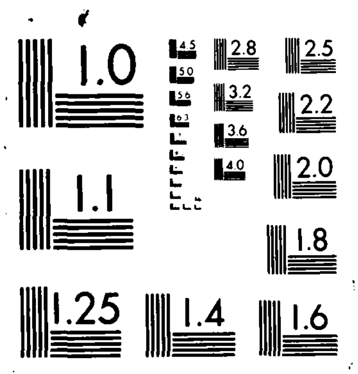
AD-A181 184 M-ESTIMATION FOR NEARLY NON-STATIONARY AUTOREGRESSIVE TIME SERIES(U) WASHINGTON UNIV SEATTLE DEPT OF STATISTICS D D COX ET AL MAR 87 TR-95 1/1

UNCLASSIFIED N00014-84-C-0169 F/G 12/3 NL

END


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

(12)

M-ESTIMATION FOR NEARLY NON-STATIONARY AUTOREGRESSIVE TIME SERIES

by

Dennis D. Cox
Isabel Llatas

TECHNICAL REPORT No. 95
March 1987

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

M-ESTIMATION FOR NEARLY NON-STATIONARY AUTOREGRESSIVE TIME SERIES

by

Dennis D. Cox
Isabel Llatas

TECHNICAL REPORT No. 95

March 1987

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

M-estimation for Nearly Non-Stationary Autoregressive Time Series

*Dennis D. Cox**

Department of Statistics
University of Illinois
Urbana, IL 61801

and

Isabel Llatas

Department of Statistics
University of Wisconsin
Madison, WI 53706

ABSTRACT

The nearly nonstationary first order autoregression is a sequence of processes where the autoregressive coefficient tends to 1 as $n \to \infty$. M-estimates of the autoregressive coefficient are considered. The process is allowed to be nongaussian, but a $2+\delta$ moment condition is assumed. The limiting distribution is not the usual normal limit but is characterized as a ratio of two stochastic integrals. The asymptotically most efficient M-estimate is not given by maximum likelihood. However, it is shown that the loss of efficiency in using maximum likelihood is no worse than about 20%, whereas the usual least squares estimator can have arbitrarily low efficiency.

* This research was partially supported by the Office of Naval Research under contract number N00014-84-C-0169, and the National Science Foundation under grant number DMS-820-2560.

Key Words and Phrases: M-estimation, time series, autoregressive, non-stationary

AMS-MOS Subject Classification (1980): Primary 62M10; Secondary 62E20, 62F12.

1. Introduction

The aim of this work is to study asymptotic properties of M-estimators of the autoregressive parameter ϕ of a nearly non-stationary first order autoregressive process, and to obtain efficient M-estimators of ϕ. We consider the sequence $\{y_n(k): 0 \le k \le n\}_{n=1}^{\infty}$ of first order autoregressive AR(1) processes

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

where we assume $\{\varepsilon(k)\}_{k=-\infty}^{\infty}$ is a sequence of *iid* random variables with mean zero and finite $(2+\delta)$-moment, for some positive δ, and ϕ_n is allowed to vary with n. Specifically, we will assume

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

for some $\beta > 0$, so that y_n tends to look like a non-stationary random walk for large n. Also we will assume that we have some knowledge on the starting value $y_n(0)$, either by considering it as a constant or by assuming is a random variable with known distribution. In principle we are interested in the asymptotic behavior of estimators of the form:

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

for some function ρ. Here, $\arg\min_{\phi}$ denotes the value of ϕ where a minimum is achieved. For example, taking $\rho(u) = u^2$ equation (1.3) gives the least squares estimator, ***LSE***, of ϕ.

It is known that the ***LSE*** of ϕ, for fixed $\phi_n \equiv \phi$ with $|\phi| < 1$ is asymptotically normal $N(0, 1-\phi^2)$, but when $\phi = 1$ the ***LSE*** is $O_p(n^{-1})$ 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 $\phi_n = 1$ (*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 convergence in distribution. Rao (1978), Dickey and Fuller (1979), and Evans and Savin (1981) have obtained representations of this limiting distribution. For the nearly non-stationary (*NNS*) model of equation (1.1), Cumberland and Sykes (1982) found that the normalized processes $n^{-1/2} y_n([nt])$ converges in distribution to an Ornstein-Uhlenbeck process defined by the Itô's 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

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

where ϕ_n is given by (1.2). Chan and Wei (1985) obtained similar results for the *NNS* model and found that when the parameter β goes to infinite 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 non-stationary 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 $|\phi| < 1$; specifically it follows from the work of these authors that under standard regularity conditions (e.g (2.A) and (2.B) below)

$$n^{1/2}(\hat{\phi}_n - \phi) \Rightarrow N(0, (1-\phi^2)v_p)$$

where

$$v_p = \frac{E\,\psi^2(\varepsilon(1))}{[E\,\dot{\psi}(\varepsilon(1))]^2}$$

$$\psi(u) = \frac{d\rho(u)}{du}; \quad \dot{\psi}(u) = \frac{d\psi(u)}{du}.$$

A simple variational argument will show the most "efficient" M-estimator (the one minimizing v_p) is obtained from $\rho = -\log(f)$ where f is the density of the ε's, *i.e.* 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 non-stationary 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 function which works "best" is a linear combination of the ***LSE*** and ***MLE*** criterion functions.

The asymptotic results that we present in this work deal with convergence in distribution 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 $[0,T]$ such that they are right continuous and the left limits exists, 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 on the Stroock and Varadhan characterization of the solution of a *SDE* as the solution of an associated *martingale problem*. For a detailed account see e.g 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.

Theorem 1 : *(7.4.1 Ethier and Kurtz (1986))*

Let $a = ((a_{ij}))$ be a continuous, symmetric, nonnegative definite $d\times d$ matrix-valued function on $\mathbb{R}^d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ be continuous. Let A be the second order differential operator on $C_c^\infty(\mathbb{R}^d)$ given by

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

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

For $n = 1, 2, \cdots$, 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 $D_{\mathbb{R}}[0,\infty)$ and $A_n(t) - A_n(s)$ is nonnegative definite for $t > s \geq 0$. Set $F_t^n = \sigma(\mathbf{X}_n(s), \mathbf{B}_n(s), \mathbf{A}_n(s) : s \leq t)$.

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

$$\mathbf{M}_n \equiv \mathbf{X}_n - \mathbf{B}_n \tag{1.6}$$

and

$$M_n^i M_n^j - A_n^{ij} \qquad i,j=1,2,\dots,d \tag{1.7}$$

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

$$\lim_{n\to\infty} E\left[\sup_{t\le\min(T,\tau_n^r)} |\mathbf{X}_n(t)-\mathbf{X}_n(t^-)|^2\right]=0 \tag{1.8}$$

$$\lim_{n\to\infty} E\left[\sup_{t\le\min(T,\tau_n^r)} |\mathbf{B}_n(t)-\mathbf{B}_n(t^-)|^2\right]=0 \tag{1.9}$$

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

$$\sup_{t\le\min(T,\tau_n^r)} |\mathbf{B}_n^i(t)-\int_0^t b_i(\mathbf{X}_n(s))ds| \xrightarrow{P} 0 \tag{1.11}$$

and

$$\sup_{t\le\min(T,\tau_n^r)} | A_n^{ij}(t)-\int_0^t a_{ij}(\mathbf{X}_n(s))ds | \xrightarrow{P} 0 \tag{1.12}$$

Suppose that $\mathbf{X}_n(0)$ *converges weakly to a random variable with distribution* ν, *then* $\{\mathbf{X}_n\}$ *converges in distribution to the solution of the martingale problem for* (A,ν) □.

Remark: By the representation mentioned before the limiting process corresponds to the diffusion with infinitesimal generator given by $\boldsymbol{A}$.

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 find the form of an optimal M-estimator.

Next, we compare the *MSE* ot the *LSE* and conditional *MLE* versus the asymptotic *MSE* of the optimal M-estimator. In Section 4 we show some results needed for the proof of the asymptotic theorem and give the proof.

2. Statement of the Main Theorem

Assume ρ in (1.3) is differentiable and set $\psi = \dot{\rho}$ 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

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

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

(2.B)

The $(2+\delta)$ order moments of $\varepsilon(1)$, $\psi(\varepsilon(1))$ and $\dot{\psi}(\varepsilon(1))$ are finite for some positive δ.

(2.C)

$E\,\psi(\varepsilon(1)) = 0$ and $E\,\dot{\psi}(\varepsilon(1)) = 1$. The assumption $E\,\dot{\psi}(\varepsilon(1)) = 1$ involves no loss of generality provided $E\,\dot{\psi}(\varepsilon(1)) \neq 0$.

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

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

Hence if we let

$$t = y_n(k+1) - \hat{\phi}_n y_n(k) \quad \text{and}$$
$$t_0 = y_n(k+1) - \phi_n y_n(k) = \varepsilon(k+1) \tag{2.3}$$

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

$$\sum_{k=1}^{n-1}\Big[y_n(k)\,\psi(\varepsilon(k+1))\Big] - (\hat{\phi}_n - \phi_n)\sum_{k=1}^{n-1} y_n^2(k)$$
$$- (\hat{\phi}_n - \phi_n)\sum_{k=1}^{n-1}\Big[y_n^2(k)(\dot{\psi}(\varepsilon(k+1)) - 1)\Big] \tag{2.4}$$
$$+ (\hat{\phi}_n - \phi_n)^2\, C \sum_{k=1}^{n-1}\Big[y_n^3(k)\alpha(k)\Big] = 0$$

The main result in this paper is summarized by the following theorem.

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

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

$$(\hat{\phi}_n - \phi_n) = \mathbf{O}_p(n^{-1}) \tag{2.5}$$

(b) *For such a sequence*

$$n(\hat{\phi}_n - \phi_n) \Rightarrow \frac{\int_0^1 Y(s)\,d\,W_2(s)}{\int_0^1 Y^2(s)\,ds} \tag{2.6}$$

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

$$dY(t) = -\beta Y(t)\,dt + dW_1(t)$$

$$Y(0) \overset{D}{=} N(0, \frac{\sigma^2}{2\beta}), \tag{2.7}$$

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

$$E\,[W_1^2(t)] = t\,E\,[\varepsilon^2(1)],$$

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

$$E\,[W_1(t)W_2(t)] = t\,E\,[\varepsilon(1)\psi(\varepsilon(1))] \quad \square$$

Remark: Implicitly stated in the assumed initial condition for the sequence of AR(1) processes is the assumption that for each n, the process is stationary. Thus is not surprising that the initial condition for the Ornstein-Uhlenbeck process of equation (2.7) is the one needed to insure the stationarity of such a process (Arnold (1974), page 135).

The weak limit in (2.6) is suggested by neglecting the last two terms of the *RHS* of (2.4), so that

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

Define

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

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

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

and

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

(with the usual convention that summation equals zero when the upper limit is smaller 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 (*i.e.* $\Delta m(k) = m(k+1) - m(k)$) and $\Delta t = n^{-1}$. Then (2.8) can be written as

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

Let $\mathbf{W}(t) = \left[W_1(t), W_2(t), W_3(t) \right]'$ be a three-dimensional Brownian motion such that variance-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 (see Cumberland and Sykes (1982)) it is natural to think of the summations in (2.11) as the Riemann-Stieltjes sums for the integrals in (2.6), and we will show in Theorem 3 below, among other things, that the two summations in (2.11) jointly converge to the corresponding integrals in (2.6).

3. Optimality

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

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

Surprisingly, we have found that this criterion leads to the finding that the optimal ψ function is a linear combination of $\eta_1(x) = x$ and $\eta_2(x) = -I_f^{-1}\dot{f}(x)/f(x)$, where f is the probability density function of the innovations (assuming it exists) and I_f is the Fisher information of the location parameter problem for the common distribution of the noise. Note that η_1 corresponds to the least squares score function while η_2 is proportional to the usual score function of the ***MLE***. The 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 immediately suggest a two stage procedure that may be useful. The first stage is to estimate ϕ_n by say the ***MLE*** ,$\hat{\phi}_{n,MLE}$, and hence β by $\beta_{n,MLE} = n\,(1-\hat{\phi}_{n,MLE})$. One can then find the optimal ψ function for the estimate $\hat{\beta}$ and the second stage consists of finding the solution of the M-estimation equation for this ψ.

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

$$Q(\psi) = \frac{L_1 - L_2}{\sigma^2} Cov^2\Big[\varepsilon(1), \psi(\varepsilon(1))\Big] + L_2 Var\Big[\psi(\varepsilon(1))\Big] \tag{3.2}$$

where

$$L_1 = E\left[\int_0^1 Y\, d\, W_1 \Big/ \int_0^1 Y^2 ds\right]^2 \quad \text{and} \quad L_2 = E\left[\int_0^1 Y^2 ds\right]^{-1} \tag{3.3}$$

Hence 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

$$Q(\psi) + \lambda_1 E\,(\psi(\varepsilon(1))) + \lambda_2\Big[E\,(\dot{\psi}(\varepsilon(1))) - 1\Big] =$$

$$\frac{L_1 - L_2}{\sigma^2}\Big[\int x\psi(x) f(x) dx\Big]^2 + L_2 \int \psi^2(x) f(x) dx$$

$$+ \lambda_1 \int \psi(x) f(x) dx + \lambda_2\Big[\int \dot{\psi}(x) f(x) dx - 1\Big]$$

equal to zero, 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

$$\Big[2\sigma^{-2}(L_1 - L_2)\int y\psi(y) f(y) dy\Big]\, x f(x) + 2L_2\psi(x) f(x) + \lambda_1 f(x) - \lambda_2 \dot{f}(x) = 0$$

whence

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

where

$$\kappa = \frac{L_2 - L_1}{\sigma^2 L_2} Cov(\psi, \varepsilon)$$

where ψ and ε are shorthand for $\psi(\varepsilon(1))$ and $\varepsilon(1)$ respectively. It is easy to see that both

$E(\varepsilon)=0$ and the constraint $E(\psi)=0$ imply $\lambda_1=0$, under the usual regularity conditions on f that allow the interchange of the integral and derivative. Thus the optimal ψ is a linear combination of the least squares and maximum likelihood criterion functions. Also the constraint $E(\dot{\psi})=1$ implies

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

Substitution of the value 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 $Cov(\psi,\varepsilon)$ for ψ in equation (3.5) gives that

$$\kappa=\frac{L_2-L_1}{L_2-L_1(1-\sigma^2 I_f)}.$$

Plugging this value in the definition of ψ gives

$$\psi(x)=\frac{(L_2-L_1)x-\sigma^2 L_1\dfrac{\dot{f}(x)}{f(x)}}{L_2-L_1(1-\sigma^2 I_f)} \tag{3.6}$$

One should note that ψ depends on β through L_1 and L_2. Further, 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. Now, it is easy to check that if L_1' and L_2' are the corresponding moments when the variance of the Brownian motion driving Y is equal to one, then $L_1=L_1'/\sigma^2$ and $L_2=L_2'/\sigma^2$, so it is enough to obtain L_1' and L_2'. Following the procedure in Williams (1942) one can obtain the moments of the ratio of powers of the numerator (to be denoted by N) and denominator (to be denoted by D) of the ratio on the ***RHS*** of equation (1.5) from the joint moment generating function of N and D. Thus, for example, if $\Lambda(s_0,s)=E\,[\exp\{-s_0D - s\,N\}]$ then

$$\int_0^\infty \Lambda(s_0,0)\,ds_0 = E\left[\int_0^\infty e^{-s_0D}\,ds_0\right] = E\left[\frac{1}{D}\right] \tag{3.7}$$

and

$$\int_0^\infty\int_t^\infty \frac{\partial^2}{\partial s^2}\Lambda(s_0,s)\Big|_{s=0} d\,s_0dt = \int_0^\infty\int_t^\infty E\left[N^2e^{-s_0D}\right] ds_0dt = E\left[\frac{N}{D}\right]^2 \tag{3.8}$$

These formal manipulations will be valid as long as the interchange of differentiation and integration are valid. From equation (4.20), Bobkoski (1983) we have that the joint ***MGF*** of N and D, when $Y(0)=0$ is given by

$$\begin{aligned}\Lambda(s_0,s) &= E\,(\exp(-s_0D - s\,N))\\ &= \exp\{\frac{\beta+s}{2}\}\left[\cosh(z) + (\beta+s)\,\mathrm{shnc}(z)\right]^{-1/2}\end{aligned} \tag{3.9}$$

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*** when the initial distribution is known are available (Llatas (1987)). The choice of $Y(0)=0$ is motivated by the convenience of checking the results obtained by numerical integration with both simulations and the approximated moments obtained by numerical integration of the explicit form of the asymptotic limiting density function obtained by Bobkoski in this special case. The fact that Λ in (3.9) is differentiable and that the terms of 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 theorem and Fubini-Tonelli theorem. In Table I we exhibit some of the values of L_1' and L_2' calculated using the integration subroutine DQAGI in QUADPACK.

Table I: Values of L_1' and L_2' obtained by numerical integration

Values of L_1' and L_2'		
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

The values obtained present a very curious feature: they fall in what seems two parallel straight lines with slope near 2 and intercept equal to 13.33 for L_1' and 5.37 for L_2' (see Figure 1). A regression line was fitted to the values in Table I assuming the two lines are indeed parallel and the regression equations are given by:

$$L_1' = 13.33 + 1.98\beta; \quad L_2' = 5.37 + 1.98\beta$$

The residuals from this regression are shown in Figure 2. Figure 2 indicates the true values would not fall in a straight line. Note the different behavior when $\beta < 1$. However over the range considered the linear approximation might be satisfactory and gives us a quick way to estimate the value of L_1' and L_2' without performing the numerical integration. This may be advantageous when considering the two step estimation procedure mentioned before. To check the values obtained by the numerical integration we performed a small Monte Carlo experiment for $\beta = 2, 10, 20$ by evaluation of the corresponding sample values of 10,000 series of sizes $n = 100, 500, 1000$. We also evaluated the second moment of the asymptotic distribution from the representation of the density of the limiting *LSE* error in Bobkoski (1983). The results are shown in Table II. The latter values are slightly smaller than the one calculated from (3.8).

Table II. Comparison of results for L_1' and L_2'

		Numerical Integration		Monte Carlo Experiment		
		MGF	Density	$n=100$	$n=500$	$n=1000$
$\beta=2$	L_1'	17.2663	17.2655	16.2126	16.9297	*
				(0.4349)	(0.4750)	*
	L_2'	9.3093	*	9.3327	9.3805	*
				(0.0735)	(0.0725)	*
$\beta=10$	L_1'	33.1105	33.1095	29.5127	31.1534	33.4475
				(0.6165)	(0.6780)	(0.7743)
	L_2'	25.1135	*	23.8719	24.7268	24.9647
				(0.1026)	(0.1056)	(0.1091)
$\beta=20$	L_1'	53.0636	53.0627	42.6071	51.5[illegible]	51.3961
				(0.7692)	(1[illegible])	(0.9723)
	L_2'	45.0643	*	40.3335	44.[illegible]96	44.4285
				(0.1204)	(0.1373)	(0.1375)

Note: Values in parenthesis are estimated standard errors for the quantity above.

The values shown for $\beta=10,20$ are obtained by integration on $[-70,\beta]$. For $\beta=2$ the range of integration is $[-35,5.70]$. As for the Monte Carlo trials, the estimated values lie within two standard deviations of the values obtained by numerical integration except when $\beta=20$, where the bias has not been overcome by the increment of the size of the series. In any case the values are close enough to support the numerical integration results. Less bias and smaller estimated standard deviation from the simulations would be ideal but unfeasible since in order to lower the value of both the bias and variance it may need more computer time than what is convenient or even allowed on the facilities used.

Now we are in the position to calculate values of Q for the score functions η_1, η_2 and ψ. By equation (3.2) and the observation about the relation between L_i and L_i' we have:

$$Q(\eta_1) = \sigma^2 L_1 = L_1'$$

$$Q(\eta_2) = (\sigma^4 I_f^2)^{-1} [L_1' - L_2'(1 - \sigma^2 I_f)]$$

$$Q(\psi) = \frac{L_1' L_2'}{L_2' - L_1'(1 - \sigma^2 I_f)}$$

Note that if I_f' is the information when $\sigma^2 = 1$ we have that $\sigma^2 I_f = I'_f$, therefore the asymptotic mean squared error for the score functions considered here does not depend on the variance of the shocks. Moreover, it depends on the probability density function of the shocks only through the information I'_f. Thus we will set $\sigma = 1$ and in this case we have $I'_f \geq 1$ (Rustagi (1976)). Consequently

$$\text{(a)} \quad \frac{Q(\eta_1)}{Q(\psi)} = \frac{L_2' - L_1'(1 - I'_f)}{L_2'} \geq 1 \tag{3.10}$$

and a minimum is obtained when $I'_f = 1$.

$$\text{(b)} \quad \frac{Q(\eta_2)}{Q(\psi)} = \frac{L_1' L_2' I'^2_f + (L_1' - L_2')^2 (I'_f - 1)}{L_1' L_2' I'^2_f} \geq 1 \tag{3.11}$$

and a maximum is obtained when $I'_f = 2$.

In Figure 3 we exhibit the ratio $Q(\eta_1)/Q(\psi)$ for the ***LSE*** for values of $I'_f = (\pi/3)^2$, 1.50, and 2.00. In Figure 4 the ratio $Q(\eta_2)/Q(\psi)$ for the ***MLE*** is shown for the same values of I'_f. Note that $I'_f = (\pi/3)^2$ corresponds to a logistic distribution with mean zero and variance 1. From these figures one can see that the ***LSE*** can be very "inefficient" while the ***MLE*** cannot be worse than 20% "inefficient" in the ***MSE*** sense.

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 (2.4) as an application of Theorem 1.

Theorem 3 : *Consider the model 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*

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

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 3-dimensional Brownian Motion defined below equation (2.11) *and Y is the Ornstein-Uhlenbeck process defined by equation* (2.7) *with initial condition having the stationary distribution.*

Proof: First of all note that we can represent $\mathbf{W}$ by

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

where $\mathbf{b}(t)$ is a 3-dimensional standard Brownian Motion with covariance $(t\,I)$ and $\Gamma = (\gamma_{ij})$ is the Cholesky factor for Σ, *i.e.* Γ is a 3×3-lower triangular matrix such that

$\Gamma\Gamma' = \Sigma$. Now the process $\mathbf{X}(t)$ satisfies the Stochastic Differential Equation:

$$d\mathbf{X}(t) = \begin{bmatrix} -\beta X_1(t) \\ 0 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 1 & 0 & 0 \\ 0 & X_1(t) & 0 \\ 0 & 0 & X_1^2(t) \end{bmatrix} d\mathbf{W}(t)$$

$$\text{(set)} = b(\mathbf{X}(t))dt + G(\mathbf{X}(t))\,d\mathbf{W}(t) \tag{4.4}$$

$$= b(\mathbf{X}(t))dt + G(\mathbf{X}(t))\,\Gamma\, d\mathbf{b}(t)$$

with initial condition $\mathbf{X}(0) = (Y(0), 0, 0)'$. The last equality in (4.4) follows by equation (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 Arnold (1974), equation (4.4) has exactly one continuous solution. Moreover the process $\mathbf{X}(t)$ is a 3-dimensional diffusion process on $[0,1]$ with drift vector $b(\mathbf{x})$ and diffusion matrix $a(\mathbf{x}) = G(\mathbf{x})\Gamma\Gamma' G'(\mathbf{x}) = G(\mathbf{x})\Sigma G'(\mathbf{x})$ (see 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, *i.e.*

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

with initial measure equal to $Law(\mathbf{X}(0))$, which should 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

3-dimensional degenerate normal $N(0,\Theta\sigma^2/2\beta)$ where Θ_{ij} equals zero 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 a $N(0,\sigma^2/2\beta)$ by an easy application of the Linderberg-Feller Central Limit Theorem to the triangular array defined by

$$T_{n,k_n}=\phi_n^k\varepsilon(-k)\quad 0\le k\le n^2$$

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

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

with $\mathbf{W}_n$ defined in equation (2.10) so it is natural to thing that $\mathbf{X}_n$ will approximate the continuous process $\mathbf{X}$. We proceed to prove this by finding 3-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 equation (4.7) it follows that

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

where

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

Since $E\,[\xi_n(k)/G_{k-1}]=0$ the predictable compensator of $\mathbf{X}_n$ is given by

$$\mathbf{B}_n(t) = \sum_{k=0}^{[nt]-1} \left\{ E\,[\Delta \mathbf{X}_n(k/n)\,/\,G_k] \right\}$$

$$= \left[-\beta \sum_{k=0}^{[nt]-1} Y_n(k/n)\Delta t\ ,0\,,0 \right] \tag{4.8}$$

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 find $\mathbf{A}_n$, the compensator of $\mathbf{M}_n(k/n)\mathbf{M}_n'(k/n)$ as

$$\mathbf{A}_n(t) = \sum_{k=1}^{[nt]} \left\{ E\,[\mathbf{M}_n(k/n)\mathbf{M}'_n(k/n) - \mathbf{M}_n((k-1)/n)\mathbf{M}'_n((k-1)/n)\,/\,G_{k-1}] \right\}$$

$$= n^{-1}\sum_{k=1}^{[nt]} E\,[\xi_n(k)\xi'_n(k)/G_{k-1}] \tag{4.10}$$

It follows from the last equality of (4.11) that the increments $\mathbf{A}_n(t) - \mathbf{A}_n(s)$, $t > s$ of the process so defined are non-negative definite.

What is left now is to verify the "continuity" conditions (1.8) to (1.10) and the "approximation" conditions (1.11) and (1.12) of Theorem 1. We start by the approximation conditions. For condition (1.11) we have just to show

$$\sup_{0 \le t \le 1} |B_{1,n}(t) - \int_0^t b_1(X_n(s))ds| \xrightarrow{P} 0$$

but the absolute value equals:

$$\beta \left| \int_0^t n^{-1/2} y_n([ns])ds - \sum_{k=0}^{[nt]-1} n^{-1/2} y_n(k)\Delta t \right| = \beta\,(t-[nt]/n)\,|Y_n(t)|$$

$$\le \frac{\beta}{n}\,|Y_n(t)| \le \frac{\beta}{n}\,\|Y_n\|_\infty \tag{4.11}$$

Since $\|Y_n\|_\infty$ is bounded in probability (Bobkosky (1983), page 25) the last quantity goes to zero as n goes to infinity. Condition (1.12) will also follow by the same type of argument and the boundeness 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 (1.10) 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

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

which follows again by (4.12).

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

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

$$\lim_{n\to\infty} E\left[n^{-2}\sup_{k\le n\tau_n^r}\left[y_n(k-1)\psi(\varepsilon(k))\right]^2\right] = 0 \tag{4.14}$$

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

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

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* $\{\eta(k)\}_{k=1}^n$ *be a sequence of iid random variables with finite* $(1+\delta)$-*moment then*

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

Proof : Let F be the cdf of $\eta(1)$. Define $x(u)=\inf\{x : F(x)\le u\}$. By the so called probability integral transformation $u=F(x)$

$$E\left[\max_{0\le k\le n}\eta(k)\right] = \int_0^1 n\,x(u)\,u^{n-1}du \tag{4.16}$$

To show (4.15) we use the Holder's inequality

$$\left|\left|\int f\,g\right|\right| \le \left[\int |f|^p\right]^{\frac{1}{p}}\left[\int |g|^q\right]^{\frac{1}{q}}$$

with $f = x(u)$, $g = n\, u^{n-1}$, $p = 1+\delta$, and $q = (1+\delta)/\delta$ to obtain

$$n^{-1} E\,[\max_{0\le k\le n} \eta(k)\,] \le n^{-1}\Big[E\left[\,|\eta(1)|^{1+\delta}\right]\Big]^{\frac{1}{1+\delta}}\, n \left[\frac{\delta}{(1+\delta)(n-1)+\delta}\right]^{\frac{\delta}{1+\delta}}$$

$$= \mathbf{O}(n^{-\frac{\delta}{1+\delta}}) \qquad \square$$

The next result proves the weak convergence of the terms on the Taylor expansion in equation (2.4) 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 4-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

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

Proof : Consider the transformation $g : C_{\mathbb{R}^3}[0,1] \rightarrow \mathbb{R}^4$ such that

$$g(\mathbf{x}) = g\left[(x_1(t), x_2(t), x_3(t))'\right]$$

$$= \left[\int_0^1 x_1^2(s)ds, \int_0^1 |x_1^3(s)|ds, x_2(1), x_3(1)\right]'$$

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. Hence $\mathbf{Z}_n$ converges weakly to $\mathbf{Z}$ by the continuity principle (Theorem 5.1 of Billingsley(1968) page 30). □

Using the asymptotic results we proceed to prove our main Theorem in the same fashion Cramer showed the asymptotic properties of the maximum likelihood estimator (Cramer (1946), chapter 33).

Proof of Theorem 2 : By means of equation (2.4) we can write $\Psi(\zeta)=0$, after multiplication by n^{-2}, in the form

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

where

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

and

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

which is bounded by 1. Theorem 3 implies that $T_{0,n}$ is $\mathbf{O}_p(n^{-1})$ and $T_{2,n}$ is $\mathbf{O}_p(n^{-1/2})$, while Lemma 2 implies that $T_{3,n}$ is $\mathbf{O}_p(n^{1/2})$ and $T_{1,n}$ converges weakly to a random variable, which is positive with probability 1 (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 $n \geq N$ there exist finite positive constants M_0, M_1, M_2, M_3 such that

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

thus with at least probability $1-\gamma$

$$n^{-2}\Psi(\zeta) > -M_0 n^{-1} - M_1(\zeta - \phi_n) - M_2 n^{-1/2} |(\zeta - \phi_n)| - M_3 n^{1/2}(\zeta - \phi_n)^2 \tag{4.19}$$

and

$$n^{-2}\Psi(\zeta) < M_0 n^{-1} - M_1(\zeta - \phi_n) + M_2 n^{-1/2} |(\zeta - \phi_n)| + M_3 n^{1/2}(\zeta - \phi_n)^2 \tag{4.20}$$

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

$$\zeta_1 = \phi_n - n^{-1}\left[\frac{2M_0}{M_1}\right]$$

$$\zeta_2 = \phi_n + n^{-1}\left[\frac{2M_0}{M_1}\right]$$

hence equation (4.19) gives:

$$n^{-2}\Psi(\zeta_1) > -M_0 n^{-1} + 2M_0 n^{-1} - n^{-3/2}\left[\frac{2M_0}{M_1}M_2 + \left[\frac{2M_0}{M_1}\right]^2 M_3\right]$$

$$> \left[M_0 - \frac{M_0}{2}\right]n^{-1} > 0$$

while (4.20) gives:

$$n^{-2}\Psi(\zeta_2) < M_0 n^{-1} - 2M_0 n^{-1} + n^{-3/2}\left[\frac{2M_0}{M_1}M_2 + \left[\frac{2M_0}{M_1}\right]^2 M_3\right]$$

$$< \left[-M_0 + \frac{M_0}{2}\right]n^{-1} < 0$$

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

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

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

For part (b) we just have to write

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

It follows from the preceding discussion that $T_{2,n} - (\hat{\phi}_n - \phi_n)T_{3,n}$ converges in probability

to zero 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 equation (4.21) to the random variable in (2.7) is guaranteed by a straightforward application of Slutsky's theorem and Theorem 5.1 in Billingsley (1968). □

5. Appendix

Let $\mathbf{W}(t)$ the 3-dimensional Brownian motion defined in Section 2. As noted before 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 3-dimensional standard Brownian Motion with covariance $(t\, \mathrm{I})$ and $\Gamma = (\gamma_{ij})$ is the Cholesky factor for Σ, *i.e.* 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

$$\int_0^1 Y(s)dW_2(s) = \gamma_{21}\int_0^1 Y(s)db_1(s) + \gamma_{22}\int_0^1 Y(s)db_2(s) \tag{5.1}$$

Note that $W_1 = \gamma_{11} b_1$ and consequently the process Y defined by the *SDE* (2.7) is independent of b_2 and b_3.

From (5.1) we have

$$Q(\psi) = (\gamma_{21})^2 E\left[\frac{\int_0^1 Y(s)db_1(s)}{\int_0^1 Y^2(s)ds}\right]^2 + (\gamma_{22})^2 E\left[\frac{\int_0^1 Y(s)db_2(s)}{\int_0^1 Y^2(s)ds}\right]^2$$

$$+ 2\gamma_{21}\gamma_{22} E\left[\frac{\left[\int_0^1 Y(s)db_1(s)\right]\left[\int_0^1 Y(s)db_2(s)\right]}{\left[\int_0^1 Y^2(s)ds\right]^2}\right] \tag{5.2}$$

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 $F_t^{(1)}$-measurable random function $h(t)$ we have:

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

and

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

This can be proven by first looking at $F_1^{(1)}$-measurable 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 $F_1^{(1)}$-measurable one obtains that $E(\int_0^1 Y(s)db_2(s) \mid F_1^{(1)}) = 0$. Thus the expectation of the cross product in (5.2) vanishes since $\int_0^1 Y(s)db_1(s)$ and $\int_0^1 Y^2(s)ds$ are $F_1^{(1)}$-measurable. 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

$$Q(\psi) = \gamma_{21}^2 E\left[\frac{\int_0^1 Y(s)db_1(s)}{\int_0^1 Y^2(s)ds}\right]^2 + \gamma_{22}^2 E\left[\int_0^1 Y^2(s)ds\right]^{-1} \tag{5.3}$$

$$= \gamma_{21}^2 L_1 + \gamma_{22}^2 L_2$$

Plugging in the values of γ_{21} and γ_{22} into (5.3) gives expression (3.2)

Figure 1: Values of L_1' and L_2'

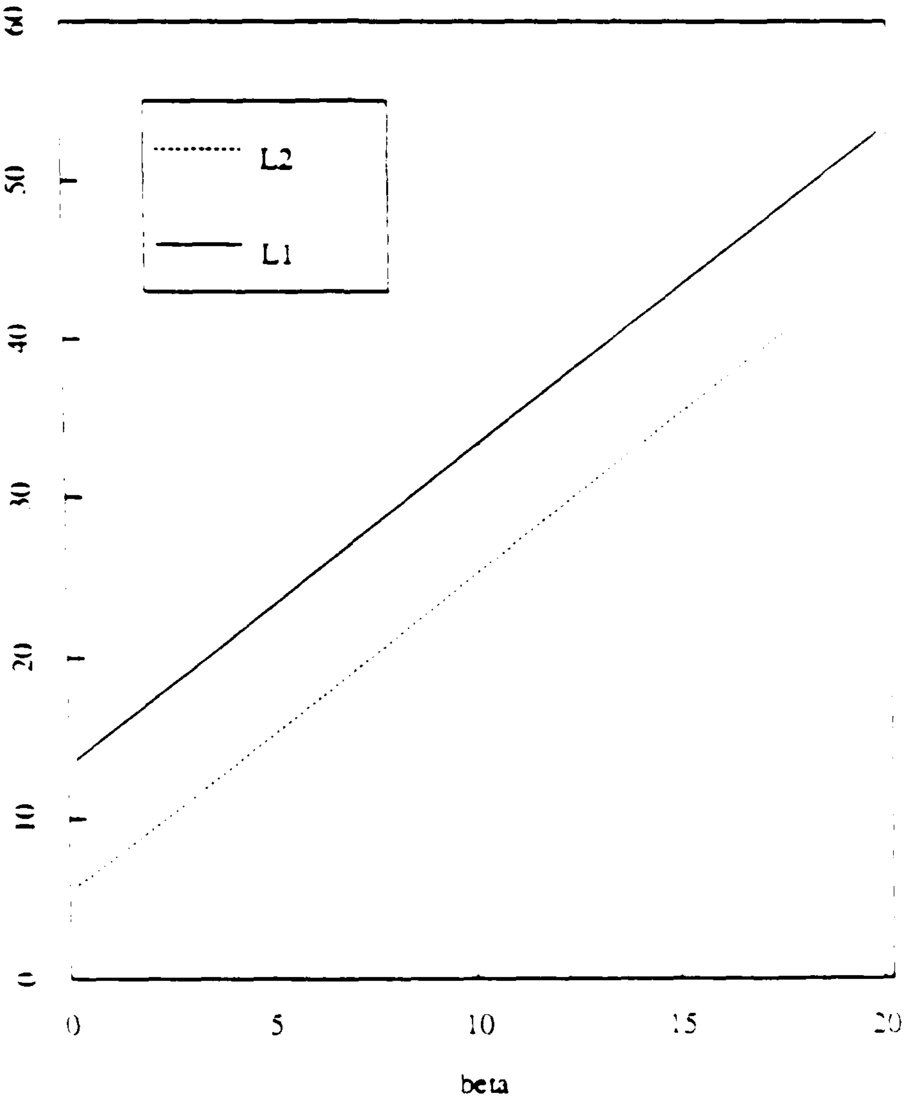


Figure 2: Residuals from linear regression

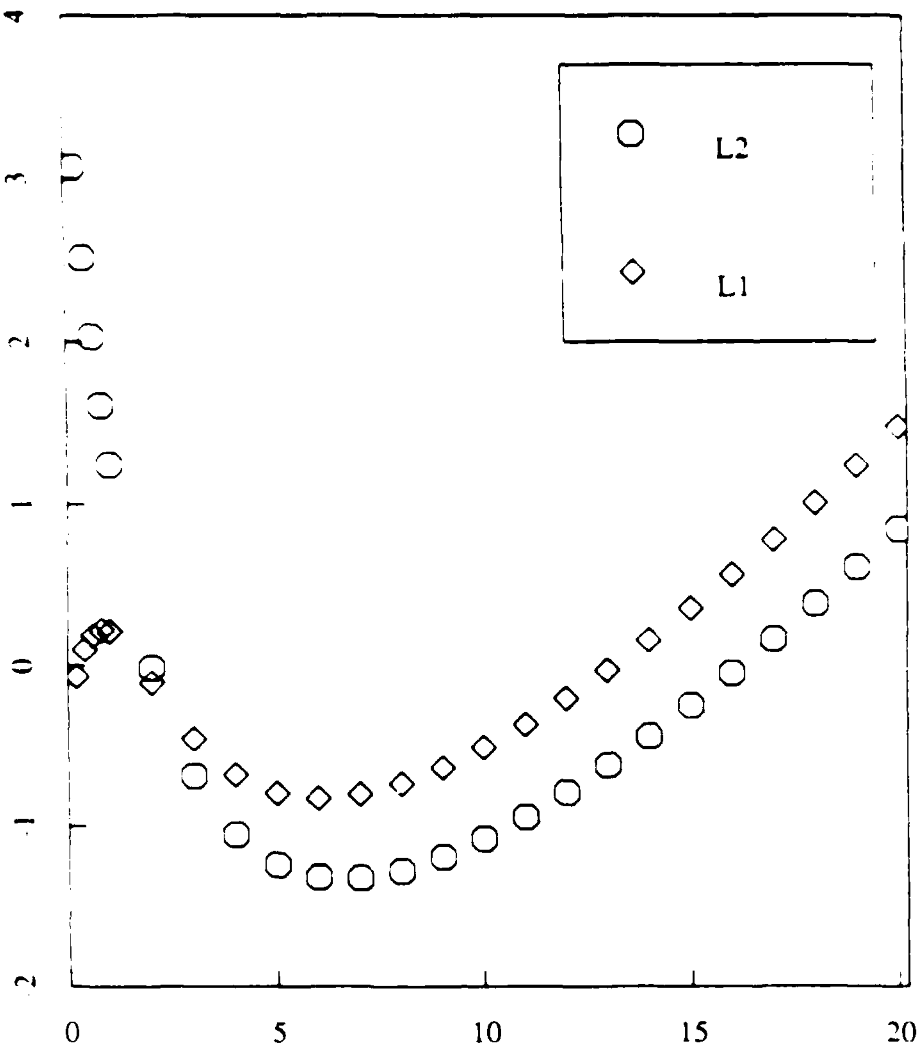


Figure 3: Comparison ***LSE*** vs "Optimal"

$Q(\eta_1)/Q(\psi)$.

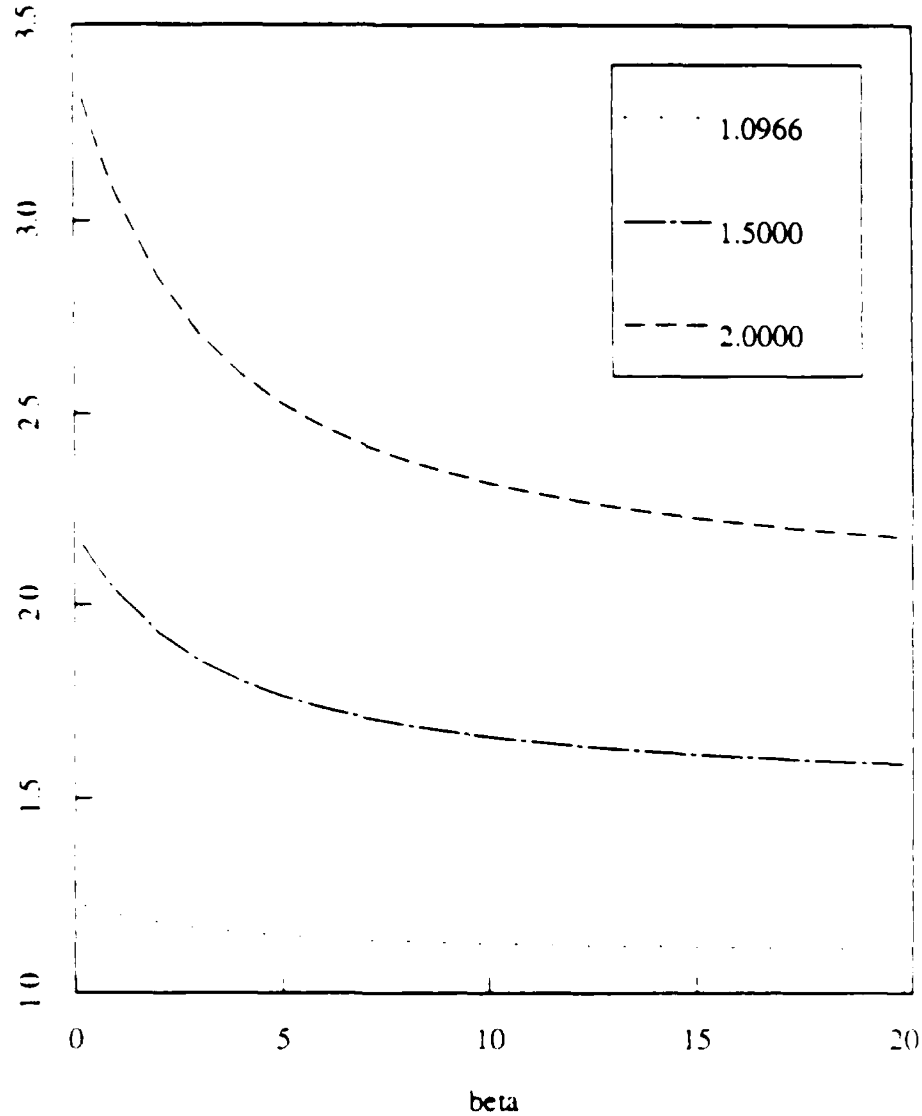
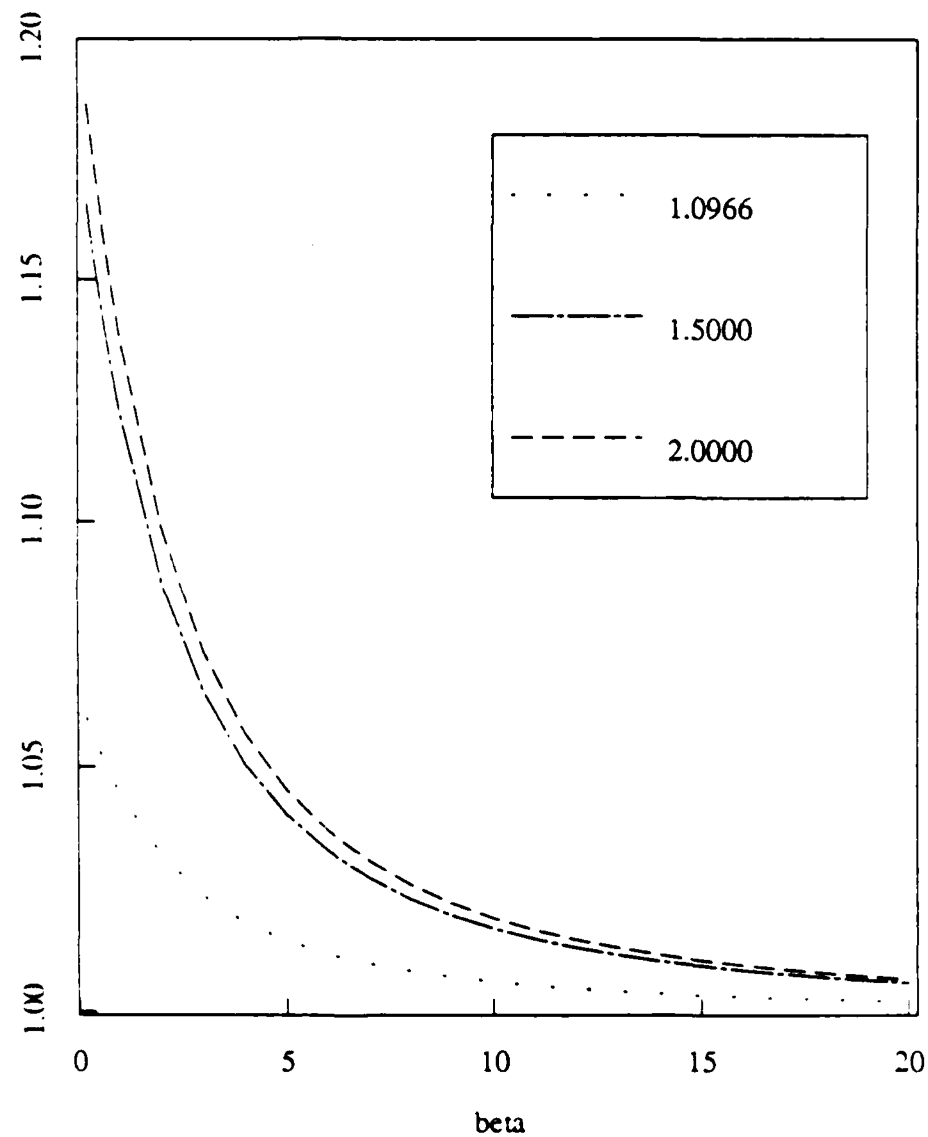


Figure 4: Comparison *MLE* vs "Optimal"

$Q(\eta_2)/Q(\psi)$.



REFERENCES

Anderson, T. W. (1959) On asymptotic distribution of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics*, **30**, 676-687.

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

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

Bobkoski, M. J. (1983) Hypothesis testing in non-stationary time series. Unpublished Ph.D. thesis. University of Wisconsin. Madison, Wisconsin.

Chan, N. H. and Wei, C. Z. (1985) Asymptotic inference of nearly non-stationary AR(1) processes. Technical Report. Department of Mathematics. University of Maryland. College Park MD.

Cramer, H. (1946) *Mathematical Methods in Statistics*. Princenton University Press.

Cumberland, W. G. and Sykes, Z. M. (1982) Weak convergence of an autoregressive process used in modeling population grown. *Journal of Applied Probability*, **19**, 450-455.

Ethier, S. N. and Kurtz, T. G. (1986) *Markov Processes: Characterization and*

Convergence. New York: Wiley.

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.. *Annals of Statistics*, **9**, 1114-1118.

Fuller, W. A. and Dickey, D. A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427-431.

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

Johnson, R. A. and Akritas, M. G. (1982) Efficiencies of tests and estimators for p-order autoregressive processes when the error distribution in nonnormal. *Annals of the Institute of Statistical Mathematics*, **34**, 579-589.

Llatas, I. (1987) Asymptotic inference for nearly non-stationary time series. Unpublished Ph. D. Thesis. University of Wisconsin. Madison, Wisconsin.

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.

Rao, M. M. (1978) Asymptotic distribution of the boundary parameter of an unstable process. *Annals of Statistics*, **6**, 185-190.

Rustagi, J. F. (1976) *Variational Methods in Statistics*. New York: Academic Press.

Stroock, D. W. and Varadhan, S. R. S. (1979) *Multidimensional Diffusion Processes*. New York: Springer-Verlag.

White, J. S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics*, **32**, 195-218.

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. *annals of mathematical statistics*, **12**, 239-241.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 95	2. GOVT ACCESSION NO. ADA181184	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) M-ESTIMATION FOR NEARLY NON-STATIONARY AUTOREGRESSIVE TIME SERIES		5. TYPE OF REPORT & PERIOD COVERED TR 12/1/86-11/30/87
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dennis D. Cox and Isabel Llatas		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0169
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, GN-22 University of Washington Seattle, Washington 98195 USA		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-661-003
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N63374 1107 NE 45th St Seattle, WA 98105		12. REPORT DATE March, 1987
		13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

M-estimation, time series, autoregressive, non-stationary

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1473 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

7-87

DTIC