3.26	3.95	4.51	5.40	1.76	1.36
T=50	0.002	0.02	0.06	0.22	1.44	2.97	3.61	4.34	5.28	1.56	1.26
T=100	0.002	0.01	0.04	0.14	1.18	2.88	3.55	4.23	5.09	1.40	1.27
T=500	0.001	0.01	0.02	0.07	0.87	2.41	3.08	3.79	4.56	1.11	1.08
T=1000	0.001	0.004	0.02	0.06	0.82	2.27	2.84	3.53	4.50	1.04	0.94

TABLE VII : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha=\exp(c/T)$; $c=-5.0$; AR Errors ; $u_t=\rho u_{t-1}+e_t$.

A: $O(T^{-1})$ Expansions

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho=0.50$											
T=25	-3.34	-1.69	-0.48	0.71	3.44	4.96	5.48	6.02	6.74	3.09	3.28
T=50	-2.72	-1.26	-0.18	0.89	3.40	4.83	5.34	5.87	6.56	3.09	2.71
T=100	-2.35	-1.01	-0.02	0.98	3.37	4.77	5.28	5.80	6.48	3.09	2.42
T=500	-2.02	-0.80	0.12	1.05	3.34	4.73	5.23	5.75	6.42	3.09	2.19
$\rho=0.90$											
T=25	3.35	3.77	4.07	4.35	5.04	6.33	6.95	7.58	8.43	5.22	0.42
T=50	3.63	3.96	4.20	4.33	5.01	6.29	6.91	7.54	8.39	5.22	0.34
T=100	3.80	4.08	4.28	4.48	4.99	6.27	6.89	7.52	8.35	5.22	0.30
T=500	3.99	4.20	4.36	4.52	4.96	6.26	6.88	7.50	8.33	5.22	0.27
$\rho=0.95$											
T=25	4.07	4.33	4.52	4.69	5.18	6.50	7.13	7.77	8.63	5.42	0.32
T=50	4.26	4.46	4.61	4.74	5.16	6.48	7.11	7.75	8.60	5.42	0.28
T=100	4.38	4.55	4.66	4.77	5.14	6.47	7.10	7.74	8.59	5.42	0.26
T=500	4.52	4.63	4.71	4.80	5.12	6.46	7.09	7.73	8.58	5.42	0.25
T=1000	4.54	4.65	4.72	4.80	5.11	6.46	7.09	7.72	8.58	5.42	0.25

B : Nearly Doubly Integrated Model.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho=0.50$											
T=25	5.0001	5.0008	5.003	5.014	5.36	6.93	7.66	8.50	9.50	5.73	0.94
T=50	5.0001	5.0008	5.004	5.014	5.32	6.79	7.47	8.16	9.11	5.67	0.79
T=100	5.0001	5.0008	5.012	5.012	5.32	6.74	7.46	8.00	9.06	5.65	0.75
$\rho=0.90$											
T=25	5.0004	5.002	5.008	5.03	5.66	7.49	8.22	9.00	9.97	6.02	1.22
T=50	5.0002	5.002	5.006	5.02	5.49	7.18	7.93	8.59	9.47	5.85	1.01
T=100	5.0002	5.001	5.004	5.02	5.40	7.07	7.85	8.61	9.57	5.78	1.02
T=500	5.0001	5.001	5.002	5.01	5.31	6.76	7.41	8.01	8.79	5.65	0.73
$\rho=0.95$											
T=25	5.001	5.004	5.017	5.06	5.84	7.64	8.39	9.11	10.10	6.16	1.27
T=50	5.0005	5.002	5.007	5.03	5.65	7.34	8.09	8.79	9.63	5.98	1.12
T=100	5.0001	5.0013	5.006	5.02	5.52	7.29	8.09	8.87	9.86	5.89	1.14
T=500	5.0001	5.0006	5.002	5.01	5.33	6.84	7.49	8.21	9.00	5.68	0.80
T=1000	5.0001	5.0007	5.003	5.01	5.32	6.74	7.44	8.14	8.94	5.65	0.76

TABLE VIII : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha=\exp(c/T)$; $c=2.0$; AR Errors ; $u_t=\rho u_{t-1}+e_t$.

A: $O(T^{-1})$ Expansions

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho=0.50$											
T=25	–5.24	–3.99	–3.11	–2.28	–0.08	0.76	1.10	1.47	2.01	–0.44	1.58
T=50	–5.11	–3.95	–3.12	–2.31	–0.08	0.76	1.09	1.46	2.00	–0.44	1.53
T=100	–5.05	–3.93	–3.12	–2.32	–0.08	0.75	1.09	1.46	1.99	–0.44	1.51
T=500	–5.00	–3.92	–3.13	–2.33	–0.08	0.75	1.09	1.46	1.99	–0.44	1.49
$\rho=0.90$											
T=25	–2.43	–2.16	–1.97	–1.54	0.03	0.93	1.36	1.83	2.52	–0.09	0.64
T=50	–2.38	–2.15	–1.98	–1.55	0.03	0.93	1.36	1.83	2.52	–0.09	0.63
T=100	–2.35	–2.15	–1.99	–1.55	0.03	0.93	1.36	1.83	2.52	–0.09	0.63
T=500	–2.33	–2.15	–2.00	–1.55	0.03	0.93	1.35	1.83	2.51	–0.09	0.63
$\rho=0.95$											
T=25	–2.19	–2.04	–1.88	–1.47	0.04	0.95	1.39	1.88	2.59	–0.06	0.59
T=50	–2.16	–2.03	–1.89	–1.47	0.04	0.95	1.39	1.88	2.58	–0.06	0.59
T=100	–2.14	–2.04	–1.90	–1.47	0.04	0.95	1.39	1.88	2.58	–0.06	0.59
T=500	–2.13	–2.04	–1.90	–1.47	0.04	0.95	1.39	1.88	2.58	–0.06	0.59
T=1000	–2.13	–2.04	–1.90	–1.47	0.04	0.95	1.39	1.88	2.58	–0.06	0.59

B : Nearly Doubly Integrated Model

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
$\rho=0.50$											
T=25	–1.99	–1.93	–1.76	–1.30	0.12	1.12	1.71	2.29	2.98	0.09	0.99
T=50	-1.99	-1.93	-1.78	-1.33	0.07	1.05	1.52	2.04	2.97	0.03	0.92
T=100	–1.99	–1.95	–1.79	–1.32	0.06	1.08	1.53	2.09	2.78	0.02	0.92
$\rho=0.90$											
T=25	–1.97	–1.87	–1.57	–0.99	0.44	1.59	2.22	2.79	3.65	0.43	1.15
T=50	–1.99	–1.92	–1.71	–1.14	0.29	1.33	1.91	2.41	3.30	0.25	1.06
T=100	–1.99	–1.91	–1.72	–1.20	0.19	1.26	1.77	2.30	3.13	0.16	1.01
T=500	–1.99	–1.95	–1.80	–1.34	0.06	1.08	1.54	2.17	3.00	0.03	0.97
$\rho=0.95$											
T=25	–1.97	–1.84	–1.47	–0.79	0.59	1.76	2.41	2.98	3.81	0.58	1.18
T=50	–1.98	–1.86	–1.58	–0.98	0.46	1.54	2.12	2.69	3.56	0.42	1.12
T=100	–1.98	–1.89	–1.65	–1.09	0.31	1.41	1.99	2.59	3.34	0.29	1.09
T=500	–1.99	–1.95	–1.75	–1.30	0.09	1.14	1.63	2.27	3.03	0.07	1.01
T=1000	–1.99	–1.94	–1.78	–1.38	0.05	0.97	1.41	1.99	2.83	–0.02	0.92

TABLE IX : The Distribution of $T(\hat{\alpha}-\alpha)$; $\alpha=\exp(c/T)$; AR Errors , $u_t=\rho u_{t-1}+e_t$, $\rho=-0.90$.

Nearly Seasonally Integrated Model.

	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%	Mean	Var
c = 0.0											
T=25	–40.46	–37.73	–34.22	–29.76	–12.63	–3.15	–2.10	–1.46	–1.02	–14.65	100.57
T=50	–65.73	–58.00	–51.37	–42.57	–14.66	–3.83	–2.71	–2.04	–1.45	–19.27	239.49
T=100	–93.98	–81.35	–68.69	–54.90	–16.45	–4.64	–3.43	–2.65	–1.94	–23.93	454.55
T=500	–131.42	–104.61	–86.49	–64.01	–16.80	–4.36	–3.19	–2.47	–1.98	–27.13	821.37
c = –5.0											
T=25	–40.81	–39.20	–37.74	–35.78	–24.88	–10.94	–7.42	–5.50	–3.09	–24.07	85.58
T=50	–81.23	–78.27	–74.59	–69.38	–45.74	–19.83	–14.64	–11.03	–7.98	–45.14	337.25
T=100	–138.55	–126.70	–118.34	–105.87	–62.86	–26.28	–19.80	–15.80	–12.35	–64.84	886.13
T=500	–218.40	–189.73	–163.90	–140.50	–68.45	–28.34	–22.30	–17.61	–14.34	–77.70	2047.31
c = 2.0											
T=25	–35.53	–29.12	–23.82	–19.02	–4.08	–2.48	–2.34	–2.27	–2.23	–7.28	51.99
T=50	–52.14	–40.84	–32.29	–22.40	–4.37	–2.52	–2.37	–2.28	–2.23	–8.91	110.75
T=100	–69.42	–53.36	–38.64	–27.14	–4.57	–2.53	–2.39	–2.30	–2.24	–10.34	189.56
T=500	–93.61	–69.86	–51.92	–33.74	–4.87	–2.54	–2.39	–2.31	–2.24	–12.68	365.75

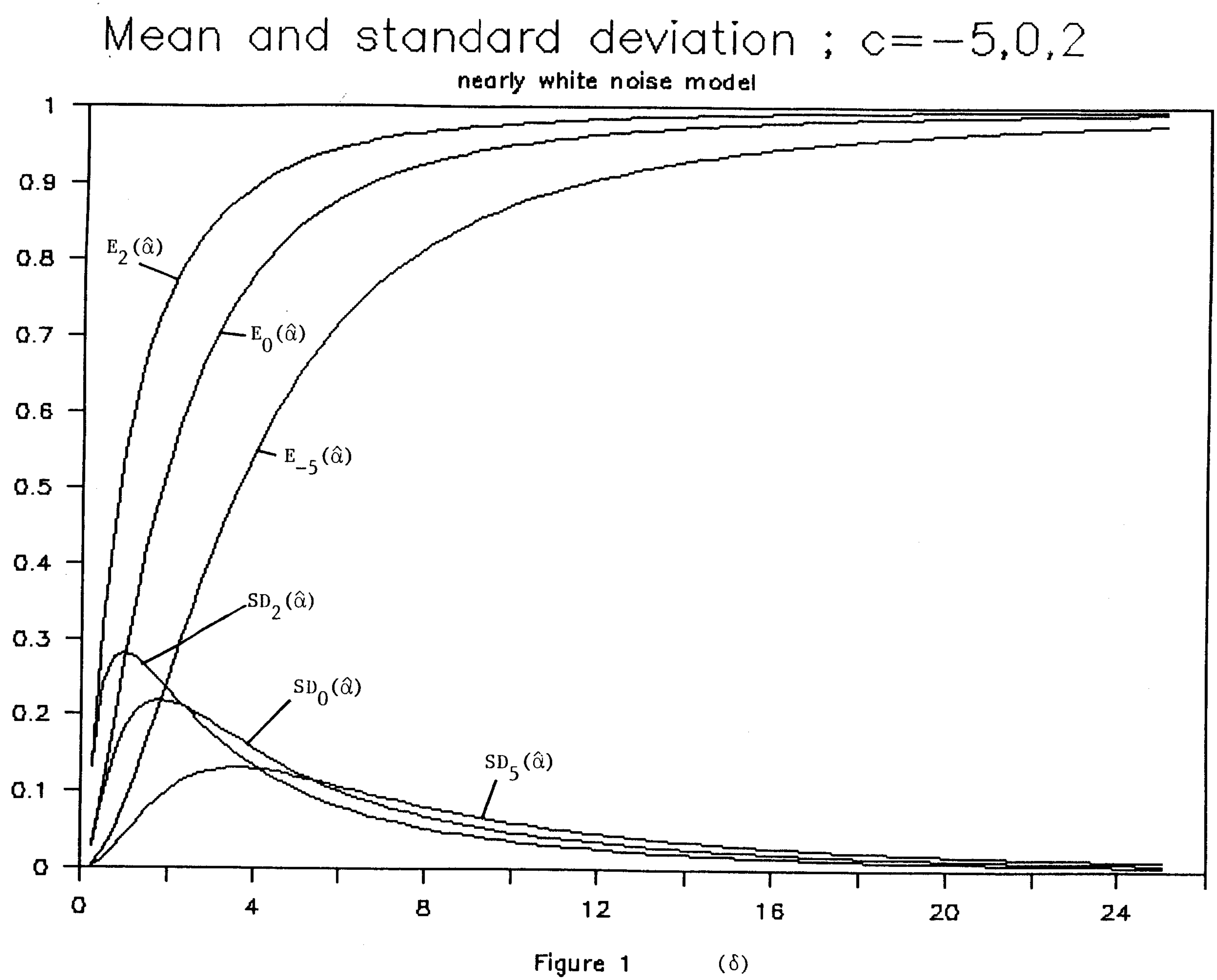

Figure 1

Notes: $E_x(\hat{\alpha})$ denotes the asymptotic mean of $\hat{\alpha}$ when c=x. Similarly, $SD_x(\hat{\alpha})$ denotes the asymptotic standard deviation of $\hat{\alpha}$ when c=x. Each curve is drawn from calculations at 100 equidistant points.