ASYMPTOTIC ESTIMATION IN NONLINEAR AUTOREGRESSIVE TIME SERIES MODELS

by

Arup Bose
Purdue University

Technical Report #88-49

Department of Statistics
Purdue University

October 1988

ASYMPTOTIC ESTIMATION IN NONLINEAR AUTOREGRESSIVE TIME SERIES MODELS

by

Arup Bose
Purdue University

Abstract. A method of studying consistency and asymptotic normality is developed for least squares estimates in autoregressive time series models. Examples where this method can be applied are discussed. In many situations, the conditions required for the validity of our results are different or less stringent compared to the conditions existing in the literature. The results are specially suited for applications in ergodic models. It has potential applications in other models and other types of estimators.

AMS 1985 subject classification. Primary 62M10, 62F12; secondary 60G10, 60G42.

Key words and phrases. Nonlinear autoregression, stationary ergodic models, least square estimate, consistency, asymptotic normality.

◇

1. Introduction. Estimation of parameters in a time series model is usually performed in one of the two following ways: by least squares/conditional least squares or by minimization of a likelihood type criterion function. Achievement of nice asymptotic properties (consistency and asymptotic normality) of these estimators is not automatic because of the diverse possibilities in the choice of the model. Tjostheim (1986) discusses the need for caution in taking these properties as granted.

For nonlinear models, there are results showing the strong (or weak) consistency and asymptotic normality of the above type of estimators. Results for conditional least squares estimators are proved in Klimko and Nelson (1978) in a general set up. An extensive theory is available for the subclass of random coefficient autoregressive models in Nicholls and Quinn (1982). The work of Tjostheim (1986) deals with both maximum likelihood (m.l.) type estimators and least squares (l.s.) estimators in a time series context and generalizes the results of Klimko and Nelson (1978). Other results dealing with individual models are also available.

It may be noted that the above results are derived under the assumption of sufficiently high degree of smoothness (sometimes differentiability up to the third order) of the penalty function. It can be visualized that a certain degree of smoothness is required. Consistency in most cases can be proved with essentially first order smoothness. To proceed a step ahead and prove asymptotic normality should not require too much of an added smoothness.

In this paper, we study l.s. estimates in nonlinear autoregressive type models and show that essentially first order smoothness is enough to guarantee consistency and asymptotic normality. Our approach is to treat the criterion function as a stochastic process indexed by the parameter. This allows us a wider choice of the "error" variables and broadens the scope of the results. Martingale arguments turn out to be the main weapon. Apart from the smoothness assumptions, we need certain moment assumptions. These are comparatively easy to verify if the model has a stationary ergodic distribution.

In section 2, we state and prove our main results. In section 3, we provide some examples and comparison with the existing results of Tjostheim (1986). We also comment on the possible extension and generalization of our method to cover other situations. The appendix contains the statements and proofs of auxiliary results used in section 2.

In a separate paper, we intend to establish results which are more suited to nonergodic models. It is plausible that our method can be applied to other types of models and penalty functions. This is under investigation. The question of estimation from a Bayesian point of view and the corresponding asymptotics will also be dealt in a forthcoming report.

2. The main results. For our purposes, an observable process (X_t) is said to be a nonlinear autoregressive process if it is generated by the equation,

$$X_t = f(\theta, X_{t-1}, \ldots, X_{t-p}, t, Y_t) + \varepsilon_t, \quad t = p, p+1, \ldots \tag{2.1}$$

Here f is a (nonlinear) function and θ is the unknown parameter. (Y_t) is an observable "input" process and (ε_t) is the sequence of unknown "error" variables. X_t can be vector valued.

A particular and simple case of the model (2.1) is given by

$$X_t = f(\theta, X_{t-1}) + \varepsilon_t, \quad t = 0, 1 \ldots \tag{2.2}$$

For ease in presentation, all our results will be stated and proved for this model with X_t being real valued. With appropriate modifications in the assumptions and with virtually no change in the proofs, the results are valid for the model (2.1). This will be made clear as we proceed.

The basic assumptions on the model are given below. Other assumptions will be introduced as we proceed.

(A1) The parameter space is assumed to be the unit ball $B = \{x \in \mathbb{R}^d : \; ||x|| \le 1\}$ of the Euclidean space $\mathbb{R}^d$. The unknown true value θ_0 is assumed to be in the interior of B.

(A2) (ε_t) is a martingale difference sequence and is independent of X_0.

(A3) $|f(\theta, x) - f(\varphi, x)| \le K(|\theta - \varphi|)J(x)$ for all θ, φ and x, with a nondecreasing function K, and J is such that $E(J(X_t)) < \infty$ for all $t = 0, 1, 2, \ldots$.

REMARK 2.1. (i) The compactness of the parameter space, in particular, that the parameter space is the unit ball is a dispensable assumption. In that case, the results can be proved for a sequence of roots of the least squares equation. However, in most practical situations, compactness is a reasonable assumption and by a reparameterization (which can be absorbed in f), the parameter space can be reduced to the unit ball. We keep the assumption (A1) for clarity of presentation.

(ii) Usually the error sequence (ε_t) is assumed to be i.i.d. or at least independent. However, as we will see, what is really crucial is the martingale difference structure of (ε_t). This broadens the scope of the results substantially.

(iii) In the same manner as that for (A1), it is enough if (A3) holds for values of θ in a neighbourhood of θ_0. By localizing the arguments, our results remain valid for one sequence of roots of the l.s. equation.

The least squares function is given by

$$Q_n(\theta) = \sum_{t=1}^{n} (X_t - f(\theta, X_{t-1}))^2$$

and this is also the conditional least squares function since $E(X_t | X_0, X_1, \ldots, X_{t-1}) = f(\theta, X_{t-1})$.

The l.s. estimate θ_n is the value of θ which minimizes $Q_n(\theta)$. We will assume that θ_n is measurable. To study the asymptotic properties of θ_n, it is useful to write $Q_n(\theta)$ in the following way.

$$\begin{aligned} Q_n(\theta) &= \sum_{t=1}^{n} [f(\theta_0, X_{t-1}) + \varepsilon_t - f(\theta, X_{t-1})]^2 \\ &= I_n(\theta) - 2Z_n(\theta) + \sum_{t=1}^{n} \varepsilon_t^2 \end{aligned}$$

where

$$I_n(\theta) = \sum_{t=1}^{n} v^2(\theta, X_{t-1})$$

$$Z_n(\theta) = \sum_{t=1}^{n} v(\theta, X_{t-1}) \varepsilon_t$$

and

$$v(\theta, x) = f(\theta, x) - f(\theta_0, x).$$

Our results are going to depend on the two functions J and K. To this end we introduce the following notion.

Definition 2.2. Two real valued nondecreasing function K and g on $\mathbb{R}^+$ are said to be α compatible if $g(0) = 0$,

$$\int_B \int_B \frac{K(|\theta - \varphi|)}{g(|\theta - \varphi|)} d\theta d\varphi \leq c,$$

and $L(x) = \int_0^x u^{-2d/\alpha} dg^{1/\alpha}(u) < \infty$ for $0 < x \leq 2$. For example if $K(x) = |x|^{d+\varepsilon}$ and $g(x) = |x|^{2d+\gamma}$ with $0 < \gamma < \varepsilon$ then K and g are $(d + \varepsilon)$ compatible.

We will also have occasion to deal with the following quantity, which we call $\delta_n^J(p)$.

$$\delta_n^J(p) = \left[\sum_{t=1}^{n} EJ^2(X_{t-1})E(\varepsilon_t^2/\mathcal{F}_{t-1})\right]^{p/2} + \sum_{t=1}^{n} |EJ(X_{t-1})\varepsilon_t|^p$$

where $\mathcal{F}_t = \sigma(X_0, X_1, \dots, X_t) = \sigma(X_0, \varepsilon_1, \dots, \varepsilon_t)$.

Our first Lemma is on the behaviour of $Z_n(\theta)$.

LEMMA 2.3. Assume (A1)–(A3) and that for some $p > 0$, $\varepsilon > 0$ and a function g, K^p and g are p compatible and $\delta_n^J(p) = o(n^p(\log n)^{-(1+\varepsilon)})$. Then for some sequence $\alpha_n = o(n)$,

$$\limsup_{n\to\infty} \sup_{\theta} \frac{|Z_n(\theta)|}{\alpha_n} \le H \text{ a.s.}$$

In particular,

$$\limsup_{n\to\infty} \sup_{\theta} \frac{|Z_n(\theta)|}{n} = 0 \text{ a.s.}$$

PROOF. Note that for every θ and φ, $(Z_t(\theta) - Z_t(\varphi), \mathcal{F}_t, t \ge 1)$ is a martingale. By a well known square function inequality (see Theorem 2.12 of Hall and Heyde (1980)),

$$\begin{aligned}\Delta_n(\theta,\varphi) &= E(\sup_{1\le t\le n} |Z_t(\theta) - Z_t(\varphi)|^p) \\ &\le c\sum_{t=1}^{n} E|v(\theta, X_{t-1})\varepsilon_t|^p + cE\left[\sum_{t=1}^{n} v^2(\theta, X_{t-1})E(\varepsilon_t^2/\mathcal{F}_{t-1})\right]^{p/2}.\end{aligned}$$

By condition (A3),

$$\Delta_n(\theta,\varphi) \le cK^p(|\theta-\varphi|)\delta_n^J(p).$$

Note that $Z_t(\theta_0) = 0$ and K^p and g are p compatible. By using Lemma 2 of appendix with $||(x_1, \dots x_n)|| = \max_{1\le t\le n} |x_t|$,

$$P(\sup_{1\le t\le n} \sup_{\theta} |Z_t(\theta)| \ge c\alpha_n) \le \delta_n^J(p)/\alpha_n^p.$$

Let $A_n = \{\sup_{2^{n-1} \le t \le 2^n} \sup_\theta |Z_t(\theta)| \ge c\alpha_{2^{n-1}}\}$. Clearly $P(A_n) \le \delta^J_{2^n}(p)/\alpha^p_{2^{n-1}}$. Note that $\delta^J_n(p) = n^p \beta^p_n (\log n)^{-(1+\varepsilon)}$ for some $\beta_n \downarrow 0$. Choose $\alpha_n = n\beta_n$. Then

$$\sum_{n=1}^{\infty} P(A_n) \le \sum_{n=1}^{\infty} \delta^J_{2^n}(p)/\alpha^p_{2^{n-1}} \le c \sum_{n=1}^{\infty} n^{-(1+\varepsilon)} < \infty.$$

This proves the Lemma. □

REMARK 2.4. (i) For $p \le 2$, the condition on $\delta^J_n(p)$ is satisfied if

$$\sup_t E(\varepsilon_t^2/\mathcal{F}_{t-1}) \le c < \infty$$

and

$$\sum_{t=1}^{n} EJ^p(X_{t-1}) = o(n^p(\log n)^{-(1+\varepsilon)}).$$

This can be seen as follows.

$$\begin{aligned} E[\sum_{t=1}^{n} J^2(X_{t-1}) E(\varepsilon_t^2/\mathcal{F}_{t-1})]^{p/2} & \\ \le cE[\sum_{t=1}^{n} J^2(X_{t-1})]^{p/2} & \\ \le c\sum_{t=1}^{n} EJ^p(X_{t-1}). & \end{aligned}$$

On the other hand,

$$\sum_{t=1}^{n} EJ^p(X_{t-1})|\varepsilon_t|^p \le \sum_{t=1}^{n} EJ^p(X_{t-1})[E(\varepsilon_t^2/\mathcal{F}_{t-1})]^{p/2}$$

proving the claim.

(ii) For $p \ge 2$, the condition is satisfied if

$$\sup_t E(|\varepsilon_t|^p/\mathcal{F}_{t-1}) \le c < \infty$$

and

$$\sum_{t=1}^{n} EJ^p(X_{t-1}) = o(n^{1+p/2}(\log n)^{-(1+\varepsilon)}).$$

This can be seen as follows.

$$E[\sum_{t=1}^{n} J^2(X_{t-1})E(\varepsilon_t^2|\mathcal{F}_{t-1})]^{p/2}$$
$$\leq cE\left(\sum_{t=1}^{n} J^2(X_{t-1})\right)^{p/2}$$
$$\leq cn^{p/2-1}\sum_{t=1}^{n} EJ^p(X_{t-1}).$$

On the other hand,

$$\sum_{t=1}^{n} E|J^p(X_{t-1})\varepsilon_t|^p = \sum_{t=1}^{n} EJ^p(X_{t-1})E(|\varepsilon_t|^p/\mathcal{F}_{t-1})$$

proving the claim.

Martingale arguments cannot be used to handle $I_n(\theta)$ and we need the following condition.

(A4) For every $\delta > 0$ $\liminf_{n\to\infty} \inf_{|\theta-\theta_0|\geq\delta} n^{-1}I_n(\theta) > 0$ (a.s. or in probability).

A sufficient condition for (A4) to hold is provided in the following Lemma. The proof is trivial and hence is omitted.

LEMMA 2.5. Assume that (A3) holds. Assume further that for every θ, $n^{-1}I_n(\theta) \longrightarrow I(\theta)$ (a.s. or in probability) and $I(\theta) \neq 0$ if $\theta \neq \theta_0$. If $n^{-1}\sum_{t=1}^{n} J^2(X_{t-1})$ is bounded (a.s. or in probability) then (A4) holds a.s. or in probability respectively.

REMARK 2.6. Condition (A4) can be viewed as an identifiability condition. The sufficient condition given in Lemma 2.5 is easy to verify if the model has a stationary ergodic solution (X_t). As already remarked, we can work with a local version of condition (A4).

The consistency of θ_n can now be established quite easily.

THEOREM 2.7. Assume the conditions of Lemma 2.3 and (A4). Then $\theta_n \longrightarrow \theta_0$ a.s. or in probability respectively.

PROOF. Note that θ_n minimizes $I_n(\theta) - 2Z_n(\theta)$. However, by Lemma 2.3,

$$\limsup_{n\to\infty} n^{-1}|Z_n(\theta)| = 0 \text{ a.s.}$$

By assumption (A4),

$$\liminf_{n\to\infty} \inf_{|\theta-\theta_0|\geq\delta} n^{-1}|I_n(\theta)| > 0.$$

The consistency in probability or a.s. now follows trivially from the above two equations.□

REMARK 2.8. (i) To prove weak consistency, it is enough to have convergence in probability in Lemma 2.3, which can perhaps be achieved with weaker conditions.

(ii) The above theorem asserts the consistency of the l.s.e. and not merely the existence of a sequence of consistent roots. The latter result would hold if the given conditions are localized or the parameter space is noncompact.

Usually, once the consistency is established, asymptotic normality follows by a Taylor's expansion argument. This is the approach used, for example, by Klimko and Nelson (1978) and Tjostheim (1986). However, they assume the existence of second and sometimes third derivatives of the l.s. function. Our result will be derived under conditions only on the first order derivatives.

To tackle the first order derivatives, we need the following conditions.

(A5) Assume that the vector $\nabla f(\theta, x)$ of partial derivatives of f w.r.t. θ_i's exist and for all θ, φ (in a neighbourhood of θ_0) and all x,

(a) $||\nabla f(\theta, x) - \nabla f(\varphi, x)|| \leq K_1(|\theta - \varphi|)J_1(x)$

(b) $||\nabla f(\theta, x)| \leq J_2(x)$

To simplify our arguments, we will assume w.l.g. that (A5) holds for all θ and φ.

LEMMA 2.9. Assume that (A5) holds and for a sequence $a_n > 0$,

$$\sum_{t=1}^{n} J_1(X_{t-1})J_2(X_{t-1})K_1(a_n)a_n^2 = o_p(1).$$

Then as $n \to \infty$,

$$\sup_{|\theta-\theta_0|\leq a_n} |I_n(\theta) - \sum_{t=1}^{n}((\theta - \theta_0)'\nabla f(\theta_0, X_{t-1}))^2| \xrightarrow{P} 0$$

PROOF. For simplicity, denote $f(\theta, x)$ by $f(\theta)$.

$$\begin{aligned}
&|I_n(\theta) - \sum_{t=1}^{n}((\theta - \theta_0)'\nabla f(\theta_0))^2| \\
&= |\sum_{t=1}^{n}[((\theta - \theta_0)'\nabla f(\theta^*))^2 - ((\theta - \theta_0)'\nabla f(\theta_0))^2]| \\
&\leq \sum_{t=1}^{n}(\theta - \theta_0)'(\nabla f(\theta^*) + \nabla f(\theta_0))(\nabla f(\theta^*) - \nabla f(\theta_0))'(\theta - \theta_0) \\
&\leq 2|\theta - \theta_0|^2 \sum_{t=1}^{n} J_1(X_{t-1})J_2(X_{t-1})K_1(|\theta - \theta_0|)
\end{aligned}$$

Hence the above supremum is bounded by

$$2a_n^2 \sum_{t=1}^{n} J_1(X_{t-1})J_2(X_{t-1})K_1(a_n),$$

proving the Lemma. □

To tackle $Z_n(\theta)$, we define $v_1(\theta,x)$ as $v_1(\theta,x) = f(\theta,x) - f(\theta_0,x) - (\theta-\theta_0)'\nabla f(\theta_0,x)$.

Lemma 2.10. Assume (A5a). Suppose that for some $\varepsilon > 0$ and for a sequence $a_n > 0$,

$$\delta_n^{J_1}(d+\varepsilon)K_1^{d+\varepsilon}(a_n)a_n^{d+2\varepsilon+1} = o(1).$$

Then

$$\sup_{|\theta-\theta_0|\le a_n} |\sum_{t=1}^{n} v_1(\theta, X_{t-1})\varepsilon_t| \xrightarrow{P} 0.$$

PROOF. By mean value theorem,

$$|v_1(\theta_1,x) - v_1(\theta_2,x)| = |(\theta_1-\theta_2)'\nabla v_1(\theta^*,x)| = |(\theta_1-\theta_2)'(\nabla f(\theta^*,x) - \nabla f(\theta_0,x))|.$$

Thus if $|\theta_1-\theta_0| \le a_n$, $|\theta_2-\theta_0| \le a_n$, we have

$$|v_1(\theta_1,x) - v_1(\theta_2,x)| \le |\theta_1-\theta_2|K_1(a_n)J_1(x).$$

Following the proof of Lemma 2.3,

$$E(\sup_{1\le t\le n} |\sum_{t=1}^{n} v(\theta_1,X_{t-1})\varepsilon_t - v(\theta_2,X_{t-1})\varepsilon_t|^{d+\varepsilon})$$
$$\le |\theta_1-\theta_2|^{d+\varepsilon}K_1^{d+\varepsilon}(a_n)\delta_n^{J_1}(d+\varepsilon).$$

We will now apply Lemma 2 of appendix with

$$K_0(x) = x^{d+\varepsilon}K_1^{d+\varepsilon}(a_n)\delta_n^{J_1}(d+\varepsilon)$$

$$g_0(x) = x^{2d+\gamma} \qquad \text{where } \gamma < \varepsilon.$$

Then

$$\int_{B(0,a_n)}\int_{B(0,a_n)} \frac{K_0(|\theta-\varphi|)}{g_0(|\theta-\varphi|)}d\theta d\varphi$$
$$\le ca_n^{\varepsilon-\gamma+1}K_1^{d+\varepsilon}(a_n)\delta_n^{J_1}(d+\varepsilon) = C_1 \text{ say},$$

$$L(x) = \int_0^x u^{-2d/(d+\varepsilon)}u^{(2d+\gamma)/(d+\varepsilon)^{-1}}du \le cx^{\gamma/(d+\varepsilon)}.$$

Thus by an application of Lemma 2 of appendix,

$$P\left(\sup_{|\theta-\theta_0|\le a_n} |\sum_{t=1}^{n} v(\theta, X_{t-1})\varepsilon_t| \ge c\lambda^{1/(d+\varepsilon)}\right) \le \frac{C_1 a_n^{d+\varepsilon}}{\lambda}.$$

Now the result follows from the assumption on $\delta_n^{J_1}(d+\varepsilon)$. □

REMARK 2.11. When $\nabla f(\theta, x)$ is Lipschitz of order 1, then the condition on $\delta_n^{J_1}(d+\varepsilon)$ is satisfied if

$$\delta_n^{J_1}(d+\varepsilon) = o(a_n^{-(2d+3\varepsilon+1)}).$$

If we choose $a_n = n^{-1/2}\log n$, we get a condition similar to the one in Lemma 2.3.

The asymptotic distribution of θ_n can now be derived very easily. For this purpose, let for a sequence $\beta_n > 0$,

$$V_n = \beta_n \left(\sum_{t=1}^{n} \varepsilon_t \frac{\partial}{\partial\theta_i} f(\theta, X_{t-1})|_{\theta=\theta_0},\ i = 1, \ldots, d\right)$$
$$D_n = \beta_n^2 \text{ Diag}\left(\sum_{t=1}^{n} (\frac{\partial}{\partial\theta_i} f(\theta, X_{t-1})|_{\theta=\theta_0})^2,\ i = 1, \ldots, d\right)$$

THEOREM 2.12. Assume that θ_n is consistent and the conditions for Lemma 2.9 and 2.10 are valid with a sequence a_n such that $\beta_n = o(a_n)$. Assume further that $V_n \xrightarrow{\mathcal{D}} V$ and D_n is positive definite. Then

$$\beta_n^{-1} D_n(\theta_n - \theta_0) \xrightarrow{\mathcal{D}} V.$$

PROOF. Let $\psi_n = \beta_n^{-1}(\theta_n - \theta_0)$. From Lemma 2.9 and 2.10, the asymptotic distribution of ψ_n is the same as that of ψ which minimizes

$$\psi' D_n \psi - 2\psi' V_n.$$

However, the minimizing ψ satisfies,

$$D_n\tilde{\psi}_n = V_n.$$

This proves the theorem. □

REMARK 2.13. (i) Martingale central limit theorems can often be used to show the convergence of V_n. In typical cases $\beta_n = n^{-1/2}$ and V_n converges with this normalization. If further, D_n converges to a positive definite matrix D then we get the convergence of $\beta_n^{-1}(\theta_n - \theta_0)$. For stationary ergodic X_t, this is the case with $\beta_n = n^{-1/2}$ and then $V_n \xrightarrow{D} V$ with V being a normal random variable. It will be interesting to find examples with other limiting distributions.

(ii) As we have already mentioned, the results are geared for application in stationary ergodic models. Suitable conditions on (ε_t) (i.i.d. with density) and f (proper boundedness) can ensure this. These conditions can be derived by applying the results of Tweedie (1975). However, note that it is not necessary that X_0 is such that the model is stationary. Only the *existence* of such a distribution is needed.

(iii) One of the limitations of our approach is that if the parameter space is of high dimension, then it necessitates the existence of higher order moments. For example, if f is Lipschitz of order 1, then we need at least $EJ^{d+\varepsilon}(X_t) < \infty$ for some $\varepsilon > 0$. Nevertheless, the situation can be salvaged in the following way. If f can be split into the form $f(\theta, x) = \sum_{i=1}^{k} f_i(\tilde{\theta}_i, x)$ where $\tilde{\theta}_i$ consists of a subset of θ_i's from the full vector θ, then all the probabilistic bounds can be obtained by working with each f_i separately with the corresponding dimension d_i of $\tilde{\theta}_i$.

(iv) Under the same set up, it is not difficult to formulate and prove a distribution result for $Q_n(\theta_0) - Q_n(\theta_n)$ as $n \to \infty$. We omit the details. See Klimko and Nelson (1978) for the arguments involved.

3. Examples and discussion. In this section we will give an application. We will also discuss a modified (weighted) least squares estimator and provide comparison with Tjostheim's condition as applied to our situation.

3.1. *Comparison with Tjostheim's conditions.* Tjostheim's (1986) conditions as applied to our set up translate into the following. For clarity, we will discuss only the scalar case (both X_t and θ are real valued).

(TA1)
$$n^{-1} \sum_{t=1}^{n} f(\theta_0, X_{t-1}) \varepsilon_t \xrightarrow{P} 0$$

(TA2)
$$\liminf_{n\to\infty} n^{-1} \left[\sum_{t=1}^{n} f'^2(\theta_0, X_{t-1}) - \sum_{t=1}^{n} f''(\theta_0, X_{t-1}) \varepsilon_t \right] > 0 \text{ a.s.}$$

(TA3)
$$\limsup_{n\to\infty} \limsup_{\delta\downarrow 0} (n\delta)^{-1} \Big| \sum_{t=1}^{n} \big[\varepsilon_t (f''(\theta^*, X_{t-1}) - f''(\theta_0, X_{t-1})) + (f'^2(\theta_0, X_{t-1}) - f'^2(\theta^*, X_{t-1})) \big] \Big| < \infty \text{ a.s.}$$

(TB1)
$$n^{-1} \sum_{t=1}^{n} \left[f'^2(\theta, X_{t-1}) - f''(\theta_0, X_{t-1}) \varepsilon_t \right] \longrightarrow V > 0 \text{ a.s.}$$

(TB2)
$$n^{-1/2} \sum_{t=1}^{n} f'(\theta_0, X_{t-1}) \varepsilon_t \xrightarrow{D} N(0, W)$$

The point θ^* is the point such that for $|\theta - \theta_0| < \delta$,

$$Q_n(\theta) = Q_n(\theta_0) + (\theta - \theta_0)' Q_n'(\theta_0) + (\theta - \theta_0)^2 Q_n''(\theta^*)$$

The conditions (TA1) – (TA3) are needed for consistency and (TB1), (TB2) are needed for asymptotic normality. These conditions clearly involve the second derivative of f and the condition (TA3) is difficult to check. Other conditions are relatively easy to verify if (X_t) is stationary ergodic. Clearly the above set of conditions overlap with ours. However, Tjostheim's set up is more general. But the set of sufficient conditions which Tjostheim provides to apply his results to conditional least squares are more stringent and involve the third derivative of f. See pages 254–255 of Tjostheim (1986).

3.2. *An example.* Let (X_t) be a process generated by the model

$$X_t = (\psi + \pi \exp(-\gamma X_{t-1}^2))X_{t-1} + \varepsilon_t, \quad t \geq 1$$

where (ε_t) are i.i.d., $E\varepsilon_t = 0$, $E\varepsilon_t^2 < \infty$.

This model is a special case of exponential autoregressive models introduced by Ozaki (1980) and Haggan and Ozaki (1981).

THEOREM 3.1. Assume that $|\psi| + |\pi| < 1$ and that γ is positive and bounded away from zero. Suppose that there exists a unique distribution for the initial variable X_0 such that $(X_t, t \geq 1)$ is strictly stationary and ergodic. [This is possible if ε_t has a density with infinite support. See Tjostheim (1986), page 256, 258.] Let $(\psi_n, \pi_n, \gamma_n)$ denote the l.s. estimate.

Then $(\psi_n, \pi_n, \gamma_n) \longrightarrow (\psi, \pi, \gamma)$ a.s. and $n^{1/2}(\psi_n, \pi_n, \gamma_n)$ is asymptotically normal.

PROOF. To prove this let $\theta = (\pi, \psi, \gamma) = (\tilde{\theta}_1, \tilde{\theta}_2)$ where $\tilde{\theta}_1 = \pi$, $\tilde{\theta}_2 = (\psi, \gamma)$.

$$f(\theta, x) = f_1(\tilde{\theta}_1, x) + f_2(\tilde{\theta}_2, x)$$

where $f_1(\tilde{\theta}_1, x) = \pi x$, $f_2(\tilde{\theta}_2, x) = \psi \exp(-\gamma x^2)x$.

To apply Theorems 2.7 and 2.12, let $d = 1$, $K(x) = J(x) = x$ for f_1 and $d = 2$, $K(x) = x$, $J(x) = c$ for f_2. For f_1, K_1 and J_1 are redundant and for f_2, $K_1(x) = x$, $J_1(x) = c$. Note that $E\varepsilon_t^2 < \infty$ implies $EX_t^2 < \infty$. All other conditions are easily verified by using the ergodic theorem and letting $a_n = n^{-1/2}$. □

REMARK 3.2. If it is only known that $\gamma \geq 0$ then we need $E|\varepsilon_t|^{6+\varepsilon} < \infty$ for some $\varepsilon > 0$ as opposed to $E\varepsilon_t^2 < \infty$. Note that Tjostheim proves the same result with $E\varepsilon_t^6 < \infty$, but apparently needs $E|\varepsilon_t|^7 < \infty$ (see Tjostheim (1986), page 257).

3.3. *Modified (weighted) least squares.* A weighted l.s. estimate θ_{wn} is obtained if we minimize

$$\sum_{t=1}^{n} \frac{(X_t - f(\theta, X_{t-1}))^2}{g(t, X_{t-1})}$$

where $g(t, X_{t-1})$ is a weight function. This would be relevant if, for instance,

$$E(\varepsilon_t^2/\mathcal{F}_{t-1}) = E[(X_t - f(\theta_0, X_{t-1}))^2/\mathcal{F}_{t-1}] = g(t, X_{t-1})$$

so that the errors are "heteroscedastic". Below, we give one example of this type of a situation. Our technique can be adapted for use in this situation too since the basic structures of $I_n(\theta)$ and $Z_n(\theta)$ remain the same.

EXAMPLE 3.3. Let (X_t) be a RCA process satisfying

$$X_t = (\theta + b_t)X_{t-1} + \varepsilon_t, \quad t = 0, 1, 2, \ldots$$

where (b_t), (ε_t) are i.i.d. with mean zero and $Eb_t^2 = \wedge$, $E\varepsilon_t^2 = \sigma^2$ are known. θ is the unknown parameter, with $|\theta_0| < 1$. We will assume that a stationary ergodic solution (X_t) exists. See Nicholls and Quinn (1982) for conditions which makes this possible.

(a) It is easily seen that

$$E((X_t - \theta X_{t-1})^2/\mathcal{F}_{t-1}) = X_{t-1}^2 \wedge + \sigma^2.$$

Hence the weighted l.s. estimate θ_{wn} of θ is obtained by minimizing

$$\sum_{t=1}^{n} \frac{(X_t - \theta X_{t-1})^2}{X_{t-1}^2 \wedge + \sigma^2}$$

Thus $\theta_{wn} = \left(\sum_{t=1}^{n} \frac{X_t X_{t-1}}{X_{t-1}^2 \wedge + \sigma^2}\right)\left(\sum_{t=1}^{n} \frac{X_{t-1}^2}{X_{t-1}^2 \wedge + \sigma^2}\right)^{-1}$. It is relatively easy to show that

$$\theta_n \to \theta_0 \quad \text{a.s.}$$

and $n^{\frac{1}{2}}(\theta_n - \theta_0) \xrightarrow{\mathcal{D}} N(0,\ \sigma_1^2)$ where

$$\sigma_1^2 = \frac{E\left(\frac{b_t X_{t-1}^2 + X_{t-1}\varepsilon_t}{X_{t-1}^2 \wedge + \sigma^2}\right)^2}{E\left(\frac{X_{t-1}^2}{X_{t-1}^2 \wedge + \sigma^2}\right)^2}.$$

(b) If θ_{1n} is the conditional l.s.e. obtained by minimizing $\sum (X_t - \theta\ X_{t-1})^2$ then it can be shown that if $E\ X_t^4 < \infty$ (guaranteed if $E(\varepsilon_t^4 + b_t^4) < \infty$) then $\theta_{1n} \to \theta_0$ a.s. and $n^{\frac{1}{2}}(\theta_{1n} - \theta) \xrightarrow{\mathcal{D}} N(0,\ \wedge + \sigma^2)$. Thus, we need stronger condition for the conditional l.s. compared to the weighted l.s.

(c) The case when $\wedge$ and σ^2 are unknown.

Tjostheim (1986) discusses this situation and uses the following maximum likelihood type penalty function

$$L(n) = \sum_{t=1}^{n} \left[\log(X_{t-1}^2 \wedge + \sigma^2) + \frac{(X_t - \theta X_{t-1})^2}{X_{t-1}^2 \wedge + \sigma^2}\right].$$

He shows that the estimates $(\wedge_n,\ \sigma_n^2)$ obtained by minimizing $L(n)$, are consistent. The estimate θ_n is given by

$$\theta_n = \left(\sum_{t=1}^{n} \frac{X_t X_{t-1}}{X_{t-1}^2 \wedge_n + \sigma_n^2}\right)\left(\sum_{t=1}^{n} \frac{X_{t-1}^2}{X_{t-1}^2 \wedge_n + \sigma_n^2}\right)^{-1}.$$

It can be shown by easy calculations that

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \left[\frac{X_{t-1}^2 b_t}{X_{t-1}^2 \wedge_n + \sigma_n^2} - \frac{X_{t-1}^2 b_t}{X_{t-1}^2 \wedge + \sigma^2} \right] \xrightarrow{P} 0$$

and

$$n^{-1} \sum_{t=1}^{n} \left[\frac{X_{t-1}^2}{X_{t-1}^2 \wedge_n + \sigma_n^2} - \frac{X_{t-1}^2}{X_{t-1}^2 \wedge + \sigma^2} \right] \longrightarrow 0 \text{ a.s.}$$

This shows that θ_n has the same asymptotic properties as that of θ_{wn} obtained in (a). Note that we need only the conditions assumed in (a). However, if we need the asymptotic normality of $(\wedge_n, \sigma_n^2)$, we also need $E(\varepsilon_t^4 + b_t^4) < \infty$. See Tjostheim (1986) for details.

CONCLUDING REMARK 3.4. Our method can be very easily extended to cover the case when X_t is vector valued and/or the autoregressive process is of order $p > 1$. It seems that the method has potential applications in other types of models and estimators. The results for a regression model of the form $X_t = f(\theta, Y_t) + \varepsilon_t$ where Y_t and ε_t are independent follow easily from our results.

This method can also be used to study the asymptotic properties of the Bayes estimator. This will be discussed in a forthcoming paper. Our method should also work for nonstationary models and this is under investigation.

It will be interesting to derive the rate at which the distribution converges. It will also be interesting to see if the above rate can be improved by using a bootstrap approximation for the distribution. The simple case of a linear autoregressive process has been worked out and the bootstrap approximation in a certain sense provides an improvement over the standard, normal approximation. See Bose (1988) for details.

APPENDIX

In this appendix we state and prove the results we have repeatedly used in sections 2 and 3.

LEMMA 1. Let p and ψ be two increasing continuous functions on $[0, \infty)$ such that $p(0) = \psi(0) = 0$ and $\psi(\infty) = \infty$. Suppose L is a normal linear space (with norm $\|\cdot\|$) and $f: \mathbb{R}^d \to L$ is a continuous function on $B = B(a, \rho) = \{x: |x - a| < \rho\} \subseteq \mathbb{R}^d$. Then

$$\iint\limits_{B \times B} \psi\left[\frac{\|f(x) - f(y)\|}{p(|x-y|)}\right] dxdy \leq B_\rho \text{ implies}$$

$$\|f(x) - f(y)\| \leq 8 \int_0^{|x-y|} \psi^{-1}\left(\frac{4^{d+2} B_\rho}{\beta^2 u^{2d}}\right) p(du)$$

where

$$\beta = \inf_{x \in B(a,\rho)} \inf_{1 < \rho^* \leq 2} \frac{|B(x, \rho^*) \cap B(a, 1)|}{\rho^d}$$

and $|A|$ = Lebesgue measure of A for $A \subseteq \mathbb{R}^d$. This Lemma is a crucial tool in proving existence of continuous versions of stochastic processes and a proof can be found on page 7 of Stroock (1982).

The next Lemma shows how the supremum of a process can be controlled in terms of the behavior of certain two dimensional moments.

LEMMA 2. Let $\{Y(\theta): \theta \in B = B(a, \rho)\}$ be a class of random variables taking values in a normed linear space L (with norm $\|\cdot\|$). Suppose that the following conditions hold.

(a) $\theta \to Y(\theta)$ is a.s. continuous

(b) there exists $\alpha > 0$ and functions g and K such that

(i) $E\|Y(\theta) - Y(\varphi)\|^\alpha \leq K(|\theta - \varphi|), \quad \forall \theta, \varphi \in B$

(ii) $\int\int_{B\times B} \frac{K(|\theta-\varphi|)}{g(|\theta-\varphi|)} d\theta d\varphi \le C_\rho$

(iii) K is non decreasing and g is continuous and increasing with $g(0) = 0$ such that

$$L(x) = \int_0^x u^{-\frac{2d}{\alpha}} dg^{\frac{1}{\alpha}}(u) < \infty, \quad \forall x > 0.$$

Then the following probability bound holds (β is as defined in Lemma 1).

$$P\Big(\sup_{\theta,\varphi\in B(a,\rho)} \frac{\|Y(\theta) - Y(\varphi)\|}{L(|\theta-\varphi|)} \ge c(\frac{\lambda}{\beta^2})^{\frac{1}{\alpha}}\Big) \le C_\rho/\lambda$$

where c is a constant depending on d. In particular if $L_\rho = \sup_{\theta,\varphi\in B} L(|\theta-\varphi|)$ then

$$P\Big(\sup_{\theta,\varphi\in B} \|Y(\theta) - Y(\varphi\| \ge c\, L_\rho(\frac{\lambda}{\beta^2})^{\frac{1}{\alpha}}\Big) \le C_\rho/\lambda.$$

If further, there exists a $\theta_0 \in B$ such that $Y(\theta_0) = 0$ then

$$P\big(\sup_{\theta\in B} \|Y(\theta)\| \ge c\, L_\rho(\frac{\lambda}{\beta^2})^{\frac{1}{\alpha}}\big) \le C_\rho/\lambda.$$

PROOF. By the given condition,

$$E\Big(\int\int_{B\times B} \Big[\frac{\|Y(\theta) - Y(\phi)\|}{g^{1/\alpha}(|\theta-\varphi|)}\Big]^\alpha d\theta d\varphi\Big)$$
$$\le \int\int_{B\times B} \frac{K(|\theta-\varphi|)}{g(|\theta-\varphi|)} d\theta d\varphi \le C_\rho.$$

Hence

$$P\Big(\int\int_{B\times B} \Big[\frac{\|Y(\theta) - Y(\phi)\|}{g^{1/\alpha}(|\theta-\varphi|)}\Big]^\alpha d\theta d\varphi \ge \lambda\Big) \le C_\rho/\lambda.$$

Note that $\psi(x) = x^\alpha$ and $p(x) = g^{1/\alpha}(x)$ satisfy conditions of Lemma 1. Hence by an application of Lemma 1, whenever

$$\int\int_{B\times B} \Big[\frac{\|Y(\theta) - Y(\varphi)\|}{g^{1/\alpha}(|\theta-\varphi|)}\Big]^\alpha d\theta d\varphi \le \lambda \text{ we have}$$

$$\|Y(\theta)-Y(\varphi)\| \le 8\int_0^{|\theta-\varphi|}\left[\frac{4^{d+2}\lambda}{\beta^2u^{2d}}\right]^{1/\alpha} dg^{1/\alpha}(u)$$
$$= c\beta^{-2/\alpha}\lambda^{1/\alpha}\int_0^{|\theta-\varphi|} u^{-2d/\alpha} dg^{1/\alpha}(u)$$
$$= c\beta^{-2/\alpha}\lambda^{1/\alpha} L(|\theta-\varphi|).$$

Thus $P\left(\sup_{\theta,\varphi\in B} \frac{\|Y(\theta)-Y(\varphi)\|}{L(|\theta-\varphi|)} \ge c\ \beta^{-2/\alpha}\lambda^{1/\alpha}\right) \le C_\rho/\lambda$ proving the first part of the Lemma. The remaining parts of the Lemma follow easily from this.

REMARK. Conditions (b) (ii) and (iii) state that K should be increasing sufficiently fast so as to accommodate a suitable increasing g with $L(x) < \infty$. These conditions will be satisfied if for instance $K(x)$ can be chosen to be $C\ x^{d+\varepsilon}$ for some $\varepsilon > 0$. In that case take

$$g(x) = x^{2d+\varepsilon/2} \text{ and then}$$
$$L(x) = c\int_0^x u^{\frac{-2d}{\alpha}+\frac{2d+\varepsilon/2}{\alpha}-1} du$$
$$= c\int_0^x u^{\frac{\varepsilon}{2\alpha}-1} du < \infty, \quad \forall x.$$

REFERENCES

BOSE, A. (1988). Edgeworth correction by bootstrap in autoregressions. To appear in *Ann. Statist.* **6**, 4(December 1988).

HAGGAN, V. and OZAKI, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189–196.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

KLIMKO, L. A. and NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629–642.

NICHOLLS, D. F. and QUINN, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*. Lecture Notes in Statistics **11**, Springer, New York.

OZAKI, T. (1980). Nonlinear time series models for nonlinear random vibrations. *J. Appl. Probab.* **17**, 84–93.

STROOCK, D. W. (1982). *Lectures on Topics in Stochastic Differential Equations*. TIFR Lecture Notes published by Springer Verlag, New York.

TJOSTHEIM, D. (1986). Estimation in nonlinear time series models. *Stoch. Proc. Appl.* **21**, 251–273.

TWEEDIE, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stoch. Proc. Appl.* **3**, 385–403.

Professor Arup Bose
Department of Statistics
Mathematical Sciences Building
Purdue University
West Lafayette, IN 47907