

Exact Distribution of the Least Squares Estimator
in a First-Order Autoregressive Model

By

Mukhtar M. Ali
Department of Economics
University of Kentucky
Lexington, KY 40506
E-mail: MMALI1@POP.UKY.EDU

Preliminary. Not to be quoted without
permission from the author.

February, 1996

Exact Distribution of the Least Squares Estimator in a First-Order Autoregressive Model

SUMMARY

This paper investigates the finite sample distribution of the least squares estimator of the autoregressive parameter in a first-order autoregressive model. Uniform asymptotic expansion for the distribution applicable to both stationary and nonstationary cases is obtained. Accuracy of approximation to the distribution by a first few terms of this expansion is then investigated. It is found that the leading term of this expansion approximates well the distribution. The approximation is, in almost all cases, accurate to the second decimal place throughout the distribution. Only rarely the accuracy improves by including further term beyond the first term of this expansion in the approximation. As a matter of fact, often the accuracy of such an approximation with additional term(s) deteriorates. An application of the finding is illustrated with examples.

Exact Distribution of the Least Squares Estimator in a First-Order Autoregressive Model

1. INTRODUCTION

Consider the first-order autoregressive model

$$(1.1) \quad y_t = \beta y_{t-1} + \epsilon_t, \quad t = 1, 2, \dots, n$$

where β is a real constant and the errors ϵ_t 's are identically independently distributed normal variables each with mean = 0 and variance = σ_ϵ^2 . The initial observation, y_0 can be treated as fixed or stochastic. When y_0 is stochastic, it is assumed to have a normal distribution with mean 0 and variance $\sigma_\epsilon^2/(1 - \beta)$, $|\beta| < 1$, and is independent of ϵ_t 's. When y_0 is fixed, the model is stationary. When y_0 is fixed, if $|\beta| < 1$, the model is stationary asymptotically. If $|\beta| = 1$, this is a well-known random walk model and if $|\beta| > 1$, the model is explosive. The random walk model implies unit root hypothesis. Recently, there has been an enormous interest in testing for unit root. They include, among others, Dickey (1976), Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984), Fuller (1976), Hasza and Fuller (1979), Perron and Phillips (1987), Phillips and Perron (1988), Schwert (1987) and Stock and Watson (1989). Diebold and Nerlove (1990) have given an excellent survey of works in this area.

The unknown parameter β is customarily estimated by its least squares estimator

$$(1.2) \quad \hat{\beta} = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2} .$$

Under the assumption that ϵ_t 's are normally distributed, this is also the maximum likelihood estimate of β . To make inference about β based on $\hat{\beta}$ one must know the distribution of $\hat{\beta}$.

Unfortunately, exact distribution of $\hat{\beta}$, in closed form is unknown. Asymptotically, it has a normal distribution (Mann and Wald, 1943) if $|\beta| < 1$, a Cauchy distribution (White, 1958) if $|\beta| > 1$ and a non-standard distribution (White, 1958, Rao, 1978) if $|\beta| = 1$. These asymptotic distributions can be used to approximate the distribution of $\hat{\beta}$. However, this approximation results in a nonsmooth transformation from a normal distribution to a non-standard distribution to a Cauchy distribution which is, to say the least nonintuitive. In addition, unless β is close to zero, these asymptotic distributions do not approximate well the true distribution in finite samples (Evans and Savin, 1981; Tsui, 1989). The non-standard limiting distribution when $|\beta| = 1$ (Rao, 1978) seems to approximate well the true distribution when $|\beta|$ is close to 1, but it is too complicated for practical use. An accurate approximation to this limiting distribution can, however be obtained from the asymptotic expansion given by Abadir (1993). Thus, the asymptotic distribution of $\hat{\beta}$ is not much of help in making inference about β .

The distribution of $\hat{\beta}$ is not known in closed form but it can be obtained numerically. The exact distribution has been numerically computed by Phillips (1977, 1978), Evans and Savin (1981) for selected values of sample size n and the autoregressive coefficient β , and comprehensively for a wide range of (n, β) by Tsui (1989) and Tsui and Ali (1994). However, these numerical approaches are often computationally demanding and expensive even with modern high speed computers. Besides the numerical methods, several authors (Dickey, 1976; Fuller, 1976; Perron, 1989) have performed Monte Carlo experiments to tabulate the distribution in the case of $\beta = 1$ and $y_0 = 0$. These distributions can be used to test for unit root hypothesis but are not much of use for further inference

about β .

As an alternative to numerically computing (or simulating) the exact distribution, several authors have attempted to obtain convenient approximations to the distribution of $\hat{\beta}$ in finite samples. Phillips (1977, 1978), Satchell (1984), and Tsui and Ali (1992) have examined approximations by Edgeworth expansion and found them, except for β close to 0 and only at the center of the distribution, to be unsatisfactory. Tsui and Ali (1992) examined also approximations by Cornish-Fisher-type (Cornish and Fisher, 1937; Fisher and Cornish, 1960; Hill and Davis, 1968) expansions and the four-parameter Pearson distributions. Accuracy of these approximations was found to depend substantially on sample size and the values of the autoregressive coefficient. None of these approximations was found to be reliable when the autoregressive coefficient is moderately large and the sample size is small. Phillips (1978) derived a saddlepoint approximation to the probability density function of $\hat{\beta}$ when $|\beta| < 1$ and the initial observation is stochastic. An approximation to the distribution can be obtained by numerically integrating this approximate density function. He found this approximation to be exceptionally accurate, certainly for sample sizes as large as 30. Unfortunately, the approximation was not defined over a sizeable region of the tail for values of the autoregressive parameter greater than 0.4. Lieberman (1994) derived an alternative saddlepoint approximation to the probability density of $\hat{\beta}$ which is, however available over the entire range of $\hat{\beta}$. From an illustrative check for the accuracy of the distribution function derived from this approximate density, the approximation was found, for the model with $|\beta| < 1$ and y_0 to be stochastic, to be excellent for both $n = 10$ and $n = 30$ and for all four β values of .2, .4, .6 and .8 that were examined. The approximation seems promising but its accuracy has been tested for only

a few values of n and of β and for the model where the initial observation is stochastic. Furthermore, this approximation can be computationally demanding (even to the point of being impractical), especially for large sample sizes. This is because a crucial step in the implementation of the approximation is to obtain a solution of a highly non-linear equation which requires either a computation of eigenvalues of an $n \times n$ matrix or repeated inversion of $n \times n$ matrices. Moreover, expensive numerical integration may be needed to obtain the distribution function from the approximate density function.

In this paper, a uniform asymptotic expansion for the distribution (not the density function) of the least squares estimator of β is obtained. This expansion is applicable for both the stationary and non-stationary cases. For ease of exposition, these cases are identified with three models, each corresponding to a different nature of y_0 . When y_0 is fixed at 0, it is named Model A; and Model B when y_0 is fixed at a non-zero real constant; and Model C when y_0 is $N(0, \sigma_\epsilon^2/(1 - \beta^2))$ and $|\beta| < 1$. An alternative expression of the joint characteristic function of the numerator and denominator of $\hat{\beta}$ as a product in trigonometric functions avoids the costly computation of eigenvalues or repeated inversion of high order matrices that requires in implementing Lieberman (1994) approximation. Accuracy of approximation to the distribution by a first few terms of this expansion is investigated for a wide range of values of n and of β and for all the three models. It is found that the very first term of this expansion approximates well the distribution, especially at the extreme tails. The approximation is, in almost all cases, accurate to the second decimal place throughout the distribution. Only rarely the accuracy improves by including further term beyond the first term of this expansion in the approximation. As a matter of fact, often the accuracy of such an approximation with additional term(s) deteriorates.

The plan of this study is as follows. The uniform asymptotic expansion for the distribution of $\hat{\beta}$ is obtained in section 2. The accuracy of the expansion to the distribution is examined in section 3. Examples illustrating the use of the results of this paper are reported in section 4. Some concluding remarks are given in section 5.

2. ASYMPTOTIC EXPANSION

There has been some confusion in the literature in regards to the sample size n . Following Hurwicz (1950), the sample size is taken as the number of stochastic y_t in (1.1). Thus, the sample size is n in Models A and B where y_0 is fixed and it is $(n+1)$ in Model C where y_0 is stochastic. The sample size in Model C would be n if (1.1) is considered for $t = 2, 3, \dots, n$ where y_1 is the initial y following $N(0, \sigma_\epsilon^2/(1 - \beta^2))$. To be consistent in sample size of n for all three models, for Model C, (1.1) is taken for $t = 2, 3, \dots, n$. Thus, for all three models, the sample size is n and the sample is y_1, y_2, \dots, y_n . From a given sample $y = (y_1, y_2, \dots, y_n)$, the least squares estimator of β is then P/Q , for Model A and Model C and $(P + y_0 y_1)/(Q + y_0^2)$ for Model B, where

$$(2.1) \quad P = \sum_{t=2}^n y_{t-1} y_t \quad \text{and} \quad Q = \sum_{t=2}^n y_{t-1}^2 .$$

For uniformity and convenience, we have taken the least squares estimator for all the three models (A, B and C) as $\hat{\beta} = P/Q$.

Let $\psi(iu, iv) = E(e^{iuP+ivQ})$ be the joint characteristic function of P and Q . Then, by Theorem 1 of Gurland (1948, p. 229), the cumulative distribution function of $\hat{\beta}$ is given as

$$(2.2) \quad G(w) = \Pr(\hat{\beta} < w) = \Pr(P/Q < w)$$

$$= \frac{1}{2} - \frac{1}{2\pi i} \lim_{\delta \rightarrow 0} \left(\int_{-\infty}^{-\delta} + \int_{\delta}^{\infty} \right) \Psi(iu, -iuw) \frac{du}{u}$$

A change of variable (new $u = \text{old } ui$) in (2.2) leads to

$$(2.3) \quad G(w) = \frac{1}{2} - \frac{1}{2\pi i} \lim_{\delta \rightarrow 0} \int_L \Psi(u, -uw) \frac{du}{u}$$

where L is a path of integration made up of two segments: from $-\infty i$ to $-\delta i$, and from δi to ∞i . It can be shown that the integrand has a simple pole at $u = 0$ and $\Psi(0, 0) = 1$, so that, by Cauchy's integral formula,

$$(2.4) \quad \frac{1}{2\pi i} \int_C \Psi(u, -uw) \frac{du}{u} = 1,$$

where C is any closed curve encircling no singularity other than $u = 0$, in the positive (counterclockwise) direction. Suppose that the curve C is a circle of radius δ , then we can rewrite (2.3) as

$$(2.5) \quad G(w) = 1 - \frac{1}{2\pi i} \int_R \Psi(u, -uw) \frac{du}{u},$$

where the new path of integration R is obtained by adding that half of the circle C for which $\text{Re}(u) > 0$ to the original path L . It can be shown that the integrand is analytic for $\text{Re}(u) > 0$. Thus, by the well known theorem of Cauchy, the path of integration R can be modified to obtain

$$(2.6) \quad G(w) = 1 - \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Psi(u, -uw) \frac{du}{u}, \quad c > 0.$$

It can be seen that the integrand in (2.6) has a simple pole at $u = 0$ and following Lieberman (1994), it can be shown that the integrand has only one saddlepoint. Defining, $h(u) = -(1/n) \ln \Psi(u, -uw)$, the integral in (2.6) is, then exactly in the form of that in equation (65) in Rice (1968). Hence, from equation (68) of Rice (1968), the uniform asymptotic expansion for $G(w)$ is given by

$$(2.7) \quad G(w) = \Phi(x\sqrt{2n}) - \phi(x\sqrt{2n}) \sum_{j=0}^{\infty} p_j n^{-j},$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are, respectively the distribution and probability density function of a standard normal variable,

$$p_j = \frac{(-1)^j \left(\frac{1}{2}\right)_j}{x^{2j+1}} \left\{ \left(\frac{x}{z}\right)^{2j+1} \left[\sum_{i=0}^{2j} (-u_1)^i \sum_{k=0}^i b_{ki} \left(j + \frac{1}{2}\right)_k \right] - 1 \right\}, \quad j \geq 0,$$

$$u_1 \text{ is the saddlepoint, } h(0) = 0, \quad h^{(r)}(u) = \frac{\partial^r h}{\partial u^r}, \quad r \geq 0, \quad h^{(0)} = h, \quad h_1^{(r)} = h^{(r)}(u_1),$$

$$h_1 = h(u_1), \quad h^{(1)}(u_1) = 0, \quad x = (-h_1)^{(1/2)} u_1 / |u_1|, \quad z = u_1 (h_1^{(2)}/2)^{(1/2)}, \quad (q)_0 = 1,$$

$$(q)_j = q(q+1) \dots (q+j-1), \quad b_{00} = 1, \quad b_{0i} = 0, \quad \text{for } i \geq 1, \quad \text{and}$$

$$b_{k+1,i+1} = \frac{1}{i+1} \sum_{s=1}^{i-k+1} d_s b_{k, i-s+1},$$

$$d_s = -2h_1^{(s+2)} / [(s+2)! h_1^{(2)}], \quad s \geq 1.$$

In deriving the expansion in (2.7), it is assumed that the origin, at which there is a simple pole for the integrand in (2.6), does not coincide with the saddlepoint. If the saddlepoint is at the origin, a classical saddlepoint analysis (Lugannani and Rice, 1980, p. 479) can be applied to obtain an asymptotic expansion for $G(w)$ as

$$(2.8) \quad G(w) = \frac{1}{2} - (2\pi n)^{-\frac{1}{2}} \left[-\theta_3 + n^{-1} \left(\frac{35}{2} \theta_3^3 - 15\theta_3 \theta_4 + 3\theta_5 \right) \right. \\ \left. - \frac{15}{2} n^{-2} \left(\frac{3003}{20} \theta_3^5 - 231\theta_3^3 \theta_4 + 63\theta_3 \theta_4^2 + 63\theta_3^2 \theta_5 \right) \right. \\ \left. - 14\theta_4 \theta_5 - 14\theta_3 \theta_6 + 2\theta_7 \right] + \dots$$

where $\theta_i = h^{(i)}(0) / (i! [h^{(2)}(0)]^{(i/2)})$, $i = 3, 4, \dots$

3. ACCURACY OF THE EXPANSION

The distribution can be computed numerically from (2.7) (or (2.8)). A major problem is the evaluation of $\psi(u, -uw)$ and hence of the function $h(u)$ that appears in the integrand in (2.6). Lieberman (1994) obtained $\psi(u, -uw)$ as a determinant of an $n \times n$ matrix depending on w . He also expressed it as an elementary function of eigenvalues of the same $n \times n$ matrix. For either case, as it requires determinant or eigenvalues of an $n \times n$ matrix for each w , the computation becomes prohibitively time consuming as the sample size n becomes large, say larger than 200. From White (1958), an expression for $\psi(u, -uw)$ can be derived in closed form. It, however involves in raising some (real) expressions to the power of sample size n . Thus, for large sample sizes, it becomes problematic to maintain reasonable numerical accuracy. Alternatively, following White (1961), $\psi(u, -uw)$ can be expressed as a polynomial of degree n , the sample size. Unfortunately, this creates computational problems in large sample sizes because this polynomial of high degree often contains a large number of terms each of which is negligible individually but significant collectively. The expression for $\psi(u, -uw)$ that is found to be most convenient is the one obtained by Tsui (1989) and Tsui and Ali (1994). This expression is in trigonometric functions and is given as

$$\begin{aligned}
 (3.1) \quad \psi(u, -uw) &= D_n(u, -uw)^{-1/2}, \text{ for Model A} \\
 &= \exp\left(-\frac{1}{2}\beta^2\alpha^2\left(1 - \frac{D_{n-1}(u, -uw)}{D_n(u, -uw)}\right)\right) D_n(u, -uw)^{-1/2}, \text{ for Model B} \\
 &= \sqrt{(1 - \beta^2)[D_n(u, -uw) - \beta^2 D_{n-1}(u, -uw)]}^{-1/2}, \text{ for Model C}
 \end{aligned}$$

where $\alpha = y_0/\sigma_e$,

$$(3.2) \quad D_n(u, -uw) = S_{n-1}(u, -uw) - (\beta + u)^2 S_{n-2}(u, -uw), \text{ and}$$

$$(3.3) \quad S_n(u, -uw) = \prod_{i=1}^n \left\{ (1 + \beta^2 + 2uw) - 2(\beta + u) \cos\left(\frac{\pi i}{n + 1}\right) \right\}.$$

Using expression (3.1), it is a routine matter to compute $h(u)$ and its various derivatives that are required to obtain the distribution function from (2.7) (or, (2.8)). In this computations, the basic function to evaluate is $S_n(u, -uw)$ and its derivatives. As $S_n(u, -uw) = \text{Sgn}[S_n(u, -uw)]|S_n(u, -uw)|$, where $\text{Sgn}(x) = 1$, if $x \geq 0$ and $= -1$, if $x < 0$, the function $S_n(u, -uw)$ and its derivatives can be obtained from those of $|S_n(u, -uw)|$ and its derivatives. In turn, the function $|S_n(u, -uw)|$ and its derivatives may be evaluated utilizing the relation

$$(3.3) \quad \ln |S_n(u, v)| = \sum_{i=1}^n \ln \left| \left(1 + \beta^2 + 2uw \right) - 2(\beta + u) \cos \left(\frac{\pi i}{n + 1} \right) \right| .$$

It may be noted that the saddlepoint u_1 is a solution to the equation

$$(3.4) \quad h^{(1)}(u_1) = 0.$$

The equation (3.4) can be solved by the iterative Newton-Raphson method starting at $u_1 = 0$ and modifying, if necessary the step size at each iteration so that $\psi(u, -uw)$ is positive.

The expansion (2.7) (or, (2.8)) can be used to compute the distribution function, $G(w)$. Unfortunately, most asymptotic expansions are nonconvergent, with the magnitude of successive terms tracking a J curve of initial decline followed by a steep rise. Fortunately, often a few of the beginning terms provide adequate approximation. To check for the accuracy of approximation to the distribution of $\hat{\beta}$ by a few terms in the expansion (2.7) (or, (2.8)), we consider three

approximations, APX-1, APX-2 and APX-3 which are obtained by truncating the expansion (2.7)

(or (2.8)) at the leading term, the second term and the third term, respectively. More specifically,

$$(3.5) \quad \text{APX-1: expansion (2.7) truncated at } j = 0 \text{ (or, expansion (2.8) truncated to the terms at most of order } O(n^{-1/2}))$$

$$(3.6) \quad \text{APX-2: expansion (2.7) truncated at } j = 1 \text{ (or, expansion (2.8) truncated to$$

the terms at most of order $O(n^{-3/2})$)

(3.7) APX-3: expansion (2.7) truncated at $j = 2$ (or, expansion (2.8) truncated to the terms at most of order $O(n^{-5/2})$)

To check for the accuracy of these approximations, the exact distributions obtained by evaluating numerically the integral in (2.2) (see Tsui (1989) and Tsui and Ali (1994)) are compared with these approximations. A fortran program is written to implement these approximations and can be obtained on request. For a thorough investigation, the distributions and the approximations were computed comprehensively for Models A, B and C. In all cases, we take various sample size $n = 10, 20, 30, 40, 50, 75, 100, 150, 200, 250, 300$ and 500 . For Models A and B, choices of autoregressive coefficient, β include $0.4, 0.6, 0.8, 0.9, 0.95, 0.99, 1.0, 1.01$. For Model B, the values of the parameter α are: 1 and 4 . As for Model C, choices of autoregressive coefficient, β are: $0.4, 0.6, 0.8, 0.9, 0.95, 0.975, 0.99$. In each case, the distribution $G(w)$ is computed at 33 percentile points, $w = x/g(n) + \beta$ ($x = -16.0, -12.0, -8.0, -6.0, -4.0, -3.5, -3.0, -2.8, -2.6, -2.4, -2.2, -2.0, -1.8, -1.6, -1.4, -1.2, -1.0, -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0$ and 4.0) where $g(n) = \sqrt{n/(1 - \beta^2)}$, if $|\beta| < 1$, $= n/\sqrt{2}$, if $|\beta| = 1.0$ and $= |\beta|^n/(\beta^2 - 1)$, if $|\beta| > 1$. It may be worth observing that in this experimentation, we have taken only positive values of β . This is because the distribution of $\hat{\beta}$ for a given β is the mirror-image of that for $-\beta$ (see Cryer, Nankervis and Savin, 1989).

Comparing the approximations with the exact distributions (all the exact and approximate distributions that are computed can be obtained upon request), some definite conclusions can be drawn. In particular, it is found that, for all cases considered, APX-1 approximates exceptionally

well, especially at the tails of the distributions where most of the inferential interests lie. Only rarely APX-2 and APX-3 improves over APX-1 and often the accuracy of APX-2 and APX-3 deteriorates compared to that of APX-1. Without any exception the accuracy of the approximations (APX-1, APX-2 or APX-3) deteriorates as β approaches 1 (larger than or smaller than 1) and improves as sample size, n increases or the parameter α in Model B (Model A is equivalent to Model B with $\alpha = 0$) increases. Thus, in these comparisons, the accuracy of these approximations is worst for $n = 10$, $\beta = 0.99, 1.0$ and 1.01 and $\alpha = 0$ in the case of Models A and B; and for $n = 10$ and $\beta = 0.99$ in the case of Model C. To support our general conclusions and in particular to show the remarkable accuracy provided by the approximation APX-1, we display the exact distributions and their approximations for some worst case scenarios, namely for $n = 10$ and $\beta = 1.0$ and 1.01 in the case of Model A (see Table 1) and for $n = 10$ and $\beta = 0.95$ and 0.99 in the case of Model C (see Table 2).

4. ILLUSTRATIVE EXAMPLES

Preceding analysis suggests that, for all practical purposes, the distribution of the least squares estimator $\hat{\beta}$, $G(w) = \Pr(\hat{\beta} < w)$ can be well approximated by the first term in the expansion (2.7) (or, (2.8)), specifically by

$$(4.1) \quad AG(w) = \Phi\left(\frac{w - \beta}{x\sqrt{2n}}\right) - \phi\left(\frac{w - \beta}{x\sqrt{2n}}\right) \left(\frac{1}{z} - \frac{1}{x}\right), \text{ if the saddlepoint } u_1 \neq 0$$

$$= 0.5 + (2\pi n)^{-1/2}\theta_3, \text{ if the saddlepoint } u_1 = 0$$

where u_1 is the saddlepoint satisfying $h^{(1)}(u) = 0$, $h(u)$ is as defined preceding the equation (2.7), $h_1 = h(u_1)$, $x = (-h_1)^{(1/2)}u_1/|u_1|$, $z = u_1(h_1^{(2)}/2)^{(1/2)}$, and $\theta_3 = h^{(3)}(u_1)/(6[h^{(2)}(u_1)]^{3/2})$. The approximation is applicable for all the models A, B and C and for all possible values of the parameter β ($|\beta| < 1$, $|\beta|$

$= 1$, or $|\beta| > 1$).

The approximation $AG(w)$ can then be used to make inference about the parameter β . In most applications, it is of interest to know whether $|\beta| < 1$, or, $|\beta| = 1$, or $|\beta| > 1$. Such inference can be made from an appropriate confidence interval estimate for β . These confidence interval estimates can be obtained using the approximation $AG(w)$. It can be shown that the lower and upper limit of the central $100(1 - \gamma)\%$ confidence interval are, respectively the solutions of

$$(4.2) \quad AG(\hat{\beta}_L) = 1 - \gamma/2, \text{ and}$$

$$(4.3) \quad AG(\hat{\beta}_U) = \gamma/2$$

where $\hat{\beta}$ is the least squares estimate of β . These equations can be solved by any of a variety of successive approximation methods (see Abramowitz and Stegun, 1970, p. 18). A fortran program which can be obtained on request is written to solve these equations using the Newton's Rule. To start the approximation method, both the lower and upper limit can be taken to be the estimate $\hat{\beta}$.

Confidence interval is a convenient tool to make inference. However, if there is a specific inference needs to be made, it may be more convenient to make such inference by testing specific hypothesis. Thus, if one is interested to know whether there is a root $= 1$ in the autoregressive polynomial or not, one may test the hypothesis $H_0: \beta = 1$, against the alternatives $H_A: \beta < 1$. We would reject H_0 if the p-value given by $\Pr(\hat{\beta} < \text{sample } \hat{\beta} \mid \beta = 1)$ is smaller than the level of significance. Using our approximation $AG(w)$ to the distribution of $\hat{\beta}$, this p-value can be computed.

To illustrate the use of these results, we have analyzed three sets of data generated from the Model A. In each case, we have chosen $n = 25$ and $\sigma_\epsilon = 1$. The first data set is generated from the model with $\beta = .95$ (asymptotically stationary model), the second with $\beta = 1$ (random-walk model)

and the third with $\beta = 1.05$ (explosive model). The three data sets, estimate $\hat{\beta}$ based on these data and the central 95% confidence interval estimates are found as follows.

Data Set 1: 0.86, 1.26, 2.39, 2.60, 2.81, 4.15, 3.36, 1.25, 1.17, 0.16, -0.09, 0.54, -0.57

-2.62, -3.10, -1.30, 0.19, 1.56, 1.60, 1.49, 3.62, 3.96, 3.03, 2.49, 3.64

$\hat{\beta} = 0.930$; 95% confidence interval: [0.772, 1.186]

The p-value, in testing $H_0: \beta = 1$ against $H_a: \beta < 1$, is 0.371

Data Set 2: 0.86, 1.31, 2.50, 2.82, 3.16, 4.64, 4.06, 2.12, 2.11, 1.15, 0.91, 1.54, 0.45,

-1.62, -2.23, -0.59, 0.83, 2.21, 2.33, 2.30, 4.51, 5.03, 4.29, 3.91, 5.18

$\hat{\beta} = 0.991$; 95% confidence interval: [0.874, 1.204]

The p-value, in testing $H_0: \beta = 1$ against $H_a: \beta < 1$, is 0.626

Data Set 3: 0.86, 1.35, 2.61, 3.06, 3.56, 5.22, 4.89, 3.20, 3.34, 2.56, 2.44, 3.19, 2.27

0.31, -0.29, 1.33, 2.83, 4.35, 4.68, 4.89, 7.34, 8.23, 7.91, 7.91, 9.58

$\hat{\beta} = 1.066$; 95% confidence interval: [0.989, 1.222]

The p-value, in testing $H_0: \beta = 1$ against $H_a: \beta < 1$, is 0.965

Based on these confidence intervals and p-values, one would not reject the hypothesis of unit root in all these three cases. However, for data set 2, unit root is located near to the center of this interval; for data set 1, it is close to the upper limit of the interval and for data set 3, it is close to the lower limit of the interval. This evidence may be interpreted to mean that the model generating the data set 1 is likely to have root (the parameter β) less than 1, that generating the data set 2 is likely to have unit root and that generating the data set 3 to have root larger than 1.

5. CONCLUDING REMARKS

Autoregressive models have been found to be most prominent among models in describing

time movement of a time series variable. Inference on the parameters of these models are then of utmost interest to the practitioners. Often, such inference has been based on the least squares estimators. As the exact distributions of these least squares estimators are rarely known, the practitioners have been forced to rely on the known asymptotic distributions for inferential purposes. There are at least two disadvantages to such procedures. First, the asymptotic distributions are either too complicated for practical use or provide impractical poor approximations to exact distributions, especially for sample sizes that are usually available in practice. Second, the form of the asymptotic distributions depend upon the parameter values. More specifically, asymptotic distribution that is appropriate when all the roots of the autoregressive polynomial are less than 1 in magnitude is different from the one that is appropriate if at least one of the roots is equal to 1, and these asymptotic distributions are, in turn different from the one appropriate when some of the roots are greater than 1 in magnitudes. This nonsmooth transformation of the asymptotic distribution from the case when the roots are less than 1 to the case when some roots are equal to 1 to the case when some roots are greater than 1 is, to say the least counterintuitive and it makes it essential prerequisite to have the knowledge of any root of the autoregressive polynomial to be equal to 1 or greater than 1 before one can use the appropriate asymptotic distribution for inference.

This paper considers the autoregressive model of order one. An uniform asymptotic expansion for the distribution of the least squares estimator for the autoregressive coefficient is derived. This is a valid expansion irrespective of the size of the root of the autoregressive polynomial. Unfortunately, this asymptotic expansion is, like most asymptotic expansions nonconvergent. Fortunately, however, it is found, after a detailed investigation that the leading term of the expansion provides excellent approximation to the exact distribution. Thus, this approximation

can be used, for all practical purposes to make inference. To implement this approximation, one is required to evaluate a certain characteristic function and its first two derivatives. A convenient closed-form expression for the characteristic function is given which should facilitate such computations. These results are illustrated with examples in obtaining p-values and constructing confidence intervals for the autoregressive parameter using this approximation.

Table 1: Cumulative Distribution of $\hat{\beta}$, $G(w) = \Pr(\hat{\beta} < w)$ for Model A, $n = 10$

x	$\beta = 1.0$				$\beta = 1.01$			
	EXACT	APX-1	APX-2	APX-3	EXACT	APX-1	APX-2	APX-3
-16.0	0.0000	0.0000	0.0000	0.0000	0.2160	0.2333	0.2169	0.2034
-12.0	0.0001	0.0001	0.0001	0.0001	0.2861	0.3045	0.2930	0.2755
-8.0	0.0047	0.0050	0.0046	0.0047	0.3800	0.3943	0.3906	0.3754
-6.0	0.0208	0.0223	0.0203	0.0211	0.4382	0.4482	0.4480	0.4344
-4.0	0.0730	0.0791	0.0705	0.0739	0.5032	0.5096	0.5121	0.5012
-3.5	0.0975	0.1059	0.0942	0.0975	0.5205	0.5263	0.5293	0.5187
-3.0	0.1293	0.1405	0.1257	0.1265	0.5382	0.5435	0.5471	0.5381
-2.8	0.1445	0.1570	0.1410	0.1399	0.5455	0.5506	0.5543	0.5469
-2.6	0.1614	0.1753	0.1584	0.1545	0.5528	0.5577	0.5616	0.5569
-2.4	0.1802	0.1954	0.1780	0.1710	0.5602	0.5649	0.5689	0.5689
-2.2	0.2011	0.2175	0.2002	0.1897	0.5678	0.5723	0.5763	0.5849
-2.0	0.2243	0.2417	0.2252	0.2115	0.5754	0.5797	0.5838	0.6092
-1.8	0.2501	0.2683	0.2533	0.2370	0.5831	0.5873	0.5870	1.1463
-1.6	0.2789	0.2974	0.2845	0.2669	0.5909	0.5948	0.5956	1.1685
-1.4	0.3111	0.3292	0.3191	0.3013	0.5988	0.6025	0.6035	1.3195
-1.2	0.3471	0.3640	0.3571	0.3401	0.6068	0.6103	0.6101	1.9005
-1.0	0.3876	0.4022	0.3984	0.3828	0.6149	0.6182	0.6131	4.2190
-0.8	0.4328	0.4441	0.4433	0.4292	0.6230	0.6261	0.6031	*(>10)
-0.6	0.4826	0.4903	0.4920	0.4812	0.6313	0.6342	0.5265	*(>10)
-0.4	0.5361	0.5414	0.5450	0.5357	0.6396	0.6428	-.1701	*(>10)
-0.2	0.5939	0.5976	0.5989	1.1909	0.6480	0.6558	*(>10)	*(>10)
0.0	0.6566	0.6586	0.6646	0.6550	0.6565	0.6585	0.6645	0.6550
0.2	0.7227	0.7226	0.7304	0.6487	0.6650	0.6624	*(>10)	*(>10)
0.4	0.7883	0.7859	0.7946	0.7874	0.6736	0.6745	1.3516	*(>10)
0.6	0.8475	0.8433	0.8531	0.8468	0.6823	0.6833	0.7737	*(>10)
0.8	0.8952	0.8904	0.8999	0.8944	0.6910	0.6917	0.7169	*(>10)
1.0	0.9300	0.9255	0.9336	0.9292	0.6997	0.7002	0.7116	-.5547
1.2	0.9536	0.9500	0.9562	0.9530	0.7085	0.7087	0.7148	0.9275
1.4	0.9691	0.9664	0.9709	0.9687	0.7173	0.7172	0.7191	1.5091
1.6	0.9792	0.9772	0.9804	0.9790	0.7260	0.7256	0.7338	0.6746
1.8	0.9858	0.9843	0.9866	0.9857	0.7348	0.7340	0.7422	0.7135
2.0	0.9902	0.9891	0.9907	0.9901	0.7435	0.7424	0.7508	0.7339
4.0	0.9995	0.9995	0.9995	0.9995	0.8265	0.8224	0.8327	0.8260

$x = (w - \beta)g(n)$, $g(n) = \sqrt{n/(1 - \beta^2)}$, if $|\beta| < 1$, $= n/\sqrt{2}$, if $|\beta| = 1$ and $= |\beta|^n/(\beta^2 - 1)$, if $|\beta| = 1$; APX-1, APX-2 and APX-3 are, respectively the approximations from the leading one, two and three terms in the expansion (2.7) or (2.8).

* the number is greater than 10 in absolute value.

Table 2: Cumulative Distribution of $\hat{\beta}$, $G(w) = \Pr(\hat{\beta} < w)$ for Model C, $n = 10$

x	$\beta = 0.95$				$\beta = 0.99$			
	EXACT	APX-1	APX-2	APX-3	EXACT	APX-1	APX-2	APX-3
-16.0	0.0001	0.0001	0.0001	0.0001	0.0086	0.0094	0.0083	0.0088
-12.0	0.0012	0.0013	0.0012	0.0012	0.0209	0.0232	0.0199	0.0215
-8.0	0.0123	0.0131	0.0120	0.0124	0.0505	0.0566	0.0481	0.0495
-6.0	0.0331	0.0357	0.0322	0.0336	0.0803	0.0899	0.0790	0.0744
-4.0	0.0852	0.0923	0.0824	0.0854	0.1332	0.1450	0.1376	0.1261
-3.5	0.1077	0.1168	0.1046	0.1062	0.1530	0.1642	0.1592	0.1469
-3.0	0.1365	0.1479	0.1336	0.1313	0.1771	0.1868	0.1848	0.1720
-2.8	0.1502	0.1625	0.1477	0.1430	0.1882	0.1971	0.1965	0.1835
-2.6	0.1654	0.1787	0.1637	0.1560	0.2003	0.2082	0.2092	0.1961
-2.4	0.1822	0.1964	0.1817	0.1708	0.2136	0.2203	0.2231	0.2099
-2.2	0.2009	0.2159	0.2019	0.1877	0.2283	0.2335	0.2382	0.2252
-2.0	0.2218	0.2373	0.2248	0.2075	0.2446	0.2481	0.2550	0.2423
-1.8	0.2451	0.2609	0.2505	0.2306	0.2627	0.2643	0.2736	0.2616
-1.6	0.2713	0.2868	0.2793	0.2575	0.2832	0.2825	0.2945	0.2835
-1.4	0.3008	0.3154	0.3114	0.2884	0.3063	0.3031	0.3181	0.3086
-1.2	0.3341	0.3470	0.3470	0.3235	0.3328	0.3266	0.3449	0.3376
-1.0	0.3719	0.3820	0.3866	0.3628	0.3634	0.3539	0.3757	0.3712
-0.8	0.4146	0.4209	0.4304	0.4064	0.3992	0.3855	0.4113	0.4104
-0.6	0.4627	0.4644	0.4790	0.4554	0.4408	0.4226	0.4523	0.4564
-0.4	0.5165	0.5131	0.5328	0.5110	0.4891	0.4659	0.4994	0.5177
-0.2	0.5753	0.5673	0.5892	0.5894	0.5439	0.5157	0.5467	1.5441
0.0	0.6385	0.6267	0.6553	0.6353	0.6037	0.5709	0.6117	0.6289
0.2	0.7041	0.6896	0.7207	0.6758	0.6654	0.6294	0.6773	0.0854
0.4	0.7683	0.7526	0.7836	0.7581	0.7246	0.6874	0.7321	0.7474
0.6	0.8261	0.8109	0.8395	0.8169	0.7776	0.7411	0.7858	0.8027
0.8	0.8738	0.8605	0.8846	0.8658	0.8222	0.7881	0.8316	0.8443
1.0	0.9101	0.8995	0.9185	0.9039	0.8583	0.8277	0.8687	0.8760
1.2	0.9365	0.9283	0.9426	0.9322	0.8868	0.8601	0.8977	0.8997
1.4	0.9551	0.9490	0.9594	0.9523	0.9092	0.8864	0.9199	0.9174
1.6	0.9680	0.9635	0.9710	0.9663	0.9266	0.9075	0.9367	0.9311
1.8	0.9770	0.9737	0.9790	0.9759	0.9404	0.9244	0.9495	0.9420
2.0	0.9832	0.9809	0.9847	0.9826	0.9512	0.9380	0.9592	0.9510
4.0	0.9988	0.9987	0.9989	0.9988	0.9912	0.9891	0.9924	0.9905

$x = (w - \beta)g(n)$, $g(n) = \sqrt{n/(1 - \beta^2)}$, if $|\beta| < 1$, $= n/\sqrt{2}$, if $|\beta| = 1$ and $= |\beta|^n/(\beta^2 - 1)$, if $|\beta| = 1$; APX-1, APX-2 and APX-3 are, respectively the approximations from the leading one, two and three terms in the expansion (2.7) or (2.8).

REFERENCES

- Abadir, K. M. (1993). "The Limiting Distribution of the Autocorrelation Coefficient Under a Unit Root", The Annals of Statistics, 21, no. 2, 1058-70.
- Abramowitz, M. and I. A. Stegun (1970). Handbook of Mathematical Functions, Dover edition, ninth printing, New York: Dover Publications, Inc.
- Cornish, E. A. and Fisher, R. A. (1937). "Moments and Cumulants in the Specification of Distribution", Revue de l'institut Internat. de Statist., 4, 307-20.
- Cryer, J. D., J. C. Nankervis and N. E. Savin (1989). "Mirror-Image Distributions in AR(1) Models", Econometric Theory, 5, 36-52.
- Dickey, D. A. (1976). Estimation and Hypothesis Testing for Nonstationary Time Series, Unpublished Ph. D. Dissertation, Iowa State University, Ames, IW.
- Dickey, D. A. and W. A. Fuller (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", J. Amer. Statist. Assn., 74, 427-31.
- Dickey, D. A. and W. A. Fuller (1981). "Likelihood Ratio Test for Autoregressive Time Series with a Unit Root", Econometrica, 49, 1057-1072. Series
- Diebold, F. X. and M. Nerlove (1990). "Unit Roots in Economic Time Series: A Selective Survey", Advances in Econometrics, 8, 3-69.
- Evans, G. B. and N. E. Savin (1981). "Testing for Unit Roots: I", Econometrica, 49, 753-79.
- Evans, G. B. and N. E. Savin (1984). "Testing for Unit Roots: II", Econometrica, 52, 1241-69.
- Fisher, R. A. and E. A. Cornish (1960). "The Percentile Points of Distributions Having Unknown Cumulants", Technometrics, 2, 209-26.
- Fuller, W. A. (1976). Introduction to Statistical Time Series, New York: Wiley
- Gurland, J. (1948). "Inversion Formula for the Distribution of Ratios", Annals of Mathematical Statistics, 19, 228-37. Mathematical
- Hasza, D. P. and W. A. Fuller (1979). "Estimation for Autoregressive Processes with Unit Roots", Annals of Statistics, 7, 1106-20.
- Hill, G. W. and A. W. Davis (1968). "Generalized Asymptotic Expansions of Cornish-Fisher

- Type", Annals of Mathematical Statistics, 39, 1264-73.
- Hurwicz, L. (1950). "Least Squares Bias in Time Series", in Statistical Inference in Dynamic Models, ed. by T. C. Koopmans, New York: Wiley, 365-83.
- Lieberman, O. (1994). "Saddlepoint Approximation for the Least Squares Estimator in First-Order Autoregression", Biometrika, 81, no. 4, 807-11.
- Lugannani, R. and S. O. Rice (1980). "Saddle Point Approximation for the Distribution of the Sum of Independent Random Variables", Adv. Appl. Prob., 12, 475-90.
- Mann, H. B. and A. Wald (1943). "On the Statistical Treatment of Linear Stochastic Difference Equations", Econometrica, 11, 173-220.
- Perron, P. (1989). "The Calculation of the Limiting Distribution of the Least Squares Estimator in a Near-Integrated Model", Econometric Theory, 5, 241-55.
- Perron, P. and P. C. B. Phillips (1987). "Does GNP have a Unit Root? A Reevaluation", Economic Letters, 23, 139-45.
- Phillips, P. C. B. (1977). "Approximations to Some Finite Sample Distributions Associated with the First Order Stochastic Difference Equation", Econometrica, 45, 463-85.
- Phillips, P. C. B. (1978). "Edgeworth and Saddlepoint Approximations in a First Order Noncircular Autoregression", Biometrika, 59, 79-84.
- Phillips, P. C. B. and P. Perron (1988). "Testing for a Unit Root in Time Series Regression", Biometrika, 75, 335-46.
- Rao, M. M. (1978). "Asymptotic Distribution of an Estimator of the Boundary Parameter of an Unstable Process", Annals of Statistics, 16, 185-90.
- Rice, S. O. (1968). "Uniform Asymptotic Expansions for Saddle Point Integrals - Applications to a Probability Distribution Occurring in Noise Theory", Bell Syst. Tech. J., 47, 1971-

Satchell, S. E. (1984). "Approximations to the Finite Sample Distributions for Non-Stable First Order Difference Equations", Econometrica, 52, 1271-88.

Schwert, G. W. (1987). "Effects of Model Specification on Tests of Unit Roots in Macroeconomic Data", J. Monetary Econ., 20, 73-103.

Stock, J. H. and W. W. Watson (1989). "Interpreting the Evidence on Money-Income Causality", J. of Econometrics, 40, 161-81.

Tsui, A. K. (1989). On the Finite Sample Distribution of a Least Squares Estimator in a First Order Autoregressive Model, Unpublished Ph. D. Dissertation, University of Kentucky, Lexington, ,

K e
n t u
c k y
.

Tsui, A. K. and M. M. Ali (1992). "Approximations to the Distribution of the Least Squares

Est
ima
tor
in a
Fir
s t
O r
der
Sta
tio
nar
y
Aut
ore
gre
ssi
v e
Mo
del
" ,
C o
m
m u
n i c
a t i o
n s
i n
S t a
t i s t
i c s
=
S i
m u
l a t i
o n ,
21 ,
no.
2 ,

Tsui, A. K. and M. M. Ali (1994). "Exact Distributions, Density Functions and Moments of

the
Lea
s t
S q
uar
e s
Est
ima
tor
in a
Fir
st -
O r
der
Aut
ore
gre
ssi
ve
Mo
del
" ,
C o
m p
ut at
ion
a l
Sta
tist
ics
&
Dat
a
An
aly
sis,
17,
4 3
3 -

White, J. S. (1958). "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case", Annals of Mathematical Statistics, 29, 1188-97.

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