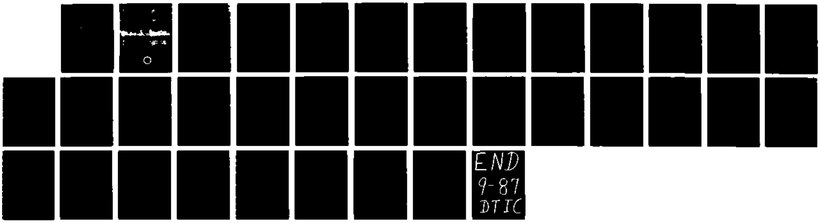
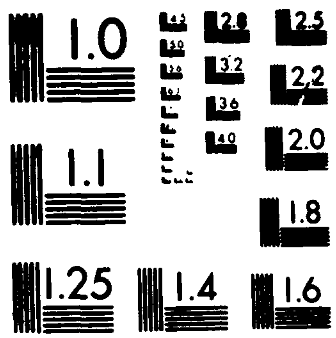
AD-A183 534 GAUSSIAN LIKELIHOOD ESTIMATION FOR NEARLY NONSTATIONARY AR(1) PROCESSES(U) WASHINGTON UNIV SEATTLE DEPT OF STATISTICS D D COX JUL 87 TR-98 N00014-84-C-0169 1/1

UNCLASSIFIED F/G 12/3 NL





MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A

GAUSSIAN LIKELIHOOD ESTIMATION FOR NEARLY NONSTATIONARY AR(1) PROCESSES

by

Dennis D. Cox

TECHNICAL REPORT No. 98

July 1987

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA

GAUSSIAN LIKELIHOOD ESTIMATION

FOR NEARLY NONSTATIONARY AR(1) PROCESSSES

by Dennis D. Cox
Department of Statistics
University of Illinois
Champaign, Illinois

This research was supported by the Office of Naval Research under Contract N00014-84-C-0169.

87 7 24 103

ABSTRACT

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 infinity. 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 than the other two, and that the conditional maximum likelihood estimator gives a very poor estimator of the process mean.

Key Words and Phrases: likelihood estimation, autoregressive processes, nearly nonstationary time series, Ornstein-Uhlenbeck process.

1. INTRODUCTION.

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

$$(1.1) \qquad [y_n(k+1) - \mu_n] = \gamma_n[y_n(k) - \mu_n] + \epsilon_n(k+1),$$
$$k = 0, 1, \ldots, n-1,$$

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

Suppose that the statistician models the process as Gaussian. Then the maximum likelihood estimate (MLE) of the parameter vector $(\mu_n, \sigma_0^2, \gamma_n)$, denoted $(\hat{\mu}_n, \hat{\sigma}_n^2, \hat{\gamma}_n)$, is a solution of a rather complicated system of equations. Assuming that $\mu_n \equiv \mu_0$ and $\gamma_n \equiv \gamma_0$ 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

$$\ell_n(\mu,\sigma^2,\gamma) := \log f_{\mu,\sigma^2,\gamma}(y(1), \ldots, y(n)|y(0))$$
$$+ \log f_{\mu,\sigma^2,\gamma}(y(0)),$$

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

$$\tilde{\ell}_n(\mu,\sigma^2,\gamma) := \log f_{\mu,\sigma^2,\gamma}(y(1), \ldots, y(n)|y(0)).$$

The MCLE, denoted $(\tilde{\mu}_n, \tilde{\sigma}^2_n, \tilde{\gamma}_n)$, is given by some simple formulae. 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 γ. If γ is close to 1, then the term

$$\log f(y(0)) = (1/2)\log[(1-\gamma^2)/\sigma^2] - (1-\gamma^2)[y(0)^2-\mu]/(2\sigma^2)$$

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. As many real series exhibit large lag one autocorrelation (hence γ 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.

Recently, there has been much interest in "nearly nonstationary" asymptotics for such time series models. See e.g. Bobkoski (1983), Chan and Wei (1985), and Tsay (1985). For the three parameter AR(1) model, this corresponds to assuming that

(1.2) $$\gamma_n = 1-\beta_0/n \; , \; \beta_0>0,$$

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

where β_0 and ν_0 are fixed. Since $\gamma_n \to 1$, in some sense the process approaches a nonstationary process as $n\to\infty$. The rationale for the particular forms of μ_n and γ_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

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

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

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

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

$$dY(t) = -\beta_0[Y(t)-\nu_0]dt + \sigma_0 dW(t), \tag{1.6}$$

$$Y(0) \overset{D}{=} N(\nu_0, \sigma_0^2/(2\beta_0)),$$

$$Y(0) \text{ independent of } \{W(t): 0 \le t \le 1\},$$

which defines an Ornstein-Uhlenbeck process. (Equality in distribution is denoted $\overset{D}{=}$.) The weak convergence of Y_n to Y follows from Lemma A.1. in the Appendix.

In Section 3 this weak convergence is used to prove convergence in (joint) distribution of the MLE $(\hat{\beta}_n, \hat{\sigma}_n^2, \hat{\nu}_n) = (n(1-\hat{\gamma}_n), \hat{\sigma}_n^2, n^{-1/2}\hat{\mu}_n)$ 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 in Section 3. 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, i.e. it can be determined exactly (with probability one) from the finite sample path $\{Y(t): 0 \le t \le 1\}$. 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 and to define the MLE in the Ornstein-Uhlenbeck model, it is necessary to develop 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 (i.e. that author derives the conditional likelihood). In Section 2, we extend those results to the case where the mean ν_0 is also unknown, and discuss the "perfect" estimability of the variance parameter σ_0^2, which results from mutual singularity of the Ornstein-Uhlenbeck measures corresponding to different variance parameters.

In Theorem 3.2 in Section 3 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. Firstly, the MLE appears to be 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 r_n are biased downward, especially for r_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.

2. THE ORNSTEIN-UHLENBECK PROCESS.

In this section we derive the likelihood for a continuous time observation $\{Y(t): 0 \le t \le 1\}$ from the Ornstein-Uhlenbeck process. The derivation is standard (see scheme (i) on p. 714 of Feigin, 1976), so it will only be sketched. The dominating measure is a Wiener process measure modified to account for starting value and scale change. Calculate the likelihood ratio of the finite dimensional vector $(Y(0), Y(1/n), Y(2/n), \ldots, Y(1))$ under the Ornstein-Uhlenbeck measure (numerator) and Wiener measure (denominator) and let $n \to \infty$ through the values $n = 2^k$.

We first derive the conditional likelihood given $Y(0)$ as it has a simpler form than the unconditional likelihood. The latter can then be obtained by modification of the former. Let $P(\cdot \mid Y(0), \nu, \sigma^2, \beta)$ 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 $Q(\cdot \mid Y(0), \sigma^2)$ denote the measure of $\sigma W(t) + Y(0)$, $0 \le t \le 1$, where W is a standard Wiener process. For the Ornstein-Uhlenbeck process we have the following integral representation valid for any $t \ge s$:

$$(2.1) \qquad Y(t)-\nu = \exp[-\beta(t-s)]\,[Y(s)-\nu] + \sigma \int_s^t \exp[-\beta(t-x)]\,dW(x).$$

See e.g. Section 8.3 of Arnold (1974). Thus, the sampled process $Y(0), Y(1/n), \ldots$ an AR(1) process with mean ν, autoregressive coefficient $\exp[-\beta/n]$, and shock variance $\sigma^2(1-\exp[-2\beta/n])$. Using this, the likelihood ratio can be shown to equal

$$\text{(2.2)} \quad \exp\left\{ \frac{n}{2}\log\left[\frac{2\beta/n}{1-\exp[-2\beta/n]}\right] - \frac{n}{2\sigma^2}\left[\left[\frac{2\beta/n}{1-\exp[-2\beta/n]} - 1\right]\right]\sum_{i=0}^{n-1}[\Delta Y(i/n)]^2 - \frac{2\beta(1-\exp[-\beta/n])}{\sigma^2(1-\exp[-2\beta/n])}\sum_{i=0}^{n-1}[Y(i/n)-\nu]\Delta Y(i/n) - \frac{\beta n(1-\exp[-\beta/n])^2}{\sigma^2(1-\exp[-2\beta/n])}\sum_{i=0}^{n-1}[Y(i/n)-\nu]^2(1/n) \right\}.$$

As in (5.3) of Feigin (1976), we have

$$\text{(2.3)} \quad \sum_{i=0}^{n-1}[\Delta Y(i/n)]^2 \xrightarrow{P} \sigma^2.$$

The convergence is $Q(\cdot \mid Y(0), \sigma^2)$-almost sure if $n = 2^k$ and $k\to\infty$, but is always true in probability by a Chebyshev argument with respect to either $P(\cdot \mid Y(0), \nu, \sigma^2, \beta)$ or $Q(\cdot \mid Y(0), \sigma^2)$. Some calculus will show that the first two terms in the exponent in (2.2) cancel each other. After computing the limits of the third and fourth terms, one obtains that the log likelihood is equal to

$$\text{(2.4)} \quad \ell(\nu,\beta \mid Y(0),\sigma^2) = -\frac{\beta}{\sigma^2}\int_0^1[Y(t)-\nu]dY(t) - \frac{\beta^2}{2\sigma^2}\int_0^1[Y(t)-\nu]^2dt.$$

For the unconditional likelihood, let $P(\cdot \mid \nu, \sigma^2, \beta)$ denote the Ornstein-Uhlenbeck measure when Y(0) is given its stationary distribution. Let $Q(\cdot \mid \sigma^2)$ 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 likelihood ratios contain extra terms in the exponent from the ratio of initial distributions. These are easy to analyze and the likelihood turns out to be

$$\ell(\nu,\beta|\sigma^2) = \frac{1}{2}\log(2\beta) + \frac{Y(0)^2}{2\sigma^2} - \frac{\beta}{\sigma^2}\left[\int_0^1 [Y(t)-\nu]dY(t) + [Y(0)-\nu]^2\right] - \frac{\beta}{2\sigma^2}\int_0^1 [Y(t)-\nu]^2 dt. \tag{2.5}$$

It is easy to solve for the MCLE for β and ν from (2.4). The results are

$$\tilde{\beta} = -\frac{\int [Y(t)-\bar{Y}]\, dY(t)}{\int [Y(t)-\bar{Y}]^2\, dt} \tag{2.6}$$

$$\tilde{\nu} = \bar{Y} + (Y(1)-Y(0))/\tilde{\beta}, \tag{2.7}$$

where

$$\bar{Y} = \int_0^1 Y(t)\, dt. \tag{2.8}$$

The MLE also exists, but is not so easy to obtain. One can solve for the minimizer over ν of $\ell(\nu,\beta|\sigma^2)$ for each fixed β, plug that back in, and then note that the resulting expression as a function of β tends to $-\infty$ as $\beta \to 0$ or $\beta \to \infty$. This shows that the MLE exists.

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 analogues 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.

THEOREM 3.1. Let $(\hat{\mu}_n, \hat{\sigma}^2_n, \hat{r}_n)$ be the MLE of (μ_n, σ^2_0, r_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^2_0 is known. Then

$$(3.1)\qquad \begin{bmatrix} n^{-1/2}\hat{\mu}_n \\ \hat{\sigma}^2_n \\ n(1-\hat{r}_n) \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \hat{\nu} \\ \sigma^2_0 \\ \hat{\beta} \end{bmatrix}.$$

PROOF. We will use the variables $n^{1/2}\nu$ in place of μ and $1-\beta/n$ in place of r. 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

$$(3.2)\qquad \ell_n(n^{1/2}\nu,\sigma^2,1-\beta/n) = -((n+1)/2)\log\sigma^2 - n/(2\sigma^2)s_n^2$$

$$+ (1/2)\log\beta + (1/2)\log(1-\beta/(2n)) - B_n(\nu)\beta/\sigma^2 - A_n(\nu)\beta^2/\sigma^2,$$

where

$$s_n^2 := \sum \Delta Y_n(k/n)^2,$$

$$A_n(\nu) := (1/2)\left\{-n^{-1}[Y_n(0)-\nu]^2 + \sum[Y_n(k/n)-\nu]^2\Delta t\right\}$$

$$B_n(\nu) := [Y_n(0)-\nu]^2 + \sum[Y_n(k/n)-\nu]\Delta Y_n(k/n).$$

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

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

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 Lemmas A.1 and A.2, and $\beta\ge0$. Since A_n and B_n are continuous and A_n is bounded below by a function of Y_n only, this implies that $\forall\epsilon>0$, $\exists\ C_1$, $C_2>0$, C_3, $C_4>0$, and N such that $\forall n\ge N$, the event

$$E_n := \left[C_1 + C_2\beta \le B_n(\hat{\nu}_n(\beta))\beta + A_n(\hat{\nu}_n(\beta))\beta^2 \right.$$

$$\left. \le C_3\beta + C_4\beta^2, \text{ for } \forall\beta\ge 0 \right]$$

satisfies

(3.3) $$P(E_n) \ge 1-\epsilon.$$

For each fixed value of β,

(3.4) $$\hat{\sigma}_n^2(\beta) := [n/(n+1)]s_n^2 +$$

$$[2/(n+1)]\left[B_n(\hat{\nu}_n(\beta))\beta + A_n(\hat{\nu}_n(\beta))\beta^2 \right]$$

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

(3.5) $$s_n^2 = \sigma_0^2\sum[\Delta W_n(k/n)]^2$$

$$- 2n^{-1}\beta_0\sigma_0\sum Y_n(k/n)\Delta W_n(k/n) + n^{-1}\beta_0^2\sum Y_n^2(k/n)\Delta t.$$

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

With a little algebra, there results

$$(3.6) \qquad 2\ell_n(n^{1/2}\hat{\nu}_n(\beta),\ \hat{\sigma}_n^2(\beta),\ 1-\beta/n) =$$

$$-(n+1)\log \hat{\sigma}_n^2(\beta) + \log \beta + \log (1-\beta/(2n))$$

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

$$(3.7) \qquad 2\ell_n(n^{1/2}\hat{\nu}_n(\beta),\ \hat{\sigma}_n^2(\beta),\ 1-\beta/n) + (n+1)\log s_n^2 \ge$$

$$(2/\tilde{\sigma}_n^2)[C_3\beta + C_4\beta^2] + \log \beta, \qquad \forall \beta \in (0,2n).$$

For all n sufficiently, the expression on the r.h.s. of (3.7) achieves a maximum at some β_n^* in $(0,2n)$, and $\beta_n^* \xrightarrow{P} \beta^*$, 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

$$(3.8) \qquad \forall \epsilon>0,\ \exists m,\ N, \text{ such that } \forall n\ge N,$$

$$P\left[2\ell_n(n^{1/2}\hat{\nu}_n(\hat{\beta}_n),\ \hat{\sigma}_n^2(\hat{\beta}_n),\ 1-\hat{\beta}_n/n) + (n+1)\log s_n^2 \ge m.\right] \ge 1-\epsilon.$$

Hence, the MLE $\hat{\beta}_n$ is with high probability in the set of $\beta\epsilon(0,2n)$ which satisfy the inequality in the event in (3.8). In view of (3.3) and (3.5), we may restrict attention to the set of β's satisfying $0\le\beta\le2n$ and for some constants C_5, $C_6>0$, and m

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

It is easy to check that G_n is maximized at 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 $< c < 0$ 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 $\beta\epsilon(0,b]$, 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,

(3.10) $\quad \forall \epsilon > 0,\ \exists\ a > 0,\ b > a,\ N$ such that $\forall n \geq N$,

$$P [\ \hat{\beta}_n \text{ exists and } a \leq \hat{\beta}_n \leq b\ ] \geq 1 - \epsilon .$$

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

$$(3.11) \qquad \hat{\nu}(\beta) := \left[2 + \beta\right]^{-1} \left[Y(0) + Y(1) + \beta \int Y(t)\ dt \right],$$

and $\hat{\sigma}_n^2(\beta) = s_n^2 + O_P(n^{-1}) \xrightarrow{P} \sigma_0^2$, uniformly in $\beta \in [0,b]$. Hence $\ell_n(n^{1/2}\hat{\nu}_n(\beta), \hat{\sigma}_n^2(\beta), 1 - \beta/n) + [(n+1)/2] \log \sigma_0^2 + n/2$ converges in probability uniformly in $\beta \in (0,b]$ to $\ell(\hat{\nu}_n(\beta), \beta)$, where

$$\ell(\nu, \beta) := (1/2) \log \beta - B(\nu)\beta/\sigma_0^2 - A(\nu)\beta^2/\sigma_0^2,$$

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

$$A(\nu) := (1/2)\int_J [Y(t)-\nu]^2\, dt.$$

Now $\ell(\nu,\beta)$ 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 follows that $\hat{\beta}_n \xrightarrow{P} \hat{\beta}$, the MLE of β in the Ornstein-Uhlenbeck setup. The proof is complete.

#

Now consider the MCLE. First, define

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

Then the MCLE's are given by

(3.13) $$\tilde{r}_n = \frac{\sum [y(t)-\bar{y}_0][y(t+1)-\bar{y}_0]}{\sum [y(t)-\bar{y}_0]^2},$$

(3.14) $$\tilde{\mu}_n = \bar{y}_0 + \frac{y(n)-y(0)}{n(1-\tilde{r}_n)},$$

(3.15) $$\tilde{\sigma}_n^2 = \frac{1}{n} \sum [y(t+1) - \tilde{r}_n y(t) - (1-\tilde{r}_n)\tilde{\nu}_n]^2.$$

The corresponding MCLE's for the Ornstein-Uhlenbeck process are given in (2.6) through (2.8). The following theorem can be proved more simply than the previous one by simply using the explicit formulae for the estimators and the results in the Appendix.

Theorem 3.2. As $n\to\infty$,

$$\begin{bmatrix} n^{-1/2}\tilde{\mu}_n \\ \tilde{\sigma}_n^2 \\ n(1-\tilde{r}_n) \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \tilde{\nu} \\ \sigma_0^2 \\ \tilde{\beta} \end{bmatrix}.$$

#

Finally, we consider some "naive" estimators. Let

$$\bar{y}_{n1} = \frac{1}{n} \sum y(t+1), \tag{3.16}$$

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

$$r_n = \frac{\sum [y(t+1)-\bar{y}_1][y(t)-\bar{y}_0]}{\left[\left\{ \sum [y(t+1)-\bar{y}_1]^2 \right\} \left\{ \sum [y(t)-\bar{y}_0]^2 \right\} \right]^{1/2}} . \tag{3.18}$$

We refer to $\bar{y}_n$, s_n^2, and r_n as the naive estimators of μ_n, σ_0^2, and r_n, respectively.

Theorem 3.3. Let

$$\bar{\beta} = \frac{\frac{1}{2}[Y(1)-Y(0)][Y(1)+Y(0)+2\bar{Y}] - \int[Y(t)-\bar{Y}]dY(t)}{\int[Y(t)-\bar{Y}]^2 dt} . \tag{3.19}$$

Then as $n\to\infty$,

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

PROOF. We will assume as in the appendix 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 $\rightarrow$ will mean $\xrightarrow{P}$ for the remainder of the proof. Now it is clear from Lemma A.1 that

$$n^{-1/2}\bar{y}_n \rightarrow \bar{Y}, \text{ and } n^{-1/2}\bar{y}_{ni} \rightarrow \bar{Y}, \quad i=0,1. \tag{3.20}$$

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

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

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

Lemma A.1 also implies that $n^{-1}S_{ni}^2 \rightarrow S^2$ as $n\rightarrow\infty$. Some algebra will show that

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

Now

$$S_{n0}(S_{n1}-S_{n0}) = \frac{S_{n0}}{S_{n1}+S_{n0}}\, n^{-1}[y(n)-y(0)][y(n)+y(0)+\bar{y}_1+\bar{y}_0]$$

$$\rightarrow \frac{1}{2}\,[Y(1)-Y(0)][Y(1)+Y(0)+2\bar{Y}].$$

If one multiplies numerator and denominator in (3.21) by n^{-1} and uses this latter along with Lemma A.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 3 estimators of β_0 and ν_0 (the naive, MCLE, and MLE) and 4 estimators of σ_0^2 ($\tilde{\sigma}_n^2$ is the ordinary sample variance). All estimators except the MLE were computed directly from the formulae. 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 starting value, and convergence was quite fast, requiring on the 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. Firstly, 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" towards 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 ommitted. 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 equation (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). This inaccuaracy in $\tilde{\nu}$ does not seem to present a problem for the other parameter estimates $\tilde{\beta}$ 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.

TABLE 1. SUMMARY OF SIMULATION RESULTS
FOR ESTIMATORS OF β_0.

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

β_0	n	Estimator	Bias	Mean Squared Error
5	100	r_n	4.38 (.03)	42.27 (.53)
		$\tilde{\beta}_n$	4.37 (.03)	43.58 (.54)
		$\hat{\beta}_n$	4.48 (.03)	39.49 (.51)
5	500	r_n	4.55 (.03)	46.11 (.57)
		$\tilde{\beta}_n$	4.55 (.03)	47.89 (.59)
		$\hat{\beta}_n$	4.22 (.03)	42.87 (.55)
2	100	r_n	4.68 (.03)	40.21 (.46)
		$\tilde{\beta}_n$	4.68 (.03)	42.00 (.48)
		$\hat{\beta}_n$	4.27 (.03)	36.21 (.44)

TABLE 2. SUMMARY OF SIMULATION RESULTS
FOR ESTIMATORS OF ν_0.

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

β_0	n	Estimator	Mean Squared Error
		$\bar{y}_n$	.032 (.000)
5	100	$\tilde{\nu}_n$	.376 (.291)
		$\hat{\nu}_n$	.029 (.000)
		$\bar{y}_n$	.032 (.000)
5	500	$\tilde{\nu}_n$	.172 (.061)
		$\hat{\nu}_n$	.030 (.000)
		$\bar{y}_n$	.139 (.001)
2	100	$\tilde{\nu}_n$	286 (257)
		$\hat{\nu}_n$	.125 (.001)

TABLE 3. SUMMARY OF SIMULATION RESULTS
FOR ESTIMATORS OF σ_0^2.

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

β_0	n	Estimator	Mean Squared Error
5	100	all	.020 (.000)
5	500	all	.0040 (.0003)
2	100	all	.020 (.000)

APPENDIX.

In this appendix we give the proofs of two technical lemmas. There is a probability space carrying probabilistic replicas of $(\epsilon_n(1), \ldots, \epsilon_n(n))$ for each n and a Wiener process $\{W(t):0\le t\le 1\}$ such that the normalized partial sum process $W_n(t)$ satisfies

$$\sup_{0\le t\le 1} |W(t) - W_n(t)| \xrightarrow{P} 0,$$

where $\xrightarrow{P}$ denotes convergence in probability. See Theorem 13.8 of Breiman (1968). We assume that our sequence of experiments is defined on this probability space, and hereafter deal only with convergence in probability. The results then transfer back to the original probability space provided one replace $\xrightarrow{P}$ with $\xrightarrow{D}$. Let $Y(t)$ denote the Ornstein-Uhlenbeck process given by the stochastic differential equation in (1.6), and Y_n the normalized AR(1) process.

LEMMA A.1.

$$\text{(A.1)} \qquad \sup_{0\le t\le 1} |Y_n(t) - Y(t)| \xrightarrow{P} 0 ,$$

PROOF. It is convenient to introduce a Gaussian step function process $\tilde{Y}_n(t)$ by

$$\tilde{Y}_n((k+1)/n) = r_n\tilde{Y}_n(k/n) + \sigma\Delta W(k/n),$$

$$\tilde{Y}_n(t) = \tilde{Y}_n([nt]/n),\ \tilde{Y}_n(0) = (2\beta_0/[n(1-r_n^2)])Y(0),$$

where $\Delta f(k/n) = f((k+1)/n) - f(k/n)$. We have the representation

$$\text{(A.2)}\qquad \tilde{Y}_n(t) = r_n^{[nt]}Y_n(0) + \int_0^{[nt]/n} r_n^{([nt]-[ns]-1)}dW(s),$$

This follows from the usual inversion formula for an AR(1) process, e.g. (2.3.3) of Fuller (1976)). Utilizing the analogous formula (2.1) for the Ornstein-Uhlenbeck process we have

$$\begin{aligned}
\text{(A.3)}\qquad & |\tilde{Y}_n(t) - Y(t)| \le \\
& |e^{-\beta_0 t} - r_n^{[nt]}(2\beta_0/(n(1-r_n{}^2)))^{1/2}||Y(0)| \\
& + r_n^{-1}|r_n^{[nt]} - e^{-\beta_0 t}||\int_0^t r_n^{-[ns]}dW(s)| \\
& + e^{-\beta_0 t}|\int_0^t (r_n^{-[ns]} - e^{\beta_0 t})\ dW(s)| \\
& + r_n^{-1}|W(t) - W([nt]/n)| \\
& := T_{n1}(t) + T_{n2}(t) + T_{n3}(t) + T_{n4}(t), \text{ say.}
\end{aligned}$$

Letting $f_n(s) := r_n^{-[ns]} - e^{\beta_0 s}$, one can show via elementary inequalities that

(A.4) $$0 \le f_n(s) \le C/n$$

for some constant C (depending on β_0). From this and (5.1.5) of Arnold (1974), we have

$$E\left\{ \sup_{0 \le t \le 1} \left| \int_0^t f_n(s)\, dW(s) \right|^2 \right\} \le 4C^2/n^2$$

and hence that $\sup T_{n3}(t) \xrightarrow{P} 0$. The proofs that $\sup T_{ni}(t) \xrightarrow{P} 0$ for $i=1,2,4$ are even easier.

Now consider

(A.5) $$Y_n(k/n) - \tilde{Y}_n(k/n) = r_n^k[Y_n(0) - \tilde{Y}_n(0)] + \sum_{i=0}^{k-1} r_n^{(k-1-i)}[\Delta W_n(i/n) - \Delta W(i/n)].$$

Lindeberg's central limit theorem can be used to show $Y_n(0) \xrightarrow{D} Y(0)$, so we may assume that our probability space carries a version of $Y_n(0)$ such that $|Y_n(0) - Y(0)| \xrightarrow{P} 0$, and then the first term on the r.h.s. of (A.5) converges to 0 in probability, uniformly in k, $0 \le k \le n$. For the second term, apply partial summation to see that it is equal in absolute value to

$$\text{(A.6)}\quad |r_n^{-1}[W_n(k/n) - W(k/n)] - (r_n^{-1} - 1)\sum_{i=1}^{k} r_n^{k-i}[W_n(i/n) - W(i/n)]|$$

$$\le r_n^{-1}(2 - r_n^k) \sup_{0\le t\le 1} |W_n(t) - W(t)|.$$

Since $r_n \to 1$, $r_n^{-1}(2-r_n^k)$ is bounded uniformly in k and n, so it follows that $\sup|Y_n(t) - \tilde{Y}_n(t)| \overset{P}{\to} 0$.

#

LEMMA A.2.

$$\text{(A.7)}\qquad \sum_{k=0}^{n-1} Y_n(k/n)\Delta W_n(k/n) \overset{P}{\to} \int_0^1 Y(t)dW(t).$$

PROOF. We have

$$\sum_{k=0}^{n-1} Y_n(k/n)\Delta W_n(k/n) = r_n^{-1}Y_n(1)W_n(1)$$

$$+ (r_n^{-1} - 1)\sum Y_n(k/n)W_n(k/n)$$

$$- (1/2)\sigma_0 r_n^{-1}W_n^2(1) - (1/2)\sigma_0 r_n^{-1}\sum [\Delta W_n(k/n)]^2$$

$$\overset{P}{\to} Y(1)W(1) + \beta\int Y(t)W(t)\,dt - (1/2)\sigma_0 W^2(1) - (1/2)\sigma_0.$$

The first equality is easily checked with some algebra. The convergence of the first three terms on the l.h.s. of the $\xrightarrow{P}$ to the first three terms on the r.h.s. of the $\xrightarrow{P}$ is immediate by Lemma A.1, and the fourth by the weak law of large numbers. By Ito's formula and the stochastic differential equation for Y we have

$$d\left[Y(t)W(t)\right] = Y(t)dW(t) - \beta Y(t)W(t)dt + \sigma_0 W(t)dW(t) + \sigma_0 dt$$

and so

$$\int Y(t)dW(t) = Y(1)W(1) + \beta\int Y(t)W(t)\,dt - (1/2)\sigma_0 W^2(1) - \sigma_0/2$$

where we used the fact

$$\int_0^1 W(t)dW(t) = (1/2)\left[W^2(1) - 1\right],$$

see e.g. Arnold (1974), page 76. This completes the proof.

#

REFERENCES.

Arnold, L. (1974) Stochastic differential equations: theory and applications, R. Oldenbourg, Munich.

Bobkoski, M. J. (1983) Hypothesis testing in nonstationary time series, unpublished Ph. D. thesis, University of Wisconsin, Madison.

Breiman, L. (1968) Probability, Addison-Wesley, Reading, Massachusetts.

Chan, N. H. and Wei, C. Z. (1985) Asymptotic inference for nearly nonstationary AR(1) processes, Technical report, Dept. of mathematics, University of Maryland, College Park, submitted to Ann. Statist.

Cumberland, W. G. and Sykes, Z. M. (1982) Weak convergence of an autoregressive process used in modelling population growth, J. Appl. Prob., 79, 355-367.

Feigin, P. D. (1976) Maximum likelihood estimation for continuous time stochastic processes, Adv. Appl. Prob., 8, 712-736.

Feigin, P. D. (1979) Some comments concerning a curious singularity, J. Appl. Prob., 16, 440-444.

Fuller, W. (1976) Introduction to stastistical time series, Wiley, New York.

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. of Statistics, Carnegie-Mellon University.

END
9-87
DTIC