The Annals of Statistics
1991, Vol. 19, No. 3, 1129–1142

GAUSSIAN LIKELIHOOD ESTIMATION FOR NEARLY NONSTATIONARY AR(1) PROCESSES[1]

By Dennis D. Cox

University of Illinois, Champaign-Urbana

An asymptotic analysis is presented for estimation in the three-parameter first-order autoregressive model, where the parameters are the mean, autoregressive coefficient and variance of the shocks. The nearly nonstationary asymptotic model is considered wherein the autoregressive coefficient tends to 1 as sample size tends to ∞. Three different estimators are considered: the exact Gaussian maximum likelihood estimator, the conditional maximum likelihood or least squares estimator and some "naive" estimators. It is shown that the estimators converge in distribution to analogous estimators for a continuous-time Ornstein–Uhlenbeck process. Simulation results show that the MLE has smaller asymptotic mean squared error then the other two, and that the conditional maximum likelihood estimator gives a very poor estimator of the process mean.

1. Introduction. Consider a sequence of statistical experiments with observation vector $(y_n(0), \dots, y_n(n))$ given by a three-parameter AR(1) process

$$[y_n(k+1) - \mu_n] = \varphi_n[y_n(k) - \mu_n] + \varepsilon_n(k+1), \quad k = 0, 1, \dots, n-1. \tag{1.1}$$

The shocks $\varepsilon_n(1), \dots, \varepsilon_n(n)$ are assumed i.i.d. with common distribution independent of n, and $E\varepsilon_n(1) = 0$, $E\varepsilon_n^2(1) = \sigma_0^2 < \infty$. We suppose that $|\varphi_n| < 1$ for all n and that $y_n(0)$ has the stationary distribution for the process. The parameters φ_n and μ_n will be allowed to vary with sample size [see (1.2) and (1.3)].

Suppose that the statistician models the process as Gaussian. Then the maximum likelihood estimate (MLE) of the parameter vector $(\mu_n, \sigma_0^2, \varphi_n)$, denoted $(\hat{\mu}_n, \hat{\sigma}_n^2, \hat{\varphi}_n)$, is a solution of a rather complicated system of equations. Assuming that $\mu_n \equiv \mu_0$ and $\varphi_n \equiv \varphi_0$ are fixed, then one can show that the MLE is asymptotically equivalent to a simpler estimator obtained by maximizing a conditional likelihood. The MLE maximizes the full log likelihood

$$l_n(\mu, \sigma^2, \varphi) := \log f_{\mu, \sigma^2, \varphi}(y(1), \dots, y(n) | y(0)) + \log f_{\mu, \sigma^2, \varphi}(y(0)),$$

whereas the maximum conditional likelihood estimator (MCLE) maximizes the conditional likelihood

$$\tilde{l}_n(\mu, \sigma^2, \varphi) := \log f_{\mu, \sigma^2, \varphi}(y(1), \dots, y(n) | y(0)).$$

Received May 1987; revised April 1990.
[1]Research supported by Office of Naval Research Contract N00014-84-C-0169.
Key words and phrases. Likelihood estimation, autoregressive processes, nearly nonstationary time series, Ornstein–Uhlenbeck process.

The MCLE, denoted $(\tilde{\mu}_n, \tilde{\sigma}_n^2, \tilde{\varphi}_n)$, is given by some simple formulas. See (3.12) through (3.15) below. Further details may be found in Fuller (1976), pages 328–332.

While the MLE and MCLE will be nearly the same with high probability for "sufficiently large n," they can be quite different for small to moderate n. Furthermore, the meaning of "large n" depends on the value of φ. If φ is close to 1, then the term

$$\log f(y(0)) = \frac{1}{2}\log\left[\frac{1-\varphi^2}{\sigma^2}\right] - \frac{(1-\varphi^2)[y(0)-\mu]^2}{2\sigma^2}$$

has a more pronounced effect on the log likelihood, and a much larger value of n is required before the classical asymptotic results are useful. This has been observed by Harvey (1981), page 135. As many real series exhibit large lag-one autocorrelation (hence φ near 1), it is worthwhile to investigate the MLE and MCLE under this condition. Furthermore, one is naturally interested in which estimator is better, or if some other estimator is even better than either of these. One would conjecture that the MLE is better than the MCLE, and we present results below which corroborate this conjecture. See also Ansley and Newbold (1980), page 181.

Recently, there has been much interest in "nearly nonstationary" asymptotics for such time series models; see, for example, Bobkoski (1983), Chan and Wei (1985), Tsay (1985) and Phillips (1987). These authors only consider asymptotics for the least squares estimator when σ^2 and μ are known. For the three-parameter AR(1) model, the nearly nonstationary setup corresponds to assuming that

$$\varphi_n = 1 - \beta_0/n, \qquad \beta_0 > 0, \tag{1.2}$$

$$\mu_n = n^{1/2}\nu_0, \tag{1.3}$$

where β_0 and ν_0 are fixed. Since $\varphi_n \to 1$, in some sense the process approaches a nonstationary process as $n \to \infty$. The rationale for the particular forms of μ_n and φ_n will be evident from the following discussion.

Define a continuous-time "step function" process $Y_n(t)$, $0 \le t \le 1$, by

$$Y_n(t) := n^{-1/2} y_n([nt]),$$

where $[\cdot]$ denotes the greatest integer. It follows from (1.1) that Y_n satisfies the difference equation

$$\Delta Y_n(k/n) = -\beta_0[Y_n(k/n) - \nu_0]\,\Delta t + \sigma_0\,\Delta W_n(k/n), \qquad 0 \le k \le n-1. \tag{1.4}$$

Here, $\Delta Y_n(k/n) := Y_n((k+1)/n) - Y_n(k/n)$ is a forward difference operator, $\Delta t := 1/n$, and

$$W_n(t) := \sigma_0^{-1} n^{-1/2} \sum_{k=1}^{[nt]} \varepsilon_n(k) \tag{1.5}$$

is a normalized partial sum process. Since W_n converges weakly to a Wiener process $W(t)$, $0 \le t \le 1$, in $D[0, 1]$, and the difference operator Δ converges in some sense to a differential operator d, one would expect that Y_n should converge to the solution of the stochastic differential equation

$$\begin{gathered} dY(t) = -\beta_0[Y(t) - \nu_0]\,dt + \sigma_0\,dW(t), \\ Y(0) =_D N(\nu_0, \sigma_0^2/(2\beta_0)), \\ Y(0) \text{ independent of } \{W(t)\colon 0 \le t \le 1\}, \end{gathered} \tag{1.6}$$

which defines an Ornstein–Uhlenbeck process [Arnold (1974)]. (Equality in distribution is denoted $=_D$.) In fact, Y_n converges weakly to Y [Bobkoski, (1983)].

In Section 3 this weak convergence is used to prove convergence in (joint) distribution of the MLE $(\hat{\beta}_n, \hat{\sigma}_n^2, \hat{\nu}_n) = (n(1 - \hat{\varphi}_n), \hat{\sigma}_n^2, n^{-1/2}\hat{\mu}_n)$ for the sequence of AR(1) processes given by (1.1), (1.2) and (1.3) to the corresponding MLE's of the parameters in the Ornstein–Uhlenbeck process given in (1.6). See Theorem 3.1. The MLE's for the continuous-time Ornstein–Uhlenbeck model are denoted $(\hat{\beta}, \hat{\nu})$. The MLE for the variance parameter is σ_0^2, that is, it can be determined exactly (with probability 1) from the finite sample path $\{Y(t)\colon\ 0 \le t \le 1\}$. Indeed, σ_0^2 is the only parameter which is consistently estimable from the sequence of AR(1) experiments.

In order to understand this phenomenon, it is necessary to investigate the likelihood (i.e., Radon-Nikodym derivative w.r.t. some dominating measure on path space) for the Ornstein–Uhlenbeck model. This has been done by Feigin (1976) for the situation where the only unknown parameter is β_0 and $Y(0)$ is taken as fixed. In Section 2 we give results when the mean ν_0 is also unknown. The "perfect" estimability of the variance parameter σ_0^2 results from mutual singularity of the Ornstein–Uhlenbeck measures corresponding to different variance parameters.

In Theorem 3.2 it is shown that the MCLE $(\tilde{\beta}_n, \tilde{\sigma}_n^2, \tilde{\nu}_n) = (n(1 - \tilde{\varphi}_n), \tilde{\sigma}_n^2, n^{-1/2}\tilde{\mu}_n)$ converges in distribution to $(\tilde{\beta}, \sigma_0^2, \tilde{\nu})$, where $\tilde{\beta}$ and $\tilde{\nu}$ denote the values of β and ν which maximize the conditional likelihood of the Ornstein–Uhlenbeck observation given the starting value $Y(0)$. Theorem 3.3 gives similar limiting distribution results for some "naive" estimators, namely the sample lag-one autocorrelation r_n as an estimator of $\varphi_n = (1 - \beta_0/n)$, a crude estimator s_n^2 of σ_0^2 and the sample mean of the $y_n(k)$'s as an estimator of μ_n.

While these results give representations for the asymptotic distribution of the estimators, it is unfortunately very difficult to carry out any calculations with the limiting distributions. Bobkoski (1983) gives some results when only β_0 is unknown and $y_n(0) = 0$. Of course, one can always resort to Monte Carlo, as we do in Section 4. The results of this paper do provide invariance principles so that fixed reference distributions can be developed for samples of different sizes, even if computation of the reference distributions is difficult. Furthermore, they allow one to obtain results about the limiting Ornstein–Uhlenbeck case by simulating discrete-time processes.

Some conclusions and conjectures can be drawn from the simulation results presented in Section 4. First, the MLE appears to be the best estimator in terms of mean squared error, but not significantly so. All the estimators of β_0 considered are biased upward, especially so for β_0 near 0. (Hence, the corresponding estimators of φ_n are biased downward, especially for φ_n near 1.) The MCLE estimator of the mean is quite bad, much worse than the sample mean or MLE. These results suggest that better estimators of β_0 may exist if one can reduce the bias.

The results of this paper indicate that it is important to make appropriate use of $y_n(0)$ to obtain asymptotically efficient estimation. Of course, this contradicts the usual statistical notion that the information in a single observation should be negligible in comparison with the information in the whole sample. It also makes clear the singular nature of the stationarity/nonstationarity boundary, as already noted in the works of Feigin (1976, 1979). Certainly, one must not take lightly the assumption of stationarity.

2. The Ornstein–Uhlenbeck process. In this section we present the likelihood for a continuous-time observation $\{Y(t)\colon\ 0 \le t \le 1\}$ from the Ornstein–Uhlenbeck process. The derivations are standard and hence omitted. See Feigin (1976) or Theorem 7.19 of Lipster and Shiryayev (1977). The dominating measure is a Wiener process measure modified to account for starting value and scale change. We also give results on the MCLE and MLE for the Ornstein–Uhlenbeck process.

We first give the conditional likelihood given $Y(0)$. Let $P(\cdot\,|Y(0), \nu, \sigma^2, \beta)$ be the Ornstein–Uhlenbeck measure on path space $C[0,1]$ with mean ν, scale σ and drift coefficient β, as in (1.6) with subscripts deleted. Let $Q(\cdot\,|Y(0), \sigma^2)$ denote the measure of $\sigma W(t) + Y(0)$, $0 \le t \le 1$, where W is a standard Wiener process. Let $l(\nu, \beta|Y(0), \sigma^2)$ denote the log of $dP(\cdot\,|Y(0), \nu, \sigma^2, \beta)/dQ(\cdot\,|Y(0), \sigma^2)$ evaluated at a path $\{Y(t)\colon\ 0 < t \le 1\}$. As is usual, we suppress the dependence on the data (path) in the log likelihood. Then

$$l(\nu, \beta|Y(0), \sigma^2) = -\frac{\beta}{\sigma^2}\int_0^1 [Y(t) - \nu]\, dY(t) - \frac{\beta^2}{2\sigma^2}\int_0^1 [Y(t) - \nu]^2\, dt. \tag{2.1}$$

For the unconditional likelihood, let $P(\cdot\,|\nu, \sigma^2, \beta)$ denote the Ornstein–Uhlenbeck measure when $Y(0)$ is given its stationary distribution. Let $Q(\cdot\,|\sigma^2)$ be the measure of $\sigma[W(t) + Z]$, $0 \le t \le 1$, where Z is a $N(0,1)$ random variable independent of $W(t)$, $0 \le t \le 1$. The corresponding log likelihood has extra terms from the initial conditions. It is given by

$$\begin{aligned} l(\nu, \beta|\sigma^2) = {}& \frac{1}{2}\log(2\beta) + \frac{Y(0)^2}{2\sigma^2} \\ & - \frac{\beta}{\sigma^2}\left(\int_0^1 Y(t) - \nu\, dY(t) + [Y(0) - \nu]^2\right) \\ & - \frac{\beta^2}{2\sigma^2}\int_0^1 [Y(t) - \nu]^2\, dt. \end{aligned} \tag{2.2}$$

It is easy to solve for the MCLE for β and ν from (2.1). The results are

$$\tilde{\beta} = -\frac{\int Y(t) - \bar{Y} dY(t)}{\int [Y(t) - \bar{Y}]^2 dt}, \tag{2.3}$$

$$\tilde{\nu} = \bar{Y} + (Y(1) - Y(0))/\tilde{\beta}, \tag{2.4}$$

where

$$\bar{Y} = \int_0^1 Y(t)\, dt. \tag{2.5}$$

The results concerning the MLE are somewhat harder to obtain.

THEOREM 2.1. *With probability* 1, *the MLE for* (β, ν) *exists and is unique.*

PROOF. For each fixed value of the variable β, maximization of $l(\nu, \beta|\sigma^2)$ over ν gives

$$\hat{\nu}(\beta) = (Y(1) + Y(0) + \beta\bar{Y})/(2 + \beta).$$

Substituting this back, one obtains an expression for $l(\hat{\nu}(\beta), \beta|\sigma^2)$ which tends to $-\infty$ as either $\beta \to 0$ or $\beta \to \infty$ almost surely. This establishes existence. We only sketch the proof of uniqueness. Taking $dl(\hat{\nu}(\beta), \beta|\sigma^2)/d\beta$ and setting to 0, one obtains, after some manipulation,

$$A\beta^4 + B\beta^3 + C\beta^2 + D\beta + E = 0, \tag{2.6}$$

where

$$A = -2\int (Y - \bar{Y})^2 dt < 0,$$

and the coefficients B, C, D and E are polynomial functions of the sufficient statistics

$$T = \left(Y(0), Y(1), \bar{Y}, \int Y^2\, dt, \int Y dY\right).$$

Using Itô's formula [see, e.g., Theorem 4.4, page 118, of Lipster and Shiryayev (1977)], one can show $\int Y dY$ is a polynomial in the other four components of T. Hence, for fixed β_0, ν_0 the coefficients of (2.6) are polynomial functions of the components of

$$S = \left(Y(0), Y(1), \bar{Y}, \int Y^2\, dt\right).$$

The l.h.s. of (2.6) is obtained by multiplying $dl(\hat{\nu}(\beta), \beta|\sigma^2)/d\beta$ by positive quantities, so there are at most two values of the MLE $\hat{\beta}$ since only two of the the four roots of (2.6) can have the right sign for the derivative to be that of a local maxima. It is clear that this can happen only if the roots of (2.6) are real and distinct, so we concentrate on this case. The two roots of (2.6) which are candidates for the MLE are the second and fourth, β_2 and β_4, when the roots are listed in ascending order. Now β_2 and β_4 are analytic functions of the coefficients and hence of S, and $l(\hat{\nu}(\beta), \beta|\sigma^2)$ is similarly an analytic function of $\beta > 0$ and S. Consider the three-dimensional manifold given by

$$\{S: l(\hat{\nu}(\beta_2(S)), \beta_2(S)|\sigma^2) = l(\hat{\nu}(\beta_4(S)), \beta_4(S)|\sigma^2)\},$$

which is the set of values of S wherein the MLE is not unique. The four-dimensional Lebesgue measure of this manifold is 0 by smoothness of the functions and an elementary argument [see, e.g., Lemma 1.4.3 of Narasimhan (1968)]. Hence, the MLE is unique a.s. if we show Law$(S) \ll m^4$. Note that the first three components

$$R = (Y(0), Y(1), \bar{Y})$$

have a nonsingular normal distribution on $\mathbb{R}^3$, and

$$Z(t) = Y(t) - E[Y(t)|R]$$

is a Gaussian process independent of R. Conditional on $R = r$, the random variable

$$\int Y^2\,dt = \int_0^1 \{Z(t) + E[Y(t)|R = r]\}^2\,dt$$

clearly has a distribution $\ll m$. Note, for instance, that by a Karhunen–Loève expansion one can represent $\int Y^2\,dt$ as a weighted sum of independent noncentral chi-squared random variables. Hence, Law$(S) \ll m^4$, which completes the proof. □

3. Main theorems. This section contains the statements and proofs of the claims that the parameter estimates for the nearly nonstationary AR(1) converge to their analogs for the Ornstein–Uhlenbeck process. The first theorem concerns the MLE and the second concerns the MCLE. The third theorem is about some "naive" estimators. Throughout this section we assume that the $\varepsilon_n(k)$'s are i.i.d. with mean 0 and variance σ^2.

We first mention a couple of facts that will be useful. As noted before, W_n converges weakly to W, and by a well-known construction [see, e.g., Theorem 13.8 of Breiman (1968)], we may in fact assume

$$\sup_{0 \le t \le 1} |W(t) - W_n(t)| \to_P 0.$$

This requires changing to a new probability space, and all results so obtained translate into weak convergence back on the original space. It is not hard then to show

$$\sup_{0 \le t \le 1} |Y_n(t) - Y(t)| \to_P 0. \tag{3.1}$$

Also, one may show

$$\sum_{k=0}^{n-1} Y_n(k/n)\,\Delta W_n(k/n) \to_P \int_0^1 Y(t)\,dW(t). \tag{3.2}$$

See Bobkoski (1983) for proofs.

THEOREM 3.1. *Let $(\hat{\mu}_n, \hat{\sigma}_n^2, \hat{\varphi}_n)$ be the MLE of $(\mu_n, \sigma_0^2, \varphi_n)$ in the AR(1) model given in (1.1) through (1.3). Let $(\hat{\nu}, \hat{\beta})$ be the MLE of (ν_0, β_0) in the Ornstein–Uhlenbeck model in (1.6) when σ_0^2 is known. Then*

$$\begin{pmatrix} n^{-1/2}\hat{\mu}_n \\ \hat{\sigma}_n^2 \\ n(1-\hat{\varphi}_n) \end{pmatrix} \to_D \begin{pmatrix} \hat{\nu} \\ \sigma_0^2 \\ \hat{\beta} \end{pmatrix}. \tag{3.3}$$

PROOF. We will use the variables $n^{1/2}\nu$ in place of μ and $1 - \beta/n$ in place of φ. Inessential constants in the log likelihood will be dropped. The first step is to eliminate ν and σ^2 from the likelihood maximization problem. The log likelihood can be written as

$$\begin{aligned} l_n\left(n^{1/2}\nu, \sigma^2, 1 - \frac{\beta}{n}\right) = & -\frac{n+1}{2}\log\sigma^2 - \frac{n}{2\sigma^2}s_n^2 \\ & + \frac{1}{2}\log\beta + \frac{1}{2}\log\left(1 - \frac{\beta}{2n}\right) \\ & - \frac{B_n(\nu)\beta}{\sigma^2} - \frac{A_n(\nu)\beta^2}{\sigma^2}, \end{aligned}$$

where

$$\begin{aligned} s_n^2 &:= \sum \Delta Y_n(k/n)^2, \\ A_n(\nu) &:= \frac{1}{2}\left\{-n^{-1}[Y_n(0) - \nu]^2 + \sum [Y_n(k/n) - \nu]^2 \Delta t\right\}, \\ B_n(\nu) &:= [Y_n(0) - \nu]^2 + \sum [Y_n(k/n) - \nu]\, \Delta Y_n(k/n). \end{aligned}$$

All summations in this proof are from $k = 0$ to $n - 1$, unless otherwise indicated. For any fixed values of σ^2 and β,

$$\hat{\nu}_n(\beta) := (2 + \beta(1 - 1/n))^{-1}\left(Y_n(0)(1 - \beta/n) + Y_n(1) + \beta\sum Y_n(k/n)\,\Delta t\right)$$

maximizes l_n over ν. Note that $\sup_{0\le\beta<\infty}|\hat{\nu}_n(\beta)|$ is bounded in probability, since all of the random variables appearing in the defining expression are bounded in probability by (3.1) and (3.2), and $\beta \ge 0$. Since A_n and B_n are continuous and A_n is bounded below by a function of Y_n only, this implies that $\forall\ \varepsilon > 0$, $\exists\ C_1, C_2 > 0$, $C_3, C_4 > 0$, and N such that $\forall\ n \ge N$, the event

$$\begin{aligned} E_n := \big[C_1 + C_2\beta \le B_n(\hat{\nu}_n(\beta))\beta + A_n(\hat{\nu}_n(\beta))\beta^2 \\ \le C_3\beta + C_4\beta^2 \text{ for } \forall\ \beta \ge 0\big] \end{aligned}$$

satisfies

$$P(E_n) \ge 1 - \varepsilon. \tag{3.4}$$

For each fixed value of β,

$$\hat{\sigma}_n^2(\beta) := \frac{n}{n+1}s_n^2 + \frac{2}{n+1}\left[B_n(\hat{\nu}_n(\beta))\beta + A_n(\hat{\nu}_n(\beta))\beta^2\right]$$

maximizes over σ^2 the function $l_n(n^{1/2}\hat{\nu}_n(\beta), \sigma^2, 1-\beta/n)$, provided $\hat{\sigma}_n^2(\beta) > 0$. Note that on the event E_n, $\hat{\sigma}_n^2(\beta) > 0$ for all n sufficiently large. Also, we have

$$\begin{aligned} s_n^2 = \sigma_0^2 \sum \left[\Delta W_n(k/n)\right]^2 - 2n^{-1}\beta_0\sigma_0 \sum Y_n(k/n)\,\Delta W_n(k/n) \\ + n^{-1}\beta_0^2 \sum Y_n^2(k/n)\,\Delta t. \end{aligned} \tag{3.5}$$

The first term on the r.h.s. of (3.5) $\to_P \sigma_0^2$ by the weak law of large numbers, while the other two terms are $O_P(n^{-1})$.

With a little algebra, there results

$$\begin{aligned} &2l_n\left(n^{1/2}\hat{\nu}_n(\beta), \hat{\sigma}_n^2(\beta), 1-\beta/n\right) \\ &\quad = -(n+1)\log \hat{\sigma}_n^2(\beta) + \log\beta + \log(1-\beta/(2n)). \end{aligned} \tag{3.6}$$

The next step of the proof consists of showing that $\hat{\beta}_n$ is bounded away from 0 and ∞ in probability. Using $\log x \le x - 1$, $\forall\, x > 0$, on the event E_n we have

$$\begin{aligned} &2l_n\left(n^{1/2}\hat{\nu}_n(\beta), \hat{\sigma}_n^2(\beta), 1-\beta/n\right) + (n+1)\log s_n^2 \\ &\quad \ge \left(2/s_n^2\right)\left[C_3\beta + C_4\beta^2\right] + \log\beta, \qquad \forall\, \beta \in (0, 2n). \end{aligned} \tag{3.7}$$

For all n sufficiently large, the expression on the r.h.s. of (3.7) achieves a maximum at some β_n^* in $(0, 2n)$, and $\beta_n^* \to_P \beta^*$, say. When β_n^* is plugged into the r.h.s. of (3.7), the resulting expression converges in probability to a constant. Since the supremum of a lower bound on the likelihood function provides a lower bound on the maximum of the likelihood, it follows that

$$\begin{aligned} &\forall\, \varepsilon > 0, \exists\, m, N \text{ such that } \forall\, n \ge N, \\ &P\Big(2l_n\left(n^{1/2}\hat{\nu}_n(\hat{\beta}_n), \hat{\sigma}_n^2(\hat{\beta}_n), 1-\hat{\beta}_n/n\right) \\ &\qquad + (n+1)\log s_n^2 \ge m\Big) \ge 1-\varepsilon. \end{aligned} \tag{3.8}$$

Hence, the MLE $\hat{\beta}_n$ is with high probability in the set of $\beta \in (0, 2n)$ which satisfy the inequality in the event in (3.8). In view of the definition of $\hat{\sigma}_n^2(\beta)$ and the lower bound in E_n and (3.5), we may restrict attention to the set of β's satisfying $0 \le \beta \le 2n$ and for some constants $C_5, C_6 > 0$, and m,

$$G_n(\beta) := -(n+1)\log\left[1 + \frac{1}{n+1}(C_5 + C_6\beta)\right] + \log\beta \ge m. \tag{3.9}$$

It is easy to check that G_n is maximized at a point $\beta_n^{**} \to C_6^{-1}$, that $G_n(\beta_n^{**}) \to -(C_5+1) - \log C_6$ and that $G_n''(\beta)$ is eventually less than $c < 0$

for all β, where c is a constant. These facts imply that there is a constant $b > 0$ such that eventually all values of β satisfying (3.9) also satisfy $\beta \le b$. Now $G_n(\beta) \to -[C_5 + C_6\beta] + \log\beta$ as $n \to \infty$, uniformly in $\beta \in (0, b]$, and the limit function crosses from above the level m at some positive value larger than β_n^{**}. For $0 < \beta \le b$, $G_n(\beta) \le C + \log\beta$ for all sufficiently large n, where C is some constant, so G_n must also cross the level m at some point in the interval $(0, \beta_n^{**})$. Hence,

$$\forall\, \varepsilon > 0, \exists\, a > 0,\ b > a,\ N \text{ such that } \forall\, n \ge N, \quad P\Big[\hat{\beta}_n \text{ exists and } a \le \hat{\beta}_n \le b\Big] \ge 1 - \varepsilon. \tag{3.10}$$

It now follows that the MLE $(\hat{\nu}_n, \hat{\sigma}_n^2, \hat{\beta}_n) = (\hat{\nu}_n(\hat{\beta}_n), \hat{\sigma}_n^2(\hat{\beta}_n), \hat{\beta}_n)$ exists with arbitrarily high probability for all n sufficiently large, and furthermore that $\hat{\beta}_n$ is bounded away from 0 and ∞ in probability. Now $\hat{\nu}_n(\beta)$ converges in probability uniformly in $\beta \in [0, b]$ to

$$\hat{\nu}(\beta) := \frac{1}{2+\beta}\left(Y(0) + Y(1) + \beta \int Y(t)\, dt\right), \tag{3.11}$$

and $\hat{\sigma}_n^2(\beta) = s_n^2 + O_P(n^{-1}) \to_P \sigma_0^2$, uniformly in $\beta \in [0, b]$. Hence, using (3.1) and (3.2), $l_n(n^{1/2}\hat{\nu}_n(\beta), \hat{\sigma}_n^2(\beta), 1 - \beta/n) + [(n+1)/2]\log\sigma_0^2 + n/2$ converges in probability uniformly in $\beta \in (0, b]$ to $l(\hat{\nu}_n(\beta), \beta)$, where

$$l(\nu, \beta) := \frac{1}{2}\log\beta - \frac{B(\nu)\beta}{\sigma_0^2} - \frac{A(\nu)\beta^2}{\sigma_0^2},$$

$$B(\nu) := [Y(0) - \nu]^2 + \int Y(t) - \nu\, dY(t),$$

$$A(\nu) := \frac{1}{2}\int [Y(t) - \nu]^2\, dt.$$

Now $l(\nu, \beta)$ is the likelihood for the Ornstein–Uhlenbeck process estimation problem (with σ_0^2 known, of course), and $\hat{\nu}(\beta)$ is clearly the MLE of ν for each fixed β. It also follows that $\hat{\beta}_n \to_P \hat{\beta}$ by the following simple fact. Suppose a sequence of functions f_n converges uniformly to f, and for each n, x_n is any maximizer of f_n. Then any limit point of the sequence $\{x_n\}$ is a maximizer of f. Uniqueness of $\hat{\beta}$ then gives the desired result. The proof is complete. □

Now consider the MCLE. First, define

$$\bar{y}_{n0} = \frac{1}{n}\sum_{t=0}^{n=1} y_n(t). \tag{3.12}$$

Then the MCLE's are given by

$$(3.13)\qquad \tilde{\varphi}_n = \frac{\sum_{t=0}^{n-1}[y_n(t) - \bar{y}_{n0}][y_n(t+1) - \bar{y}_{n0}]}{\sum_{t=0}^{n-1}[y_n(t) - \bar{y}_{n0}]^2},$$

$$(3.14)\qquad \tilde{\mu}_n = \bar{y}_{n0} + \frac{y_n(n) - y_n(0)}{n(1-\tilde{\varphi}_n)},$$

$$(3.15)\qquad \tilde{\sigma}_n^2 = \frac{1}{n}\sum_{t=0}^{n-1}[y_n(t+1) - \tilde{\varphi}_n y_n(t) - (1-\tilde{\varphi}_n)\tilde{\mu}_n]^2.$$

The corresponding MCLE's for the Ornstein–Uhlenbeck process are given in (2.3) through (2.5). The following theorem can be proved more simply than the previous one by simply using the explicit formulas for the estimators and (3.1) and (3.2).

THEOREM 3.2. *As* $n \to \infty$,

$$\begin{pmatrix} n^{-1/2}\tilde{\mu}_n \\ \tilde{\sigma}_n^2 \\ n(1-\tilde{\varphi}_n) \end{pmatrix} \to_D \begin{pmatrix} \tilde{\nu} \\ \sigma_0^2 \\ \tilde{\beta} \end{pmatrix}.$$

Finally, we consider some "naive" estimators. Let

$$(3.16)\qquad \bar{y}_{n1} = \frac{1}{n}\sum_{t=0}^{n-1} y_n(t+1),$$

$$(3.17)\qquad \bar{y}_n = \frac{1}{n+1}\sum_{t=0}^{n} y_n(t),$$

$$(3.18)\qquad s_n^2 = \frac{1}{n}\sum_{t=0}^{n-1}[y_n(t+1) - y_n(t)]^2,$$

$$(3.19)\qquad r_n = \frac{\sum_{t=0}^{n-1}[y_n(t+1) - \bar{y}_{n1}][y_n(t) - \bar{y}_{n0}]}{\left(\left\{\sum_{t=0}^{n-1}[y_n(t+1) - \bar{y}_{n1}]^2\right\}\left\{\sum_{t=0}^{n-1}[y_n(t) - \bar{y}_{n0}]^2\right\}\right)^{1/2}}.$$

We refer to $\bar{y}_n$, s_n^2 and r_n as the naive estimators of μ_n, σ_0^2 and φ_n, respectively. Note that r_n is precisely the bivariate correlation of the pairs $(y_n(0), y_n(1)), (y_n(1), y_n(2)), \dots, (y_n(n-1), y_n(n))$, and so $-1 < r_n < 1$, a.s. It is important to use $\bar{y}_{n0}$ and $\bar{y}_{n1}$ in the definition of r_n (rather than $\bar{y}_n$) to ensure r_n is in $(-1, 1)$.

THEOREM 3.3. *Let*

$$(3.20)\qquad \bar{\beta} = \frac{\frac{1}{2}[Y(1) - Y(0)][Y(1) + Y(0) + 2\bar{Y}] - \int Y(t) - \bar{Y} dY(t)}{\int [Y(t) - \bar{Y}]^2\, dt}.$$

Then as $n \to \infty$,

$$\begin{pmatrix} n^{-1/2}\bar{y}_n \\ s_n^2 \\ n(1-r_n) \end{pmatrix} \to_D \begin{pmatrix} \bar{Y} \\ \sigma_0^2 \\ \bar{\beta} \end{pmatrix}.$$

PROOF. We will assume as in (3.1) and (3.2) that all convergences are taking place on a common probability space so that we may use convergence in probability rather than convergence in distribution, and $\to$ will mean $\to_P$ for the remainder of the proof. Now it is clear from (3.1) that

$$n^{-1/2}\bar{y}_n \to \bar{Y} \quad \text{and} \quad n^{-1/2}\bar{y}_{ni} \to \bar{Y}, \qquad i = 0, 1.$$

Also, $s_n^2 \to \sigma_0^2$ as already noted below (3.5). Thus, we need only take care of the convergence result on r_n. Put

$$S_{ni}^2 = \frac{1}{n}\sum\left[y(t+i) - \bar{y}_{ni}\right]^2, \qquad i = 0, 1,$$

$$S^2 = \int\left[Y(t) - \bar{Y}\right]^2 dt.$$

The result (3.1) also implies that $n^{-1}S_{ni}^2 \to S^2$ as $n \to \infty$. Some algebra will show that

$$n(1-r_n) = \frac{nS_{n0}(S_{n1} - S_{n0}) - \Sigma\left[y(t) - \bar{y}_0\right]\Delta y(t)}{S_{n0}S_{n1}}. \tag{3.21}$$

Now

$$\begin{aligned} S_{n0}(S_{n1} - S_{n0}) &= \frac{S_{n0}}{S_{n1} + S_{n0}}\frac{1}{n}[y(n) - y(0)][y(n) + y(0) + \bar{y}_1 + \bar{y}_0] \\ &\to \frac{1}{2}[Y(1) - Y(0)]\left[Y(1) + Y(0) + 2\bar{Y}\right]. \end{aligned}$$

If one multiplies the numerator and denominator in (3.21) by n^{-1} and uses this latter along with (3.2) the desired result follows. □

4. Monte Carlo results. Tables 1 through 3 present the results of a simulation study of the various estimators. The simulation program used the IMSL subroutine GGNML to generate $n+1$ pseudorandom variates which were used to construct AR(1) sample paths according to the model (1.1). We considered three estimators of β_0 and ν_0 (the naive, MCLE and MLE) and four estimators of σ_0^2 ($\tilde{\sigma}_n^2$ is the ordinary sample variance). All estimators except the MLE were computed directly from the formulas. The MLE was computed by a Newton type algorithm using finite difference approximations to the derivatives of the log likelihood function as a function of β with ν and σ^2 substituted out, as in the proof of Theorem 3.1. The naive estimator was used as the starting value, and convergence was quite fast, requiring on the

TABLE 1
Summary of simulation results for estimators of β_0

β_0	n	Estimator	Bias	Mean squared error
5	100	r_n	4.38 (0.03)	42.27 (0.53)
		$\tilde{\beta}_n$	4.37 (0.03)	43.58 (0.54)
		$\hat{\beta}_n$	4.48 (0.03)	39.49 (0.51)
5	500	r_n	4.55 (0.03)	46.11 (0.57)
		$\tilde{\beta}_n$	4.55 (0.03)	47.89 (0.59)
		$\hat{\beta}_n$	4.22 (0.03)	42.87 (0.55)
2	100	r_n	4.68 (0.03)	40.21 (0.46)
		$\tilde{\beta}_n$	4.68 (0.03)	42.00 (0.48)
		$\hat{\beta}_n$	4.27 (0.03)	36.21 (0.44)

Note: For all cases, $\nu_0 = 1$ and $\sigma_0^2 = 1$. Estimated standard errors are shown in parentheses next to the figure.

TABLE 2
Summary of simulation results for estimators of ν_0

β_0	n	Estimator	Mean squared error
5	100	$\bar{y}_n$	0.032 (0.000)
		$\tilde{\nu}_n$	0.376 (0.291)
		$\hat{\nu}_n$	0.029 (0.000)
5	500	$\bar{y}_n$	0.032 (0.000)
		$\tilde{\nu}_n$	0.172 (0.061)
		$\hat{\nu}_n$	0.030 (0.000)
2	100	$\bar{y}_n$	0.139 (0.001)
		$\tilde{\nu}_n$	286 (257)
		$\hat{\nu}_n$	0.125 (0.001)

Note: For all cases, $\nu_0 = 1$ and $\sigma_0^2 = 1$. Estimated standard errors are shown in parentheses next to the figure.

TABLE 3
Summary of simulation results for estimates of σ_0^2

β_0	n	Estimator	Mean squared error
5	100	All	0.020 (0.000)
5	500	All	0.0040 (0.0003)
2	100	All	0.020 (0.000)

Note: For all cases, $\nu_0 = 1$ and $\sigma_0^2 = 1$. Estimated standard errors are shown in parentheses next to the figure.

average less than two iterations of the Newton algorithm. The results were compared with those of the SAS statistical package on selected sample paths in order to validate the program. All results are based on 25,000 Monte Carlo replications.

The results indicate that the MLE is the best of the estimators considered in terms of mean squared error, although not by much in comparison with the naive. Two surprising results emerge. First, all estimators of β_0 are badly biased, with the bias becoming worse as β_0 becomes smaller. It should be possible to find improved estimators of β_0 by "shrinking" toward 0, with the amount of "shrinkage" becoming larger as say the sample lag-one autocorrelation becomes larger. The bias in the estimators of the other parameters was negligible compared to the variance and so is omitted. A second surprising result is the poor performance of the MCLE of the location ν_0, particularly as β_0 becomes smaller. This is also the widely used least squares estimator of location. The main problem here is the term $(y(n) - y(0))/\tilde{\beta}$ [see (3.14)], which severely inflates the variance. Results presented by Bobkoski (1983) indicate that there is some probability of obtaining $\tilde{\beta}$ close to 0 (it may even be negative, which is why $\tilde{\beta}$ was not used as the starting value for the iterative calculation of the MLE). Alternatively, if one considers the Ornstein–Uhlenbeck observation, with β known, the Fisher information about ν in the conditional likelihood is

$$I(\nu|Y(0);\beta) = \beta^2/\sigma^2,$$

whereas for the full likelihood (with β known)

$$I(\nu;\beta) = (\beta^2 + 2\beta)/\sigma^2.$$

For small β, the latter is much larger than the former. This inaccuracy in $\tilde{\nu}$ does not seem to present a problem for the other parameter estimates β or $\tilde{\sigma}^2$. As the MCLE is in general the worst of the estimators, we suggest that one use either the naive estimators or the full MLE, until something better is found.

REFERENCES

ANSLEY, C. and NEWBOLD, P. (1980). Finite sample properties of estimators for autoregressive moving average models. *J. Econometrics* **13** 159–184.

ARNOLD, L. (1974). *Stochastic Differential Equations: Theory and Applications*. R. Oldenbourg, Munich.

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

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

CHAN, N. H. and WEI, C. Z. (1985). Asymptotic inference for nearly nonstationary AR(1) processes. Technical Report, Dept. Mathematics, Univ. Maryland, College Park.

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

CUMBERLAND, W. G. and SYKES, Z. M. (1982). Weak convergence of an autoregressive process used in modelling population growth. *J. Appl. Probab.* **79** 355–367.

FEIGIN, P. D. (1976). Maximum likelihood estimation for continuous time stochastic processes. *Adv. in Appl. Probab.* **8** 712–736.

FEIGIN, P. D. (1979). Some comments concerning a curious singularity. *J. Appl. Probab.* **16** 440–444.

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

HARVEY, A. C. (1981). *Time Series Models*. Halsted, New York.

LIPSTER, R. S. and SHIRYAYEV, A. N. (1977). *Statistics of Random Processes. I. General Theory*. Springer, New York.

NARASIMHAN, R. (1968). *Analysis on Real and Complex Manifolds*. Masson and Cie, Paris.

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

TIAO, G. C. and TSAY, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Statist.* **11** 856–871.

TSAY, R. S. (1985). Asymptotic inference of least squares estimates in nearly nonstationary time series. Technical Report, Dept. Statistics, Carnegie Mellon Univ.

DEPARTMENT OF STATISTICS
101 ILLINI HALL
UNIVERSITY OF ILLINOIS
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820