The Annals of Statistics
1991, Vol. 19, No. 3, 1109–1128

MAXIMUM LIKELIHOOD TYPE ESTIMATION FOR NEARLY NONSTATIONARY AUTOREGRESSIVE TIME SERIES

BY DENNIS D. COX[1] AND ISABEL LLATAS

University of Illinois, Champaign-Urbana and Universidad Simon Bolivar

The nearly nonstationary first-order autoregression is a sequence of autoregressive processes $y_n(k+1) = \phi_n y_n(k) + \varepsilon(k+1), 0 \le k \le n$, where the $\varepsilon(k)$'s are iid mean zero shocks and the autoregressive coefficient $\phi_n = 1 - \beta/n$ for some $\beta > 0$, so that $\phi_n \to 1$ as $n \to \infty$. We consider a class of maximum likelihood type estimators called M estimators, which are not necessarily robust. The estimates are obtained as the solution $\hat{\phi}_n$ of an equation of the form

$$\sum_{k=0}^{n-1} y_n(k)\psi(y_n(k+1) - \phi y_n(k)) = 0,$$

where ψ is a given "score" function. Assuming the shocks have $2 + \delta$ moments and that ψ satisfies some regularity conditions, it is shown that the limiting distribution of $n(\hat{\phi}_n - \phi_n)$ is given by the ratio of two stochastic integrals. For a given shock density f satisfying regularity conditions, it is shown that the optimal ψ function for minimizing asymptotic mean squared error is not the maximum likelihood score in general, but a linear combination of the maximum likelihood score and least squares score. However, numerical calculations under the constraint $y_n(0) = 0$ show that the maximum likelihood score has asymptotic efficiency no lower than 40%.

1. Introduction. The aim of this work is to study asymptotic properties of a class of estimators of the autoregressive parameter ϕ of a nearly nonstationary first-order autoregressive process, and to obtain efficient estimators of ϕ. The class of estimators considered are those obtained by solving nonlinear equations including likelihood equations. We refer to them as "M estimators," but they should not be confused with robust M estimators since robustness is not one of our concerns. We consider the sequence $\{y_n(k): 0 \le k \le n\}_{n=1}^{\infty}$ of first-order autoregressive AR(1) processes given by

$$y_n(k) = \phi_n y_n(k-1) + \varepsilon(k), \tag{1.1}$$

where $\{\varepsilon(k)\}_{k=-\infty}^{\infty}$ is a sequence of iid random variables with mean 0, variance σ^2 and finite $(2+\delta)$-moment for some $\delta > 0$. ϕ_n is allowed to vary with n.

Received May 1987; revised April 1990.
[1]Research supported by Office of Naval Research Contract N00014-84-C-0169.
AMS 1980 *subject classifications*. Primary 62M10; secondary 62F12, 62E20, 60F17.
Key words and phrases. Non-Gaussian time series, autoregressive processes, maximum likelihood estimation, asymptotic efficiency.

Specifically,

$$\phi_n = 1 - \frac{\beta}{n} \tag{1.2}$$

for some $\beta > 0$, so that y_n tends to look like a nonstationary random walk for large n. This is the so-called *nearly nonstationary* (NNS) autoregressive process. It has been investigated by Bobkoski (1983), Chan and Wei (1987), Jeganathan (1987), Phillips (1987) and Cox (1990). Also we will assume that we have some knowledge of the starting value $y_n(0)$, either by considering it as a constant or by assuming it is a random variable with known distribution. In principle we are interested in the asymptotic behavior of estimators of the form

$$\hat{\phi}_n = \arg\min_{\phi} \sum_{k=0}^{n-1} \rho(y_n(k+1) - \phi y_n(k)) \tag{1.3}$$

for some criterion function ρ. Here, $\arg\min_\phi$ denotes the value of ϕ where a minimum is achieved. Taking $\rho(u) = u^2$, (1.3) gives the least squares estimator (LSE) of ϕ.

It is well known that for fixed $\phi_n \equiv \phi \in (0, 1)$, the LSE converges to ϕ at rate $O_p(n^{-1/2})$ and is asymptotically $N(0, 1 - \phi^2)$, with the usual centering and scaling. But when $\phi = 1$ the LSE converges to 1 at rate $O_p(n^{-1})$ and the normal approximation fails [see, e.g., Fuller (1976), Section 8.5]. White (1958) was able to represent the asymptotic distribution of the estimation error when $\phi_n \equiv 1$ [i.e., $\beta = 0$ in (1.2)] as

$$n(\hat{\phi}_n - 1) \Rightarrow \frac{\int_0^1 W(s)\, dW(s)}{\int_0^1 W^2(s)\, ds},$$

where W denotes a standard Brownian motion process and $\Rightarrow$ denotes weak convergence. Rao (1978), Dickey and Fuller (1979) and Evans and Savin (1981) have obtained representations for this limiting distribution. For the NNS model of (1.1), Cumberland and Sykes (1982) found that the normalized processes $n^{-1/2}y_n([nt])$ converge weakly to an Ornstein–Uhlenbeck process defined by the Itô stochastic differential equation (SDE)

$$dY(t) = -\beta Y(t)\, dt + \sigma\, dW(t). \tag{1.4}$$

Bobkoski (1983) independently proved the latter result, and based on this convergence obtained

$$n(\hat{\phi}_n - \phi_n) \Rightarrow \frac{\int_0^1 Y(s)\, dW(s)}{\int_0^1 Y^2(s)\, ds}, \tag{1.5}$$

where ϕ_n is given by (1.2). Chan and Wei (1987) obtained similar results for the NNS model and found that when the parameter β goes to ∞ the asymptotic distribution of the "t statistic" $[\sum_{k=1}^{n-1} y_n^2(k)]^{1/2}(\hat{\phi}_n - \phi_n)$ is standard normal, which is in agreement with intuition, since for large β it takes longer for the NNS behavior to manifest itself.

In this work we obtain the weak limit of the M estimator when $\phi_n = 1 - \beta/n$. Martin and Jong (1977) showed that the generalized M estimator is asymptotically normal when $\phi_n \equiv \phi$ with $|\phi| < 1$. Although these authors make certain boundedness assumptions (e.g., on the derivative of ρ), one can adapt their work to show that under standard regularity conditions [e.g., (2.A) and (2.B) below]

$$n^{1/2}(\hat{\phi}_n - \phi_n) \Rightarrow N(0, (1 - \phi^2)V_\rho), \tag{1.6}$$

where

$$V_\rho = \frac{E\psi^2(\varepsilon(1))}{\left[E\dot{\psi}(\varepsilon(1))\right]^2}, \qquad \psi(u) = \frac{d\rho(u)}{du}, \qquad \dot{\psi}(u) = \frac{d\psi(u)}{du}. \tag{1.7}$$

A simple variational argument will show the most "efficient" M estimator (the one minimizing V_ρ) is obtained from $\rho = -\log(f)$, where f is the density of the ε's, that is, when $\hat{\phi}$ is the maximum likelihood estimator, MLE, conditioned on the initial value $y_n(0)$. Other efficiency results for the stationary AR(1) process when the errors are not normal can be found in Johnson and Akritas (1982). For the nearly nonstationary model where ϕ_n is given by (1.2), a similar calculation based on the limit theorems presented here indicates that the MLE will generally *not* be the most "efficient" M estimator. Indeed, the ρ function which works "best" is a linear combination of the LSE and MLE criterion functions.

We comment briefly on various difficulties not considered here. In order that (1.3) produce an estimator which is useful in practice, it is necessary to include scale estimation [for the scale σ^2 of the $\varepsilon(i)$'s]. We do not consider that here, although it follows from Cox (1990) that consistent estimates of scale are readily available, and we will assume henceforth that the scale is known, although further investigation is called for. Although consistent estimation of σ^2 is easy, one cannot in fact consistently estimate β, for reasons discussed in Cox (1990). This fact will have important ramifications later; see Section 3. One would also be interested in cases where the $\varepsilon(i)$'s have infinite variance, but we cannot treat this with the current setup. Finally, it is important to note that the estimator in (1.3) with $\rho = -\log(f)$ is not really the exact MLE if one assumes the process is stationary, since the full MLE would include a term in the log likelihood from the initial distribution of $y_n(0)$. This point is treated in some detail in Cox (1990) for the Gaussian likelihood. In the terminology of that paper, the estimator of (1.3) should be referred to as the "conditional M estimate" [or "conditional MLE" when $\rho = -\log(f)$]. We have tried to include the term from the initial distribution (which is virtually impossible to even compute in the non-Gaussian case) but have been unable to complete the analysis. Essentially, there is a nontrivial loss of information in regarding $y_n(0)$ as a known constant [i.e., conditioning on $y_n(0)$]. Nonetheless, our results on the inefficiency of the MLE are still valid as one may treat any datum as a constant rather than a random variable whose distribution contains information about the parameters of interest, and, of course, in some cases it may be necessary to treat it as constant.

The asymptotic results that we present in this work deal with weak convergence of a sequence of stochastic processes with sample paths in $D_{\mathbb{R}^d}[0, T]$, the space of $\mathbb{R}^d$-valued functions defined on $[0, T]$ which are right continuous and have left limits, to a process with sample paths in $C_{\mathbb{R}^d}[0, T]$, the space of continuous $\mathbb{R}^d$-valued functions on $[0, T]$. The sequence of processes we investigate here are solutions of stochastic difference equations. In a natural way one might expect that if the difference equation "converges" in some sense to a (stochastic) differential equation, then the solutions of these equations would be "near" each other.

We base our proofs in the Stroock and Varadhan characterization of the solution of a SDE as the solution of an associated *martingale problem*. For a detailed account, see, for example, Ethier and Kurtz (1986), Section 5.3, or Stroock and Varadhan (1979), Chapter 6. We obtain the asymptotic results of later sections from the following diffusion approximation theorem due to Ethier and Kurtz. Here, $C_c^\infty(\mathbb{R}^d)$ denotes the space of functions $f\colon \mathbb{R}^d \to \mathbb{R}$ which are infinitely differentiable and have compact support.

THEOREM 1 [Ethier and Kurtz (1986), 7.4.1]. *Let $a = ((a_{ij}))$ be a continuous, symmetric, nonnegative definite $d \times d$ matrix-valued function on $\mathbb{R}^d$ and $b\colon \mathbb{R}^d \to \mathbb{R}^d$ be continuous. Let $\mathbf{A}$ be the second-order differential operator on $C_c^\infty(\mathbb{R}^d)$ given by*

$$\mathbf{A} f = \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij}\, \partial_i \partial_j f + \sum_{i=1}^{d} b_i\, \partial_i f, \qquad f \in C_c^\infty(\mathbb{R}^d)$$

and suppose the $C_{\mathbb{R}^d}[0, \infty)$-martingale problem for $\mathbf{A}$ is well posed.

For $n = 1, 2, \dots$, let $\mathbf{X}_n$ and $\mathbf{B}_n$ be processes with sample paths in $D_{\mathbb{R}^d}[0, \infty)$ and let $\mathbf{A}_n = ((A_n^{ij}))$ be a symmetric $d \times d$ matrix-valued process such that A_n^{ij} has sample paths in $D_{\mathbb{R}}(\mathbb{R})$ and $\mathbf{A}_n(t) - \mathbf{A}_n(s)$ is nonnegative definite for $t > s \ge 0$. Set $F_t^n = \sigma(\mathbf{X}_n(s), \mathbf{B}_n(s), \mathbf{A}_n(s)\colon 0 \le s \le t)$.

Let $\tau_n^r = \inf\{t\colon |\mathbf{X}_n(t)| \ge r \text{ or } |\mathbf{X}_n(t^-)| \ge r\}$ and suppose

$$\mathbf{M}_n \equiv \mathbf{X}_n - \mathbf{B}_n \tag{1.8}$$

and

$$M_n^i M_n^j - A_n^{ij}, \qquad i, j = 1, 2, \dots, d, \tag{1.9}$$

are local $\{F_t^n\}$ martingales, and that for each $r > 0$ and $T > 0$.

$$\lim_{n \to \infty} E\left[\sup_{t \le \min(T, \tau_n^r)} |\mathbf{X}_n(t) - \mathbf{X}_n(t^-)|^2 \right] = 0, \tag{1.10}$$

$$\lim_{n \to \infty} E\left[\sup_{t \le \min(T, \tau_n^r)} |\mathbf{B}_n(t) - \mathbf{B}_n(t^-)|^2 \right] = 0, \tag{1.11}$$

$$\lim_{n \to \infty} E\left[\sup_{t \le \min(T, \tau_n^r)} \left|A_n^{ij}(t) - A_n^{ij}(t^-)\right| \right] = 0, \tag{1.12}$$

$$\sup_{t \le \min(T, \tau_n^r)} \left| \mathbf{B}_n^i(t) - \int_0^t b_i(\mathbf{X}_n(s))\, ds \right| \to_P 0 \tag{1.13}$$

and

$$\sup_{t \le \min(T, \tau_n^r)} \left| \mathbf{A}_n^{ij}(t) - \int_0^t a_{ij}(\mathbf{X}_n(s))\, ds \right| \to_P 0, \tag{1.14}$$

where the latter three relations hold for $1 \le i,\ j \le d$. *Suppose that* $\mathbf{X}_n(0)$ *converges weakly to a random variable with distribution* ν, *then* $\{\mathbf{X}_n\}$ *converges weakly to the solution of the martingale problem for* $(\mathbf{A}, \nu)$.

REMARK. By the representation mentioned before, the limiting process corresponds to the diffusion with infinitesimal generator given by **A**.

The rest of the paper is organized as follows: In Section 2 we formalize our problem and state the asymptotic theorem. In Section 3 we derive an expression for the asymptotic mean squared error, MSE, and derive the form of the optimal M estimator. In Section 4 we give the proof of the main theorem.

2. Statement of the main theorem.

Assume ρ in (1.3) is differentiable and set $\psi = \rho$ as before. Also assume that the following statements for the ψ function hold:

(2.A) ψ is continuously differentiable and satisfies the second-order Lipschitz condition

$$\psi(t) - \psi(t_0) - (t - t_0)\psi(t_0) = C(t - t_0)^2 \alpha(t, t_0), \tag{2.1}$$

where C is a positive constant and $|\alpha(t, t_0)| < 1$.

(2.B) $\varepsilon(1), \varepsilon(2), \dots$ is an iid sequence with $E\varepsilon(i) = 0$, $E\varepsilon(i)^2 = \sigma^2 > 0$ and the $(2 + \delta)$-order moments of $\varepsilon(1)$, $\psi(\varepsilon(1))$ and $\psi(\varepsilon(1))$ are finite for some $\delta > 0$.

(2.C) $E\psi(\varepsilon(1)) = 0$ and $E\dot{\psi}(\varepsilon(1)) = 1$.

The assumption $E\dot{\psi}(\varepsilon(1))$ involves no loss of generality provided $E\dot{\psi}(\varepsilon(1)) \neq 0$. This normalization simplifies some of our formulas below. Note that a version of ψ which includes scale is given by $\psi_\sigma(x) = \sigma\psi_1(x/\sigma)$, where ψ_1 is a fixed score function independent of σ. One would use this score when simultaneously estimating scale, which is generally necessary in practice.

Now, for $\hat{\phi}_n$ to be a solution of (1.3), it is necessary that

$$\Psi_n(\hat{\phi}_n) \equiv \sum_{k=0}^{n-1} y_n(k)\psi\big(y_n(k+1) - \hat{\phi}_n y_n(k)\big) = 0. \tag{2.2}$$

The main result in this paper is the following theorem.

THEOREM 2. *Suppose assumptions* (2.A) *through* (2.C) *hold. Let* $\phi_n = 1 - \beta/n$ *with* β *a positive real constant. Then under the model* (1.1) *with* $y_n(0) \equiv_D \sum_{l=0}^{\infty} \phi_n^l \varepsilon(-l)$:

(a) *There exists a sequence* $\{\hat{\phi}_n\}$ *of solutions of* (2.2) *such that*

$$\big(\hat{\phi}_n - \phi_n\big) = O_p(n^{-1}). \tag{2.3}$$

(b) *For such a sequence*

$$n(\hat{\phi}_n - \phi_n) \Rightarrow \frac{\int_0^1 Y(s)\, dW_2(s)}{\int_0^1 Y^2(s)\, ds}, \tag{2.4}$$

where $Y(t)$ is the Ornstein–Uhlenbeck process defined by the stochastic differential equation

$$\begin{aligned} dY(t) &= -\beta Y(t)\, dt + dW_1(t), \\ Y(0) &\equiv_D N\left(0, \frac{\sigma^2}{2\beta}\right), \end{aligned} \tag{2.5}$$

and $[W_1(t), W_2(t)]'$ is a two-dimensional Brownian motion with

$$E[W_1^2(t)] = tE[\varepsilon^2(1)], \qquad E[W_2^2(t)] = tE[\psi^2(\varepsilon(1))],$$
$$E[W_1(t)W_2(t)] = tE[\varepsilon(1)\psi(\varepsilon(1))].$$

REMARK. Implicitly stated in the assumed initial condition for the sequence of AR(1) processes is that for each n the process is stationary. Thus it is natural that the initial condition for the Ornstein–Uhlenbeck process of (2.5) is the one needed to ensure the stationarity of such a process [Arnold (1974), page 135]. The result of the theorem holds whenever the distribution of $n^{-1/2}y_n(0)$ has a weak limit, then the distribution for $Y(0)$ in (2.5) should be changed to this limiting initial distribution. In particular, if $y_n(0) = 0$ for all n, then, of course, $Y(0) = 0$ should be used in (2.5). See the remark after the proof of Theorem 3 below.

If we let

$$\begin{aligned} t &= y_n(k+1) - \hat{\phi}_n y_n(k), \\ t_0 &= y_n(k+1) - \phi_n y_n(k) = \varepsilon(k+1) \end{aligned} \tag{2.6}$$

in (2.1), then (2.2) becomes, with $\alpha(k) = \alpha(t, t_0)$,

$$\begin{aligned} &\sum_{k=0}^{n-1} [y_n(k)\psi(\varepsilon(k+1))] - (\hat{\phi}_n - \phi_n) \sum_{k=0}^{n-1} y_n^2(k) \\ &\quad - (\hat{\phi}_n - \phi_n) \sum_{k=0}^{n-1} \{y_n^2(k)[\dot{\psi}(\varepsilon(k+1)) - 1]\} \\ &\quad + (\hat{\phi}_n - \phi_n)^2 C \sum_{k=0}^{n-1} [y_n^3(k)\alpha(k)] = 0. \end{aligned} \tag{2.7}$$

The weak limit in (2.4) is suggested by neglecting the last two terms of the right-hand side of (2.7), so that

$$n(\hat{\phi}_n - \phi_n) \approx \frac{\sum_{k=0}^{n-1}[y_n(k)\psi(\varepsilon(k+1))]}{n^{-1}\sum_{k=0}^{n-1} y_n^2(k)}. \tag{2.8}$$

The scaling n^{-1} (rather than $n^{-1/2}$) is nonstandard and results from the near nonstationarity. We will now give a heuristic justification for this scaling and the representation of the limiting distribution given in (2.4). Define

$$\eta(k) = \left[\varepsilon(k), \psi(\varepsilon(k)), \dot{\psi}(\varepsilon(k)) - 1\right]'$$

and let Σ be the covariance matrix of the random vector $\eta(1)$. Now we define the stochastic processes $Y_n(t)$ and $\mathbf{W}_n$ for t in $[0, 1]$ by

$$Y_n(t) = n^{-1/2} y_n([nt]) \tag{2.9}$$

and

$$\begin{aligned} \mathbf{W}_n &= [W_{1,n}(t), W_{2,n}(t), W_{3,n}(t)]' \\ &= n^{-1/2} \sum_{k=1}^{[nt]} \eta(k) \end{aligned} \tag{2.10}$$

(with the usual convention that summation equals 0 when the upper limit is less than the lower). The $W_{3,n}$ component does not appear in the limiting distribution but is used in the proof.

Let Δ be the usual forward difference operator, that is, $\Delta m(k) = m(k+1) - m(k)$ and $\Delta t = n^{-1}$. Then (2.8) can be written as

$$n\left(\hat{\phi}_n - \phi_n\right) \approx \frac{\sum_{k=0}^{n-1}[Y_n(k/n)\, \Delta W_{2,n}(k/n)]}{\sum_{k=0}^{n-1} Y_n^2(k/n)\, \Delta t}. \tag{2.11}$$

Let $\mathbf{W}(t) = [W_1(t), W_2(t), W_3(t)]'$ be a three-dimensional Brownian motion such that covariance matrix of the random vector $\mathbf{W}(t)$ is $t\Sigma$. It can be proven by means of the martingale central limit theorem [see, e.g., Ethier and Kurtz (1986), Section 7.1] that the process $\mathbf{W}_n$ defined in (2.10) converges weakly to $\mathbf{W}$. Since Y_n converges to Y [Cumberland and Sykes (1982)], the summations in (2.11) are approximately Riemann–Stieltjes sums for the integrals in (2.4), and we will show in Theorem 3 that the two summations in (2.11) jointly converge to the corresponding integrals in (2.4).

3. Optimality. We now explore the optimality of M estimators under a natural criterion. Our approach is to minimize an asymptotic mean squared error

$$Q(\psi) = E\left\{\left[\frac{\int_0^1 Y(s)\, dW_2(s)}{\int_0^1 Y^2(s)\, ds}\right]^2\right\}. \tag{3.1}$$

Surprisingly, we have found that this criterion leads to the finding that the optimal ψ function is a linear combination of $\psi_{\mathrm{LS}}(x) = x$ and $\psi_{\mathrm{ML}}(x) = -I_f^{-1}\dot{f}(x)/f(x)$, where f is the probability density function of the innovations and I_f is the Fisher information of the location problem for f. (We assume throughout this section that f exists and satisfies the usual regularity conditions.) Note that ψ_{LS} corresponds to the least squares score function while ψ_{ML} is proportional to the usual score function for the MLE. The optimal ψ function so obtained is not directly useful as an estimator since the coefficients of the linear combination depend on the unknown parameter β. Nonetheless, it does suggest a two-stage procedure that may work well. The first stage is to estimate ϕ_n by say the MLE, $\hat{\phi}_{n,\mathrm{MLE}}$, and hence β by $\hat{\beta}_{n,\mathrm{MLE}} = n(1 - \hat{\phi}_{n,\mathrm{MLE}})$. One can then find the optimal ψ function for the estimate $\hat{\beta}_{n,\mathrm{MLE}}$ and the second stage consists of finding the solution of the M-estimation equation for this ψ. However, since β is not consistently estimable, it is not clear that this will lead to an asymptotically efficient estimator, or even an estimator that will improve over the MLE.

To prove the claim, we can think of Q as a functional on $\mathbf{L}^2(f) = \{\xi\colon \int \xi^2(x) f(x)\,dx < \infty\}$. We would like to find the minimizer of Q on $\mathbf{L}^2(f)$ subject to the constraints in (2.C), that is, $\int \xi(x) f(x)\,dx = 0$ and $\int \dot{\xi}(x) f(x)\,dx = 1$. We have shown in the appendix that Q can be written as

$$Q(\psi) = (L_1 - L_2)C^2(\psi) + L_2 V(\psi), \tag{3.2}$$

where

$$L_1 = E\left\{\left[\frac{\int_0^1 Y\,dW_1}{\int_0^1 Y^2\,ds}\right]^2\right\}, \qquad L_2 = \sigma^2 E\left\{\left[\int_0^1 Y^2\,ds\right]^{-1}\right\}, \tag{3.3}$$

$$C(\psi) = \mathrm{Cov}[\varepsilon, \psi]/\sigma^2, \quad \text{and} \quad V(\psi) = \mathrm{Var}[\psi]/\sigma^2$$

Here ψ and ε are shorthand for $\psi(\varepsilon(1))$ and $\varepsilon(1)$, respectively. Note that all quantities above are scale invariant. Thus Q is a positive-definite quadratic functional and since the constraints are linear, the solution to the minimization problem is obtained by setting the first variation (with respect to ψ) of the Lagrangian $Q(\psi) + \lambda_1 E[\psi] + \lambda_2\{E[\dot{\psi}] - 1\}$ equal to 0 and choosing the multipliers λ_1 and λ_2 so that the constraints hold. This operation followed by an integration by parts leads to the equation

$$2\sigma^{-2}(L_1 - L_2)C(\psi) + 2\sigma^{-2}L_2\psi(x) f(x) + \lambda_1 f(x) - \lambda_2 \dot{f}(x) = 0,$$

whence

$$\psi(x) = \kappa x + \frac{\lambda_2\sigma^2}{2L_2}\frac{\dot{f}(x)}{f(x)} - \frac{\lambda_1\sigma^2}{2L_2}, \tag{3.4}$$

where

$$\kappa = (1 - L_1/L_2)C(\psi).$$

It is easy to see that both $E(\varepsilon) = 0$ and the constraint $E(\psi) = 0$ imply $\lambda_1 = 0$. Thus the optimal ψ is a linear combination of the least squares and maximum likelihood score functions. Also the constraint $E(\dot{\psi}) = 1$ implies

$$\frac{\lambda_2\sigma^2}{2L_2} = I_f^{-1}(\kappa - 1).$$

Substitution of the values of the multipliers into (3.4) gives

$$\psi(x) = \kappa x + (\kappa - 1) I_f^{-1} \frac{\dot{f}(x)}{f(x)}. \tag{3.5}$$

Calculating $\mathrm{Cov}(\psi, \varepsilon)$ for ψ in (3.5) gives κ and plugging this into (3.5) yields the optimal score function

$$\psi_*(x) = \frac{(L_2 - L_1)x - \sigma^2 L_1[\dot{f}(x)/f(x)]}{L_2 - L_1(1 - \sigma^2 I_f)}. \tag{3.6}$$

One should note that ψ_* depends on β through L_1 and L_2. Furthermore, evaluation of L_1 and L_2 is nontrivial since they are expectations of rational functions of random integrals whose distribution is nontrivial to describe. Following Williams (1941), one can obtain the moments of the ratio of powers of the numerator random variable within the brackets in the definition of L_1 in (3.3), to be denoted by N, and the denominator random variable, to be denoted by D, from the joint moment generating function of N and D. Thus, for example, if $\Lambda(s_0, s) = E[\exp\{-s_0 D - sN\}]$ then

$$\int_0^\infty \Lambda(s_0, 0)\, ds_0 = E\left[\int_0^\infty e^{-s_0 D}\, ds_0\right] = E\left[\frac{1}{D}\right] \tag{3.7}$$

and

$$\int_0^\infty \int_t^\infty \frac{\partial^2}{\partial s^2} \Lambda(s_0, s)\bigg|_{s=0} ds_0\, dt = \int_0^\infty \int_t^\infty E[N^2 e^{-s_0 D}]\, ds_0\, dt = E\left[\frac{N}{D}\right]^2. \tag{3.8}$$

The formal manipulations of interchanging differentiation and integration will be justified shortly. From (4.20) of Bobkoski (1983), we have that the joint MGF of N and D, when $Y(0) = 0$, is given by

$$\begin{aligned} \Lambda(s_0, s) &= E[\exp(-s_0 D - sN)] \\ &= \exp\left\{\frac{\beta + s}{2}\right\}[\cosh(z) + (\beta + s)\mathrm{shnc}(z)]^{-1/2}, \end{aligned} \tag{3.9}$$

where

$$z = (\beta^2 + 2\beta s + 2s_0)^{1/2} \quad \text{and} \quad \mathrm{shnc}(z) = \frac{\sinh(z)}{z}.$$

Expressions for the MGF for other initial conditions are available [Llatas (1987)]. The choice of $Y(0) = 0$ is motivated by convenience for checking the results; see the discussion after Table 1. By the remarks after the statement of Theorem 2, $Y(0) = 0$ is a permissible choice if $y_n(0) = 0$ for all n. The fact that Λ in (3.9) is differentiable and that the terms in these derivatives will be eventually dominated by e^{-Ks_0}, where K is a positive constant, as $s_0 \to \infty$, allow us to interchange the order of the integration and differentiation in both (3.7) and (3.8) by application of the dominated convergence and Fubini theorems. In Table 1 we exhibit some of the values of L_1 and L_2 calculated using the integration subroutine DQAGI in QUADPACK.

TABLE 1
Values of L_1 and L_2 obtained by numerical integration

β	L_1	L_2
0.200	13.698232	5.921848
0.400	14.104907	6.285748
0.600	14.507015	6.653889
0.800	14.905686	7.025686
1.000	15.301856	7.400631
2.000	17.266291	9.309338
3.000	19.228876	11.252599
4.000	21.198798	13.214063
5.000	23.175399	15.186088
6.000	25.156913	17.164780
7.000	27.141975	19.147965
8.000	29.129653	21.134334
9.000	31.119311	23.123046
10.000	33.110506	25.113539
11.000	35.102916	27.105415
12.000	37.096305	29.098390
13.000	39.090494	31.092254
14.000	41.085346	33.086846
15.000	43.080753	35.082042
16.000	45.076630	37.077746
17.000	47.072908	39.073881
18.000	49.069531	41.070385
19.000	51.066453	43.067207
20.000	53.063637	45.064306

A plot of the values obtained presents a very curious feature: They fall in what seem to be two parallel straight lines. Regression lines were fit to the values in Table 1 assuming the two lines are parallel, and the fitted equations are given by

$$L_1 \approx 13.33 + 1.98\beta, \qquad L_2 \approx 5.37 + 1.98\beta.$$

These are very accurate in the range $0.2 < \beta \leq 20.0$. The relative error in L_1 is less than 0.2% in this range, and the relative error in L_2 is less than 3.05%. Over the range $2.0 < \beta \leq 20.0$, the relative error in L_2 is less than 0.51%. These approximations could be useful in the two-stage procedure proposed above.

We checked the numerical integrations used to produce Table 1 in two ways. When $Y(0) = 0$, L_1 can also be computed from the asymptotic density of the estimation error for the LSE using the density given in Bobkoski (1983), again by numerical integration. The results so obtained for $\beta = 2$, 10 and 20 agree with those of Table 1 to two decimal places. We also used simulation of the NNS AR(1) process for sample sizes $n = 100$, 500 and 1000, for the same values of β. We generated 10,000 realizations of each such series. The Monte Carlo results agreed with those of Table 1 with 95% confidence, except for $\beta = 20$, where we believe the bias of the finite sample estimators has not been overcome. The existence of such simple and accurate approximations to L_1 and L_2 suggests conjectures that may lead to fruitful research. Further details may be found in Llatas (1987).

Now we are in the position to calculate values of Q for the score functions $\psi_{\mathrm{LS}}(x) = x$, $\psi_{\mathrm{ML}}(x) = -I_f \dot{f}(x)/f(x)$ and ψ_*. By (3.2) we have

$$Q(\psi_{\mathrm{LS}}) = L_1, \qquad Q(\psi_{\mathrm{ML}}) = L_1/I'_f,$$

where $I'_f = \sigma^2 I_f$ is the information when $\sigma^2 = 1$. An interesting feature of the NNS asymptotic relative efficiency of the MLE and LSE is that it is the same as in the classical setting. See (1.6) and (1.7). Also,

$$Q(\psi_*) = \frac{L_1 L_2}{L_2 + L_1(I'_f - 1)}.$$

Note that $I'_f \geq 1$ with equality only for the normal density [Rustagi (1976)]. The inefficiency of the LSE w.r.t. the optimal score function is

$$Q(\psi_{\mathrm{LS}})/Q(\psi_*) = 1 + (I'_f - 1)L_1/L_2, \tag{3.10}$$

and a minimum is obtained when $I'_f = 1$, that is, the Gaussian case. The maximum is obtained when $I'_f = \infty$, where the inefficiency is ∞. The inefficiency of the MLE w.r.t. the optimal score is $1/I'_f$ times the inefficiency of the LSE. Again, the minimum is obtained when $I'_f = 1$ and the maximum when $I'_f = \infty$, but for the MLE, the maximum inefficiency is bounded by L_1/L_2. Recall that L_1 and L_2 depend on β. Assume again that $Y(0) = 0$, and then the largest value of L_1/L_2 occurs when $\beta = 0$ when it is approximately $13.33/5.37 \approx 1/0.40$, that is, the efficiency of the MLE is bounded below by about 40%. For fixed I'_f, the maximum inefficiency of both the LSE and MLE occurs at $\beta = 0$.

4. The large sample behavior of $\hat{\phi}_n$. In this section we will prove Theorem 2. First, we establish the joint limiting distribution of the sums in (2.7) as an application of Theorem 1.

THEOREM 3. *Consider the model* (1.1) *with initial value* $y_n(0)$ *as in the statement of Theorem* 2. *Suppose that assumptions* (2.A) *to* (2.C) *hold. Consider the sequence of processes on* $D_{\mathbb{R}^3}[0, 1)$ *defined by*

$$\mathbf{X}_n(t) = \begin{bmatrix} n^{-1/2} y_n([nt]) \\ n^{-1} \sum_{k=1}^{[nt]} [y_n(k-1)\psi(\varepsilon(k))] \\ n^{-3/2} \sum_{k=1}^{[nt]} y_n^2(k-1)\left[\dot{\psi}(\varepsilon(k)) - 1\right] \end{bmatrix}. \tag{4.1}$$

Then $\mathbf{X}_n \Rightarrow \mathbf{X}$ *as* $n \to \infty$, *where* $\mathbf{X}$ *is the continuous process on* $[0,1]$ *given by*

$$\mathbf{X}(t) = \left[Y(t), \int_0^t Y(s)\,dW_2(s), \int_0^t Y^2(s)\,dW_3(s)\right]', \tag{4.2}$$

where $\mathbf{W}$ *is the three-dimensional Brownian motion defined below* (2.11) *and* Y *is the Ornstein–Uhlenbeck process defined by* (2.5) *with initial condition having the stationary distribution.*

PROOF. First, note that we represent $\mathbf{W}$ by

$$\mathbf{W}(t) = \Gamma \mathbf{b}(t), \tag{4.3}$$

where $\mathbf{b}(t)$ is a three-dimensional standard Brownian motion with covariance (tI) and $\Gamma = (\gamma_{ij})$ is the Cholesky factor for Σ, that is, Γ is a 3×3 lower triangular matrix such that $\Gamma\Gamma' = \Sigma$. Now the process $\mathbf{X}(t)$ satisfies the SDE:

$$\begin{aligned} d\mathbf{X}(t) &= \begin{bmatrix} -\beta X_1(t) \\ 0 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 1 & 0 & 0 \\ 0 & X_1(t) & 0 \\ 0 & 0 & X_1^2(t) \end{bmatrix} d\mathbf{W}(t) \\ &= b(\mathbf{X}(t))\,dt + G(\mathbf{X}(t))\,d\mathbf{W}(t) \\ &= b(\mathbf{X}(t))\,dt + G(\mathbf{X}(t))\Gamma\,d\mathbf{b}(t) \end{aligned} \tag{4.4}$$

with initial condition $\mathbf{X}(0) = (Y(0), 0, 0)'$. The last equality in (4.4) follows by (4.3) and Itô's formula [Arnold (1974), page 90].

The functions b and G do not depend directly on time and they have continuous partial derivatives of first order that are bounded on $\{|\mathbf{x}| \le M\}$ for all $M > 0$. Consequently, by Corollary 6.3.3 of Arnold (1974), (4.4) has exactly one continuous solution. Moreover, the process $\mathbf{X}(t)$ is a three-dimensional diffusion process on $[0,1]$ with drift vector $b(\mathbf{x})$ and diffusion matrix $a(\mathbf{x}) = G(\mathbf{x})\Gamma\Gamma'(\mathbf{x})G'(\mathbf{x}) = G(\mathbf{x})\Sigma G'(\mathbf{x})$ [Arnold (1974), Theorem 9.3.1., page 152]. In this case $a(\mathbf{x})$ equals

$$a(\mathbf{x}) = \begin{bmatrix} \sigma_{11} & \sigma_{12}x_1 & \sigma_{13}x_1^2 \\ \sigma_{12}x_1 & \sigma_{22}x_1^2 & \sigma_{23}x_1^3 \\ \sigma_{13}x_1^2 & \sigma_{23}x_1^3 & \sigma_{33}x_1^4 \end{bmatrix}. \tag{4.5}$$

Thus $\mathbf{X}(t)$ is a solution of the associated martingale problem for the infinitesimal operator of the diffusion, that is,

$$D = \sum_{i=1}^3 b_i(\mathbf{x}) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 a_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i\,\partial x_j} \tag{4.6}$$

with initial measure equal to Law($\mathbf{X}(0)$), which should be equal to the weak limit of Law($\mathbf{X}_n(0)$) to have the appropriate limiting distribution. We claim that Law($\mathbf{X}(0)$) is the three-dimensional degenerate normal $N(0, \Theta\sigma^2/2\beta)$, where Θ_{ij} equals 0 unless $i = j = 1$. Our claim follows from the definition of $\mathbf{X}_n(0)$ and the fact that

$$Y_n(0) = \sum_{k=0}^{\infty} \phi_n^k (n^{-1/2}\varepsilon(-k))$$

converges weakly to a random variable distributed as an $N(0, \sigma^2/2\beta)$ by an easy application of the Linderberg–Feller central limit theorem to the triangular array defined by

$$T_{n,k} = n^{-1/2}\phi_n^k \varepsilon(-k), \qquad 0 \le k \le n^2.$$

Now, $\mathbf{X}_n$ is a solution of the following stochastic difference equation

$$\Delta \mathbf{X}_n\left(\frac{k}{n}\right) = \begin{bmatrix} -\beta X_{1,n}(k/n) \\ 0 \\ 0 \end{bmatrix} \Delta t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & X_{1,n}(k/n) & 0 \\ 0 & 0 & X_{1,n}^2(k/n) \end{bmatrix} \Delta W_n\left(\frac{k}{n}\right) \tag{4.7}$$

with $\mathbf{W}_n$ defined in (2.10) so it is natural to think that $\mathbf{X}_n$ will approximate the continuous process $\mathbf{X}$. We proceed to prove this by finding three-dimensional processes $\mathbf{B}_n(t)$ and 3×3 matrix-valued processes $\mathbf{A}_n(t)$ such that the conditions of Theorem 1 are satisfied. From (4.7) it follows that

$$\Delta \mathbf{X}_n\left(\frac{k}{n}\right) = \begin{bmatrix} -\beta Y_n(k/n) \\ 0 \\ 0 \end{bmatrix} \Delta t + n^{-1/2}\xi_n(k+1),$$

where

$$\xi_n(k) = \left[\varepsilon(k), n^{-1/2} y_n(k-1)\psi(\varepsilon(k)), n^{-1} y_n^2(k-1)\left[\dot{\psi}(\varepsilon(k)) - 1\right]\right]'.$$

Let $G_k = \sigma(\mathbf{X}_n(j/n)\colon 0 \le j \le k)$. Since $E[\xi_n(k) | G_{k-1}] = 0$ the predictable compensator of $\mathbf{X}_n$ is given by

$$\begin{aligned} \mathbf{B}_n(t) &= \sum_{k=0}^{[nt]-1} \{E[\Delta \mathbf{X}_n(k/n) | G_k]\} \\ &= \left[-\beta \sum_{k=0}^{[nt]-1} Y_n(k/n)\,\Delta t, 0, 0\right]' \end{aligned} \tag{4.8}$$

and writing $\mathbf{X}_n(k/n) = \Delta \mathbf{X}_n((k-1)/n) + \mathbf{X}_n((k-1)/n)$, one can see that

$$\mathbf{M}_n(k/n) = \mathbf{X}_n(k/n) - \mathbf{B}_n(k/n) = n^{-1/2}\xi_n(k) + \delta_n(k), \tag{4.9}$$

where $\delta_n(k)$ is G_{k-1} measurable. Thus one can show $\mathbf{A}_n$, the compensator of $\mathbf{M}_n(k/n)\mathbf{M}_n'(k/n)$, is

$$\begin{aligned} \mathbf{A}_n(t) &= \sum_{k=0}^{[nt]-1} \{E[\mathbf{M}_n((k+1)/n)\mathbf{M}_n'((k+1)/n) \\ &\qquad\qquad - \mathbf{M}_n(k/n)\mathbf{M}_n'(k/n) | G_k]\} \\ &= n^{-1} \sum_{k=0}^{[nt]-1} E[gx_n(k+1)\xi_n'(k+1) | G_k]. \end{aligned} \tag{4.10}$$

It follows from the last equality of (4.10) that the increments $\mathbf{A}_n(t) - \mathbf{A}_n(s)$, $t > s$, of the process so defined are nonnegative definite.

What is left now is to verify the "continuity" conditions (1.10) to (1.12) and the "approximation" conditions (1.13) and (1.14) of Theorem 1. We start with the approximation conditions. For condition (1.13) we must show

$$\sup_{0\le t\le 1}\left|B_{1,n}(t) - \int_0^t b_1(X_n(s))\,ds\right| \to_P 0.$$

But the absolute value equals

$$\begin{aligned}\beta&\left|\int_0^t n^{-1/2}y_n([ns])\,ds - \sum_{k=0}^{[nt]-1} n^{-1/2}y_n(k)\,\Delta t\right|\\ &= \beta(t-[nt]/n)|Y_n(t)| \le \frac{\beta}{n}|Y_n(t)| \le \frac{\beta}{n}\|Y_n\|_\infty.\end{aligned} \tag{4.11}$$

Since $\|Y_n\|_\infty$ is bounded in probability [Bobkoski (1983), page 25], the last quantity goes to 0 as $n \to \infty$. Condition (1.14) will also follow by the same type of argument and the boundedness of $\|Y_n^q\|_\infty$ for $q = 0, 1, 2, 3, 4$. To prove the continuity conditions, let τ_n^r be the stopping time defined in Theorem 1. Thus for $t < \tau_n^r$ we have $|\mathbf{X}_n(t)| < r$ and, in particular,

$$|Y_n(t)| < r \quad \text{for } t < \tau_n^r. \tag{4.12}$$

Hence the continuity condition (1.12) for $\mathbf{A}_n$ is easily verified when we note that it reduces to proving that

$$\lim_{n\to\infty} n^{-1}E\left[\sup_{t\le\tau_n^r}|Y_n(([nt]-1)/n)|^j\right] = 0 \quad \text{for } j = 1,2,3,4, \tag{4.13}$$

which is obvious by (4.12) since we are evaluating the process at a time point strictly smaller than τ_n^r. In the same way, the condition for $\mathbf{B}_n$ reduces to

$$\lim_{n\to\infty}(\beta/n)^2E\left[\sup_{t\le\tau_n^r}Y_n^2(([nt]-1)/n)\right] = 0, \tag{4.14}$$

which follows again by (4.12).

Finally, for the condition on the $\mathbf{X}_n$ process, it is sufficient to verify

$$\lim_{n\to\infty}E\left[n^{-1}\sup_{k\le n\tau_n^r}\left[\varepsilon^2(k) - (2\beta/n)\varepsilon(k)y_n(k-1) + (\beta/n)^2y_n^2(k-1)\right]\right] = 0,$$

$$\lim_{n\to\infty}E\left[n^{-2}\sup_{k\le n\tau_n^r}\left[y_n(k-1)\left[\dot\psi(\varepsilon(k))-1\right]\right]^2\right] = 0,$$

$$\lim_{n\to\infty}E\left[n^{-3}\sup_{k\le n\tau_n^r}\left[y_n^2(k-1)\left[\dot\psi(\varepsilon(k))-1\right]\right]^2\right] = 0.$$

But each one of those conditions holds by (4.12), Lemma 1 and the assumption on the moments of ε, $\psi(\varepsilon)$ and $\dot\psi(\varepsilon)$. Hence Theorem 1 guarantees the weak convergence of $\mathbf{X}_n$ to $\mathbf{X}$. □

REMARK. In the proof of Theorem 3 it is not necessary to make the assumption that $y_n(0)$ has the stationary distribution. The result will follow as soon as $Y_n(0)$ has a weak limit. In particular, the result is true when one assumes $Y_n(0)$ to be constant.

LEMMA 1. *Let* $\{\eta(k)\}_{k=1}^{\infty}$ *be a sequence of iid random variables with finite* $1+\delta$*-moment. Then*

$$n^{-1}E\Big[\max_{0\le k\le n}\eta(k)\Big]\to 0\quad \text{as } n\to\infty. \tag{4.15}$$

PROOF. We have the left-hand side of (4.15) is bounded above by

$$\begin{aligned}E\Big[n^{-1-\delta}\max_{0\le k\le n}|\eta(k)|^{1+\delta}\Big]&\le n^{-\delta}n^{-1}\sum_{k=1}^{n}E|\eta(k)|^{1+\delta}\\&=O(n^{-\delta}).\end{aligned}$$

□

We next prove the weak convergence of the terms in the Taylor expansion in (2.7) and, in particular, the joint convergence of $[\sum_{k=1}^{n-1}Y_n^2(k/n)\,\Delta t,\ \sum_{k=1}^{n-1}Y_n(k/n)\,\Delta W_{2,n}(k/n)]'$ to the random vector $[\int_0^1 Y^2(s)\,ds,\ \int_0^1 Y(s)\,dW_2(s)]'$.

LEMMA 2. *Under model* (1.1) *and assumptions* (2.A) *to* (2.C), *the sequence of four-dimensional random vectors*

$$\mathbf{Z}_n=\begin{bmatrix}\sum_{k=1}^{n-1}Y_n^2(k/n)\,\Delta t\\ \sum_{k=1}^{n-1}|Y_n^3(k/n)|\,\Delta t\\ \sum_{k=1}^{n-1}Y_n(k/n)\,\Delta W_{2,n}(k/n)\\ \sum_{k=1}^{n-1}Y_n^2(k/n)\,\Delta W_{3,n}(k/n)\end{bmatrix}$$

converges weakly to

$$\mathbf{Z}=\Big[\int_0^1 Y^2(s)\,ds,\int_0^1|Y^3(s)|\,ds,\int_0^1 Y(s)\,dW_2(s),\int_0^1 Y^2(s)\,dW_3(s)\Big]'.$$

PROOF. Consider the transformation $g\colon C_{\mathbb{R}^3}[0,1]\to\mathbb{R}^4$ such that

$$\begin{aligned}g(x)&=g([x_1(t),x_2(t),x_3(t)]')\\&=\Big[\int_0^1 x_1^2(s)\,ds,\int_0^1|x_1^3(s)|\,ds,x_2(1),x_3(1)\Big]'.\end{aligned}$$

It is easy to see that this is a continuous transformation. Now let $\mathbf{Z}_n=g(\mathbf{X}_n)$

and $\mathbf{Z} = g(\mathbf{X})$, where $\mathbf{X}_n$ and $\mathbf{X}$ are the processes defined in Theorem 3. Then $\mathbf{Z}_n$ converges weakly to $\mathbf{Z}$ by the continuity principle [Theorem 5.1 of Billingsley (1968), page 30]. □

We now prove our main theorem in the same fashion that Cramér showed the asymptotic properties of the maximum likelihood estimator [Cramér (1946), Chapter 33].

PROOF OF THEOREM 2. By means of (2.7) we can write $\Psi_n(\zeta) = 0$, after multiplication by n^{-2}, in the form

$$n^{-2}\Psi_n(\zeta) = T_{0,n} - (\zeta - \phi_n)T_{1,n} - (\zeta - \phi_n)T_{2,n} + (\zeta - \phi_n)^2 T_{3,n} = 0, \tag{4.16}$$

where

$$\begin{aligned} T_{0,n} &= n^{-1} \sum_{k=1}^{n-1} Y_n(k/n)\,\Delta W_{2,n}(k/n), \\ T_{1,n} &= \sum_{k=1}^{n-1} Y_n^2(k/n)\,\Delta t, \\ T_{2,n} &= n^{-1/2} \sum_{k=1}^{n-1} Y_n^2(k/n)\,\Delta W_{3,n}(k/n), \\ T_{3,n} &= n^{1/2} C\theta_n \sum_{k=1}^{n-1} |Y_n^3(k/n)|\,\Delta t \end{aligned} \tag{4.17}$$

and

$$\theta_n = \left[\sum_{k=1}^{n-1} |Y_n^3(k/n)|\right]^{-1} \sum_{k=1}^{n-1} Y_n^3(k/n)\alpha(k),$$

which is bounded by one. Theorem 3 implies that $T_{0,n}$ is $O_p(n^{-1})$ and $T_{2,n}$ is $O_p(n^{-1/2})$, while Lemma 2 implies that $T_{3,n}$ is $O_p(n^{1/2})$ and $T_{1,n}$ converges weakly to a random variable, which is positive with probability one. (This last claim follows from the fact that if $Y = 0$ a.e. then necessarily $W = 0$ a.e. which is a contradiction.) Hence if γ is an arbitrarily small positive number there exists an N such that for all $n > N$ there exist finite positive constants M_0, M_1, M_2, M_3 such that

$$\begin{aligned} P\big[|T_{0,n}| < n^{-1}M_0\big] &> 1 - \frac{\gamma}{4}, \\ P\big[|T_{1,n}| > M_1\big] &> 1 - \frac{\gamma}{4}, \\ P\big[|T_{2,n}| < n^{-1/2}M_2\big] &> 1 - \frac{\gamma}{4}, \\ P\big[|T_{3,n}| < n^{1/2}M_3\big] &> 1 - \frac{\gamma}{4}. \end{aligned} \tag{4.18}$$

Thus with at least probability $1 - \gamma$:

$$(4.19)\qquad \begin{aligned} n^{-2}\Psi_n(\zeta) > -M_0 n^{-1} - M_1(\zeta - \phi_n) - M_2 n^{-1/2}|\zeta - \phi_n| \\ - M_3 n^{1/2}(\zeta - \phi_n)^2 \quad \text{for } \zeta < \phi_n \end{aligned}$$

and

$$(4.20)\qquad \begin{aligned} n^{-2}\Psi_n(\zeta) < M_0 n^{-1} - M_1(\zeta - \phi_n) + M_2 n^{-1/2}|\zeta - \phi_n| \\ + M_3 n^{1/2}(\zeta - \phi_n)^2 \quad \text{for } \zeta \geq \phi_n. \end{aligned}$$

Now, choose n large enough so that

$$n^{-1/2}\left[\frac{2M_0}{M_1}M_2 + \left(\frac{2M_0}{M_1}\right)^2 M_3\right] < \frac{M_0}{2}$$

and for such n, let

$$\zeta_1 = \phi_n - n^{-1}\frac{2M_0}{M_1}, \qquad \zeta_2 = \phi_n + n^{-1}\frac{2M_0}{M_1}.$$

Equation (4.19) gives

$$\begin{aligned} n^{-2}\Psi_n(\zeta_1) &> -M_0 n^{-1} + 2M_0 n^{-1} - n^{-3/2}\left[\frac{2M_0}{M_1}M_2 + \left(\frac{2M_0}{M_1}\right)^2 M_3\right] \\ &> \left(M_0 - \frac{M_0}{2}\right)n^{-1} > 0, \end{aligned}$$

while (4.20) gives

$$\begin{aligned} n^{-2}\Psi_n(\zeta_2) &< M_0 n^{-1} - 2M_0 n^{-1} + n^{-3/2}\left[\frac{2M_0}{M_1}M_2 + \left(\frac{2M_0}{M_1}\right)^2 M_3\right] \\ &< \left(-M_0 + \frac{M_0}{2}\right)n^{-1} < 0. \end{aligned}$$

Thus, since $\Psi_n(\zeta)$ is continuous, the equation $\Psi_n(\zeta) = 0$ will, with probability exceeding $1 - \gamma$, have a root, $\hat{\phi}_n$, between ζ_1 and ζ_2 as we wished. Moreover,

$$|\hat{\phi}_n - \phi_n| < \frac{4M_0}{M_1}n^{-1} \quad \text{with probability } 1 - \gamma$$

and consequently the proof of part (a) is complete.

For part (b) we note

$$(4.21)\qquad n(\hat{\phi}_n - \phi_n) = \frac{nT_{0,n}}{T_{1,n} + T_{2,n} - (\hat{\phi}_n - \phi_n)T_{3,n}}.$$

It follows from the preceding discussion that $T_{2,n} - (\hat{\phi}_n - \phi_n)T_{3,n}$ converges in probability to 0 while, by Lemma 2, $(nT_{0,n}, T_{1,n})'$ jointly converges to $(\int_0^1 Y(s)\,dW_2(s), \int_0^1 Y^2(s)\,ds)'$. Thus the weak convergence of the right-hand side of (4.21) to the random variable in (2.4) is guaranteed by a straightforward application of Slutsky's theorem and Theorem 5.1 in Billingsley (1968). □

APPENDIX

Let $\mathbf{W}(t)$ be the three-dimensional Brownian motion defined in Section 2. As noted in the proof of Theorem 3, we can represent this process by

$$\mathbf{W}(t) = \Gamma \mathbf{b}(t),$$

where $\mathbf{b}(t)$ is a three-dimensional standard Brownian motion with covariance (tI) and $\Gamma = (\gamma_{ij})$ is the Cholesky factor for Σ, that is, Γ is a 3×3 lower triangular matrix such that $\Gamma\Gamma' = \Sigma$. Using this representation, we can prove that $Q(\psi)$ can be expressed as in (3.2). By Itô's theorem [Arnold (1974), page 90], we can write

$$\text{(A.1)} \qquad \int_0^1 Y(s)\, dW_2(s) = \gamma_{21}\int_0^1 Y(s)\, db_1(s) + \gamma_{22}\int_0^1 Y(s)\, db_2(s).$$

Note that $W_1 = \gamma_{11} b_1$ and, consequently, the process Y defined by the SDE (2.5) is independent of b_2 and b_3.

From (A.1) we have

$$\text{(A.2)} \qquad \begin{aligned} Q(\psi) &= (\gamma_{21})^2 E\left[\frac{\int_0^1 Y(s)\, db_1(s)}{\int_0^1 Y^2(s)\, ds}\right]^2 + (\gamma_{22})^2 E\left[\frac{\int_0^1 Y(s)\, db_2(s)}{\int_0^1 Y^2(s)\, ds}\right]^2 \\ &\quad + 2\gamma_{21}\gamma_{22} E\left[\frac{\left(\int_0^1 Y(s)\, db_1(s)\right)\left(\int_0^1 Y(s)\, db_2(s)\right)}{\left(\int_0^1 Y^2(s)\, ds\right)^2}\right]. \end{aligned}$$

Define $F_t = \sigma(\mathbf{b}(s), 0 \le s \le t)$ and $F_t^{(1)} = \sigma(b_1(s), 0 \le s \le t)$. We claim that for any $F_t^{(1)}$-adapted random function $h(t)$ we have

$$E\left[\int_0^1 h(s)\, db_2(s) \middle| F_1^{(1)}\right] = 0$$

and

$$E\left[\left(\int_0^1 h(s)\, db_2(s)\right)^2 \middle| F_1^{(1)}\right] = \int_0^1 h^2(s)\, ds.$$

This can be proven by first looking at $F_t^{(1)}$-adapted step functions and making use of the fact that b_1 and b_2 are independent. Then the usual limiting argument gives the result. Consequently, since $\{Y(t): 0 \le t \le 1\}$ is $F_t^{(1)}$ adapted, one obtains that $E(\int_0^1 Y(s)\, db_2(s) | F_1^{(1)}) = 0$. Thus the expectation of the cross product in (A.2) vanishes since $\int_0^1 Y(s)\, db_1(s)$ and $\int_0^1 Y^2(s)\, ds$ are $F_t^{(1)}$ adapted.

Also

$$E\left[\frac{\int_0^1 Y(s)\,db_2(s)}{\int_0^1 Y^2(s)\,ds}\right]^2 = E\left\{\left[\int_0^1 Y^2(s)\,ds\right]^{-2} E\left[\left(\int_0^1 Y(s)\,db_2(s)\right)^2 \middle| F_1^{(1)}\right]\right\}$$
$$= E\left[\int_0^1 Y^2(s)\,ds\right]^{-1}.$$

From all this discussion Q reduces to

$$\text{(A.3)} \qquad \begin{aligned} Q(\psi) &= \gamma_{21}^2 E\left[\frac{\int_0^1 Y(s)\,db_1(s)}{\int_0^1 Y^2(s)\,ds}\right]^2 + \gamma_{22}^2 E\left(\int_0^1 Y^2(s)\,ds\right)^{-1} \\ &\equiv \frac{\gamma_{21}^2 L_1}{\sigma^2} + \frac{\gamma_{22}^2 L_2}{\sigma^2}. \end{aligned}$$

Plugging the values of γ_{21} and γ_{22} into (A.3) gives expression (3.2). □

Acknowledgment. We would like to thank a helpful referee who pointed out an error in the original manuscript and suggested a shorter proof for Lemma 1.

REFERENCES

ANDERSON, T. W. (1959). On asymptotic distribution of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

ARNOLD, L. (1974). *Stochastic Differential Equations: Theory and Applications*. Wiley, New York.

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

BOBKOSKI, M. J. (1983). Hypothesis testing in non-stationary time series. Ph.D. dissertation, Univ. Wisconsin, Madison.

CHAN, N. H. and WEI, C. Z. (1987). Asymptotic inference of nearly non-stationary AR(1) processes. *Ann. Statist.* **15** 1050–1063.

COX, D. (1990). Gaussian likelihood estimation for nearly nonstationary AR(1) processes. Technical Report 98, Dept. Statistics, Univ. Washington.

CRAMÉR, H. (1946). *Mathematical Methods in Statistics*. Princeton Univ. Press.

CUMBERLAND, W. G. and SYKES, Z. M. (1982). Weak convergence of an autoregressive process used in modeling population growth. *J. Appl. Probab.* **19** 450–455.

DICKEY, D. A. and FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427–431.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

EVANS, G. B. and SAVIN, N. E. (1981). The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model. *Ann. Statist.* **9** 1114–1118.

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

JEGANATHAN, P. (1987). On the asymptotic behavior of least squares estimators in AR time series with roots near the unit circle. Technical Report 150, Dept. Statistics, Univ. Michigan, Ann Arbor.

JOHNSON, R. A. and AKRITAS, M. G. (1982). Efficiencies of tests and estimators for p-order autoregressive processes when the error distribution is nonnormal. *Ann. Inst. Statist. Math.* **34** 579–589.

LLATAS, I. (1987). Asymptotic inference for nearly non-stationary time series. Ph.D. dissertation, Univ. Wisconsin, Madison.

MARTIN, D. and JONG, J. (1977). Asymptotic properties of robust generalized M-estimates for the first order autoregressive parameter. Bell Lab. Memorandum, Murray Hill, N.J.

PHILLIPS, P. C. B. (1987). Toward a unified asymptotic theory for autoregression. *Biometrika* **74** 535–547.

RAO, M. M. (1978). Asymptotic distribution of the boundary parameter of an unstable process. *Ann. Statist.* **6** 185–190.

RUSTAGI, J. F. (1976). *Variational Methods in Statistics.* Academic, New York.

STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer, New York.

WHITE, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29** 1188–1197.

WILLIAMS, J. D. (1941). Moments of the ratio of the mean square successive differences to the mean square difference in samples from a normal universe. *Ann. Math. Statist.* **12** 239–241.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820

DEPARTAMENTO DE MATEMATICAS
PURES Y APLICADAS
UNIVERSIDAD SIMON BOLIVAR
APARTADO 89.000, CARACAS 108-A
VENEZUELA