ON ADAPTIVE ESTIMATION IN THE NON–LINEAR THRESHOLD AR(1) MODEL

by

Beong–Soo So
Purdue University

Technical Report #90-45

Department of Statistics
Purdue University

August 1990
Revised May 1991

*

ON ADAPTIVE ESTIMATION IN THE NON–LINEAR THRESHOLD AR(1) MODEL

by

Beong–Soo So
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399, U.S.A.

Running title: Estimation in Non-linear AR(1) Model

⋆

ON ADAPTIVE ESTIMATION IN THE NON–LINEAR THRESHOLD AR(1) MODEL

by

Beong–Soo So
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399, U.S.A.

Abstract. In this paper we consider non–linear threshold AR(1) processes, where the parameter consists of the autoregressive parameter and a shape parameter $f(\cdot)$, the density of the errors. On the basis of the local asymptotic normality of the model, we construct the efficient estimator for the euclidean parameter under nuisance shape parameter $f(\cdot)$ of the error distribution. This adaptive estimator seems to be asymptotically as good as estimators constructed if f is regarded as fixed, even if f is not symmetric.

Keywords and phrases: Non–linear threshold autoregressive process; local asymptotic normality; local asymptotic minimax bounds; adaptive estimator.

Running title: Estimation in Non-linear AR(1) Model

1. Introduction

In this paper we consider stochastic processes (X_t) defined by the following difference equation;

$$X_t = \sum_{k=1}^{p} [\phi(0,k) + \phi(1,k)X_{t-1}] I(X_{t-1} \in R_k) + \epsilon_t, \quad t \in Z, \tag{1.1}$$

where $I(A)$ is the indicator of the set A,

$$R_k = (\gamma_{k-1}, \gamma_k], \ 1 \le k \le p,$$

$$-\infty = \gamma_0 < \gamma_1 < \gamma_2 < \ldots < \gamma_p = \infty$$

and random variables ϵ_t, $t \in Z$, are assumed to be independent and identically distributed with Lebesgue density f.

Such non–linear autoregressive (AR) processes, which are called as the first–order threshold AR models or SETAR(1) models in the time–series literature, are known to be practically relevant in the analysis of many time–series data. See for example Tong (1983).

The parameter $(\phi(i,k);\ i = 0,1,\ k = 1,\ldots,p; f) = (\phi; f)$ is assumed to obey the following regularity conditions.

(1.2)

(i) $\phi(1,1) < 1,\ \phi(p,1) < 1,\ \phi(1,1)\phi(1,p) < 1.$

(ii) f is absolutely continuous with positive finite Fisher information $I(f) = \int (f'/f)^2 f dx < \infty.$

(iii) $f(x),\ \forall x \in R,\ \int x^2 f(x)dx = \sigma^2 < \infty,$

$$\int x f(x) dx = 0.$$

It is well–known that condition (i) is sufficient for the *geometric ergodicity* of the process $\{X_t\}$ of (1.1) and is almost necessary for the *ergodicity*. See Chan and Tong (1985) for further details. Without loss of generality, we can reparametrize the model (1.1) and rewrite the model in the form;

$$X_t = \theta_0 + \sum_{k=1}^{p} [\theta(0,k) + \theta(1,k)X_{t-1}] I(X_{t-1} \in R_k) + \epsilon_t, \ t \in Z, \tag{1.3}$$

where we assume $\theta(0,1) \equiv 0$ without loss of generality in order to ensure *identifiability* of the model (1.3).

One relevant issue in this model is the efficient estimation of the euclidean parameter $(\theta(i,k);\ i = 0,1;\ k = 1,\ldots,p) = \theta$ when the exact shape f of the error distribution is regarded as a nuisance parameter. Previous results on this subject include adaptive estimation in the linear AR(p) model by Kreiss (1986, 1987) and the corresponding result by So (1990) for the non–linear threshold AR(1) model with symmetric error density $f(\cdot)$. Our main new results in this paper are the proof of the fact that adaptation is possible for the estimation of θ and the construction of the efficient estimator which is adaptive for all densities f of the distribution of the white noise $(\epsilon_t : t \in Z)$ and not only for symmetric ones.

This paper is organized as follows:

In section 2 we prove *local asymptotic normality* (LAN) under a local parametrization for our model. While Kreiss (1987) obtains similar results for the linear AR(p) model, we need to employ special properties of the non–linear threshold AR(1) models to ensure the desired local asymptotic normality.

Section 3 deals with *asymptotic minimax bounds* for estimation of θ under nuisance function f and the corresponding *convolution theorem* for the regular estimators. As is noted by Kreiss (1987) in the linear AR(p) models, this lower bound is shown to be equal to the bound obtained when f is regarded fixed.

In section 4 we construct a sequence of adaptive estimators of θ which achieves this lower bound for all densities f of the error distribution and not only for symmetric ones. The methods are different from those used in Kreiss (1986) and So (1990) and more similar to the ideas of Stone (1975) and Kreiss (1987). While Kreiss (1987) uses similar construction in the linear AR(p) model, his proof cannot be carried over to the non–linear model considered here and our proof depends heavily on the special *non–linear* structure of the threshold AR(1) model.

To make the paper more inviting to read all technical proofs are given in section 5.

2. Local Asymptotic Normality

We consider the following local parameter $(h, \beta) \in H_n$, with

$$\begin{aligned} \theta_n &= \theta + n^{-1/2} h, \; h \in R^{2p} \\ f_n &= [(1 - \int \beta^2 d\lambda / n)^{1/2} \sqrt{f} + \beta/\sqrt{n}]^2, \\ & \beta \in L_2(R, \lambda), \; \beta \perp \sqrt{f}, \end{aligned} \tag{2.1a}$$

where $H_n \subset R^{2p} \times L_2(\lambda)$ is chosen so that (1.2) (i) is fulfilled and $\int \beta^2 d\lambda < n$. We note that, as $n \to \infty$, $H_n \to H = R^{2p} \times \{\beta \in L_2(\lambda) | \beta \perp \sqrt{f}\}$.

Remark 2.1. As is noted in Kreiss (1987), it is enough to consider parametrization of f_n via β, such that $\int [\sqrt{n}(\sqrt{f_n} - \sqrt{f}) - \beta]^2 d\lambda \to 0$.

See also Begun et al. (1983) for similar parametrization.

Finally we define on H the following scalar–product:

$$< (h_1, \beta_1), \; (h_2, \beta_2) >= h_1^T \Gamma I(f) h_2 + 4 \int \beta_1 \cdot \beta_2 d\lambda, \tag{2.1b}$$

where Γ is the $(2p \times 2p)$–covariance matrix of an SETAR(1)–process with parameter θ.

Now the density of the distribution $P_{n,(h,\beta)}$ of $(X_0, \dots, X_n)$ can be expressed in the form.

$$g(X_0; h, \beta) \cdot \prod_{i=1}^{n} f_n(X_i - \theta_n^T X(i-1)), \tag{2.2}$$

where the abbreviation

$$X(i-1) = (1, I(X_{i-1} \in R_1), \; X_{i-1} I(X_{i-1} \in R_1), \dots, X_{i-1} I(X_{i-1} \in R_p))$$

is used.

To obtain local asymptotic normality, we also assume the following

$$\begin{aligned} & g_n(X_0; h_n, \beta_n) \to g(X_0; 0, 0) \text{ in} \\ & P_{n,0} - \text{probability, if } (h_n, \beta_n) \to (h, \beta). \end{aligned} \tag{2.3}$$

Now we can establish the following local asymptotic normality (LAN) for the model.

THEOREM 2.1. *Let* $(\beta_n, h_n) \to (h, \beta) \in H$. *Then*

$$\log(dP_n(h_n,\beta_n)/dP_{n,0}) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\dot{\varphi}(e_i)h^T X(i-1)+$$

$$2\beta(e_i)/\sqrt{f(e_i)}\} + \tfrac{1}{2}\{h_1^T\Gamma h_1 I(f) + 4\|\beta(h)\|^2\} = o_{p_{n,0}}(1), \tag{2.4}$$

and

$$\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\dot{\varphi}(e_i)h^T X(i-1) + 2\frac{\beta(e_i)}{\sqrt{f(e_i)}}\}|P_{n,0}\right)$$

$$\Rightarrow N(0, h_1^T\Gamma h_1 I(f) + 4\|\beta(h)\|^2), \tag{2.5}$$

where $\phi = -f'/f$, $e_i = X_i - \theta^T X(i-1)$, *and* $\beta(h) = \beta + \dot{\varphi}\sqrt{f}/2(h_0 + h_1^T\overline{X}_1(j-1))$, $h = (h_0, h_1)$. □

As well–known consequences of (2.4) and (2.5) we have

COROLLARY 2.2. *Under the same assumptions as theorem* 2.1

$$\{P_n(h_n,\beta_n); n \in Z^+\} \text{ and } \{P_{n,0}; n \in Z^+\} \tag{2.6}$$

are contiguous if (h_n) *is bounded and*

$$\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\dot{\varphi}(e_i)h^T X(i-1) + \frac{2\beta(e_i)}{\sqrt{f(e_i)}}\}|P_n(h_n,\beta_n)\right)$$

$$\Rightarrow N(h_1^T\Gamma h_1 I(f) + 4\|\beta(h)\|^2), h_1^T\Gamma h_1 I(f) + 4\|\beta(h)\|^2), \tag{2.7}$$

$$\mathcal{L}\left(\sum_{j=1}^{n}\dot{\varphi}(e_j)[X_1(j-1) - \overline{X}_1] - \Gamma^{1/2}\sqrt{I(f)}h_1|P_n(h_n,\beta_n)\right)$$

$$\Rightarrow N(0, \Gamma I(f)), \tag{2.8}$$

where $X(j-1) = (1, X_1(j-1)),\ h = (h_0, h_1)$

$\overline{X}_1 = E\{X_1(j-1)\}$ and $\Gamma = E[X_i - \overline{X}_1]^T[X_1 - \overline{X}_1]$.

□

3. Locally asymptotic minimax (LAM) bounds

We note that tangent spaces for $\theta_0,\ \theta_1,\ f$ are given by

$$P_{\theta_0} = \left\{ \sum_{j=1}^{n} \dot{\varphi}(e_j)h/\sqrt{n};\ h \in R \right\}, \tag{3.1}$$

$$P_{\theta_1} = \left\{ \sum_{j=1}^{n} \dot{\varphi}(e_j)X_1^T(j-1)h_1/\sqrt{n};\ h \in R^{2p-1} \right\}, \tag{3.2}$$

$$P_f = \left\{ \sum_{j=1}^{n} [\beta(e_j)/\sqrt{f(e_j)}]/\sqrt{n};\ \beta \in L_2(\lambda) \right\}, \tag{3.3}$$

respectively. Here $P_{\theta_0},\ P_{\theta_1},\ P_f$ denote subspaces of the Hilbert space $L_2(R^n, \mathcal{B}^n, \lambda^n)$.

Since we are interested in estimating θ_1 under the nuisance parameters (θ_0, f), we note that the *effective score function* Δ_n^* for θ_1 is given by

$$\Delta_n^* = \sum_{j=1}^{n} \dot{\varphi}(e_j)[X_1(j-1) - \overline{X}_1(j-1)]/\sqrt{n}, \tag{3.4}$$

and the corresponding *effective information* matrix I_* for θ_1 is given by

$$I_* = I(f)\Gamma, \tag{3.5}$$

where $X(j-1) = (I(X_{j-1} \in R_1); X_1(j-1))$, $\overline{X}_1(j-1) = EX_1(j-1)$ and Γ is a covariance matrix of $X_1(j-1)$. One important consequence of the above results is the following local asymptotic minimax (LAM) lower bound.

THEOREM 3.1. (LAM bound) *Suppose that $\ell(\cdot)$ is lower semicontinuous and subconvex. Then*

$$\begin{aligned} &\lim_{c\to\infty} \liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{(h,\beta)\in K_c} \int \ell(\sqrt{n}(\hat{\theta}_n - \theta_1))dP_n(h,\beta) \\ &\geq E\ell((I(f)\Gamma)^{-1/2}Z), \end{aligned} \tag{3.6}$$

where K_c denotes compact subset of the set $\{(h,\beta) \in H\colon \|(h,\beta)\| < c\}$ and Z is a standard normal random variable on R^{2p-1}. □

Now we introduce definition of the *regular* estimators.

Definition 3.1: $\{\hat\theta_n\}$ is said to be regular sequence of estimators if for any sequence $\theta_n = \theta_0 + h_n/\sqrt{n}$, $f_n^{1/2} = f_0^{1/2} + \beta_n/\sqrt{n}$ such that $(h_n, \beta_n) = (h, \beta) + o(1)$,

$$\mathcal{L}[\sqrt{n}(\hat\theta_n - \theta)|P_n(h_n,\beta_n)] \Rightarrow \mathcal{L}(V) \tag{3.7}$$

where $\mathcal{L}(V)$ does not depend on the choice of sequences (h_n, β_n). For regular estimators we have the following convolution theorem.

THEOREM 3.2. (Convolution theorem). *Let $\{\hat\theta_n\}$ be a sequence of regular estimates.*

$$\mathcal{L}(\sqrt{n}(\hat\theta_n - \theta)|P_n(h_n,\beta_n)) \Rightarrow \mathcal{L}(Z_* + V) \tag{3.8}$$

where $Z_* \sim N(0, (I(f)\Gamma)^{-1})$ and V is some random variable which is independent of Z_*. □

Above results suggest the two alternative definitions of the optimality of the sequence of estimators; namely.

Definition 3.2: A sequence of estimator which achieves LAM lower bound in (3.6) is said to be LAM–*efficient.*

Definition 3.3: A regular sequence of estimators whose asymptotic distribution is $N(0, I_*^{-1})$ is said to be *regular–efficient.*

In order to show that a sequence of estimators $\{T_n\}$ is efficient in either of the two senses, it suffices to show that they are *asymptotically linear*; namely

$$T_n = \theta_1 + I_*^{-1}\Delta_n^*(\theta)/\sqrt{n} + o_{p_{n,0}}(1), \tag{3.9}$$

where $\Delta_n^*(\theta) = \sum_{j=1}^{n} \dot\varphi(e_j(\theta))[X_1(j-1) - \overline{X}_1]/\sqrt{n}$ is the effective score function for θ_1.

Remark 3.1. We note that regularity of the estimator $\{T_n\}$ follows immediately from that of the sequence $\{\theta_1 + I_*^{-1}\Delta_n^*(\theta)/\sqrt{n}\}$ which in turn is implied by (2.8).

4. Construction of the adaptive estimator

In order to construct the efficient estimator, we assume the existence of the $\sqrt{n}$–consistent estimator $\{\tilde{\theta}_n\}$. Actually Chan and Tong (1985) proved that usual least–squares estimator $\{\overline{\theta}_n\}$ is $\sqrt{n}$–consistent for the SETAR(1) model. For technical reasons, we also use discretized version of the $\overline{\theta}_n$ which is formally defined as a point in $n^{-1/2}Z^{2p}$ closed to $\tilde{\theta}_n$. Furthermore we suppose $\dot{\varphi}(\cdot)$ satisfies the following conditions:

$$\text{(4.1)} \qquad \begin{aligned} &\text{(i)} \quad \lim_{h\to 0}\int [\varphi(x+h)-\varphi(x)]^2 f(x)dx = 0 \\ &\text{(ii)} \quad \lim_{h\to 0}\int h^{-1}[\varphi(x+h)-\varphi(x)]f(x)dx = -I(f)/2. \end{aligned}$$

Then we have the following lemma.

LEMMA 4.1. (Existence of efficient estimator for fixed f). *Assume $\{\overline{\theta}_n\}$ is a sequence of discrete $\sqrt{n}$–consistent estimators. Then $\tilde{\theta}_n$ defined below is asymptotically linear.*

$$\tilde{\theta}_n = \overline{\theta}_n + \hat{\Gamma}(\overline{\theta}_n)/I(f)\cdot\Delta_n(\overline{\theta}_n)/\sqrt{n}, \tag{4.2}$$

where $\hat{\Gamma}(\theta) = \sum_{j=1}^{n} X_1(j-1)X_1(j-1)^T/n$, $\Delta_n(\overline{\theta}_n) = \sum_{j=1}^{n}\dot{\varphi}(e_j(\overline{\theta}_n))\cdot(X_1(j-1)-\overline{X}_1)/\sqrt{n}$, *and* $\overline{X}_1 = \sum_{j=1}^{n} X_1(j)/n$. □

We note that the estimator $\tilde{\theta}_n$ depends on the nuisance parameter f. Thus natural question is whether it is possible to construct estimators, which has the same property as $\tilde{\theta}_n$, but does not depend on f. In order to obtain such an *adaptive* estimator, we have to estimate $\Delta(\overline{\theta}_n)$ and $I(f)$ consistently. Because of the lack of symmetry of $f(\cdot)$, we cannot use the same technique as given in So (1990) but will use the ideas given in Stone (1975) and Kreiss (1987) with appropriate changes to the non–linear situation considered here. First we introduce the following notations

$$\varphi(X;\sigma) = \exp(-x^2/2\sigma^2)/\sigma\sqrt{2\pi},\ x \in R$$

$$f_\sigma(x) = \int \varphi(x-y;\sigma)f(y)dy$$

$$= \int f(y-\sigma x)\varphi(y;1)dy, \tag{4.3}$$

where g is a density restricted to $[-1,1]$ with $g(x) \le g(0) = 1$ and g is continuous.

Then we can establish the following results.

THEOREM 4.2. *Let the sequence* $\{\theta_n\}$ *satisfy*

$$\sqrt{n}(\theta_n - \theta) = O(1). \tag{4.4}$$

Suppose $\sigma_n \to 0,\ c_n \to \infty,\ \int (f'/f)^4 f(x)dx < \infty$ *and* $\dot\varphi$ *is continuous a.e. and for* $k, K > 0$,

$$|\dot\varphi(x) - \dot\varphi(y)| \le K(1 + |x-y|^k).$$

Then we have

$$\Delta_n(\theta_n) - \sqrt{n}\int -\frac{f'_{\sigma_n}(x)}{f_{\sigma_n}(x)} \cdot \sum_{j=1}^{n}\{[\varphi(x - e_j(\theta_n);\sigma_n) - f_{\sigma_n}(x)]$$

$$(X_1(j-1) - \overline{X})\}g(x/c_n)dx/n = o_{p_{n,0}}(1), \tag{4.5}$$

where $e_j(\theta) = X_j - \theta^T X(j-1),\ P_{n,0} = P_n(0,0)$. □

and

THEOREM 4.3. *Let the sequence* $\{\theta_n\}$ *satisfy* (4.4) *and let*

$$\hat f_{\sigma_n}(x) = \tfrac{1}{n}\sum_{j=1}^{n}\varphi(x - e_j(\theta_n){:}\,\sigma_n). \tag{4.6}$$

Suppose $\sigma_n = o(1),\ c_n \to \infty,\ \delta > 0, c_n^2/(n^{1-\delta}\sigma_n^{4+\delta}) \to 0$ *and* $\int (f'/f)^4 f d\lambda < \infty$. *Then we have*

$$\sqrt{n}\int -\frac{\hat f'_{\sigma_n}(x)}{\hat f_{\sigma_n}(x)}\tfrac{1}{n}\Sigma\{[\varphi(x - e_j(\theta_n){:}\,\sigma_n) - \hat f_{\sigma_n}(x)]\}$$

$$(X_1(j-1) - \overline{X}_1)g(x/c_n)dx -$$

$$\sqrt{n}\int -\frac{f'_{\sigma_n}(x)}{f_{\sigma_n}(x)}\tfrac{1}{n}\Sigma\{[\varphi(x - e_j(\theta_n){:}\,\sigma_n) - f_{\sigma_n}(x)]\} \tag{4.7}$$

$$(X_1(j-1) - \overline{X}_1)g(x/c_n)dx$$

$$= o_{p_{n,0}}(1).$$

□

Above results imply the following main theorem of this paper immediately.

THEOREM 4.4 (existence of adaptive estimators). *Under the assumptions of Theorem 4.3 we have*

$$\hat{\theta}_n = \overline{\theta}_n + \hat{\Gamma}_n^{-1}/\hat{I}_n \ \hat{\Delta}(\overline{\theta}_n)/\sqrt{n} \tag{4.8}$$

is asymptotically linear and efficient. □

Here $\hat{\Delta}_n(\overline{\theta}_n)$ is the abbreviation for the l.h.s. in (4.7) and $\hat{I}_n$ denotes consistent estimator of $I(f)$.

Remark 4.1. We note that consistent estimator $\hat{I}_n$ of $I(f)$ are available. For example we can use estimator $\hat{I}_n = \sum_{j=1}^{n} \hat{q}_{nj}(e_j(\overline{\theta}_n);\ \overline{\theta}_n)/n$ where $\hat{q}_{nj}(e(\theta);\theta)$ is defined in So (1990).

Remark 4.2. Our construction follow very closely that of Kreiss (1987). But the proofs of the key substitution theorem 4.1, 4.2 for the non–linear time series model considered here are completely different and depend heavily on the non–linear property of the SETAR(1) model.

Remark 4.3. We also note that our results open the way for the construction of adaptive estimator for many other interesting classes of non–linear semiparametric time–series models; for example, we can get the exact same optimality results for the process $\{X_t\}$ defined by the difference equation;

$$X_t = \theta^T X(t-1) + T(X_{t-1}) + \epsilon_t,\ t \in Z,$$

where $T(\cdot)$ is an arbitrary bounded function on R.

5. Proofs

We start with the following auxiliary lemmas

LEMMA 5.1. *Let conditions* (i), (iii) *of* (1.2) *be satisfied. Suppose* $\theta_n = \theta + h_n/\sqrt{n}$ *and* h_n *is bounded. Then there exists constants* $M > 0,\ \rho < 1$ *such that*

$$\sup_{n\geq 1} E_n \|p^n(X_0|\cdot) - \pi_n(\cdot)\| \leq M\rho^n, \tag{5.1}$$

where $p^n(A|x) = p(X_n \in A|X_0 = x)$, $\pi_n(\cdot)$ *is the stationary initial distribution of the process* $\{X_t(\theta_n): t \in Z\}$ *and* $\|\cdot\|$ *is the variation norm of the signed measure.* □

PROOF. We first note that there exist $g(\cdot) \geq 0,\ k > 0,\ 0 < \delta < 1$ such that

$$\sup_n \sup_{|x|>k} E_n[g(y)|x] \leq (1-\delta)g(x). \tag{5.2}$$

See Chan and Tang (1985) for the details of the choices of $g(\cdot)$, δ and k. (5.2) in turn implies (5.1) by the same arguments in the proofs of theorem 2.3 and 2.5 of Nummelin and Tuominen (1982). □

LEMMA 5.2. *If* $\{X_t\}$ *satisfies conditions* (i) (iii) *of* (1.2), *then*

$$\max_{1\leq j\leq 1} |X(j)|/\sqrt{n} = o_{p_{n,0}}(1)$$

$$\sum_{j=1}^{n} X(j)X^T(j)/n = \Gamma \text{ nonsingular} \tag{5.3}$$

as $n \to \infty$. □

PROOF. The first property follows from $EX_t^2 < \infty$, which is implied by the theorem 2.3 of Chan and Tong (1985). The second follows from the ergodic theorem. □

LEMMA 5.3. *Let* $m: R \to R$ *be a square-integrable function,* $K = \int m^2(x)f(x)dx$. *Then*

$$E_n \|\frac{1}{\sqrt{n}} \sum_{j=1}^{n} m(e_j(\theta_n))[X(j-1) - \overline{X}]\|^2 = O(1)\cdot K, \tag{5.4}$$

with $\theta_n = \theta + h_n/\sqrt{n},\ h_n = O(1)$. □

PROOF. The left–hand side equals

$$O(1)\sum_{j=1}^{n} E_n\, m^2(e_j(\theta))/n$$

$$+\,2\sum_{j>\ell} E_n[m(e_j)m(e_\ell)(X(j-1)-\overline{X})(X(\ell-1)-\overline{X})]/n. \tag{5.5}$$

We note that first term in (5.5) is $K \cdot O(1)$ and the second term is bounded by

$$2\sum_{i<j} K^{1/2}|E_n m(e_i)(X(i-1)-\overline{X})(X(j-1)-\overline{X})|/n. \tag{5.6}$$

In order to get the bound for the (5.6), we now apply the corollary 2.5 in chap. 7 of Ethier and Kurtz (1986) to obtain the bound.

$$K^{1/2}\sum_{i<j} 2^3 \cdot \alpha^{1/u}(\mathcal{F}_{j-1}\|\mathcal{F}_i)\|m(e_i)(X(i-1)-\overline{X})\|_2\|X(j-1)-\overline{X}\|_{2+\delta} \tag{5.7}$$

where $1/u + 1/(2+\delta) + 1/2 = 1$ and $\alpha(\mathcal{F}_{j-1}\|\mathcal{F}_i) = 1/2 \cdot \sup_{A\in\mathcal{F}_{j-1}} \|p(A|\mathcal{F}_i) - p(A)\|_1$. By lemma 5.1 this in turn is bounded above by

$$2K\sum_{i<j} M p_u^{(j-i)}/n \cdot O(1) = KO(1). \tag{5.8}$$

This completes the proof. □

LEMMA 5.4. *If* $\{X_t\}$ *satisfied the condition* (i) *of* (1.2) *and* $\int x^{2+\delta} f(x)dx < \infty$ *for some* $\delta > 0$*, then*

$$\sum_{j=1}^{n}[X(j) - \overline{X}(j)]/\sqrt{n} = Op_{n,0}(1). \tag{5.9}$$

□

PROOF. We note that condition (i) of (1.2) implies *geometric ergodicity* of the process $\{X_t\}$ which in turn implies *strong mixing* property of $\{X_t\}$. Then we apply standard central limit theorem for stationary process to finish the proof. □

PROOF OF THEOREM 2.1.

It follows closely the corresponding proof of the theorem 2.1 in Kreiss (1987) with minor modifications. □

PROOF OF COROLLARY 2.2

We use the following orthogonality:

$$\Delta_n^*(\theta) \perp \sum_{j=1}^{n} \beta(e_j(\theta))/\sqrt{f(e_j)} \tag{5.10}$$

and Cramer–Wald device to obtain the desired result. See (3.1) in Kreiss (1987) for similar constructions in the linear AR model.

PROOF OF THEOREM 3.1

The assertion follows from essentially the same arguments as in the proof of Theorem 3.2 in Begun et al. (1983) by using LAN property of the process $\{X_t\}$. □

PROOF OF THEOREM 3.2

It follows from essentially the same proof as in the proof of theorem 3.1 in Begun et al. (1983) using LAN of the process $\{X_t\}$ and standard contiguity arguments. □

PROOF OF LEMMA 4.1

By using essentially the same arguments in the proof of theorem 2.1 in Beran (1976), we can show

$$\Delta_n^*(\theta_n) - \Delta_n^*(\theta) + \Gamma \cdot I(f)(\theta_n - \theta) = o_{p_{n,0}}(1). \tag{5.11}$$

Then we note that

$$\sqrt{n}(\overline{X}_1 - \mu) \cdot \sum_{j=1}^{n} \{\dot\varphi(e_j(\theta_n)) - \dot\varphi(e_j(\theta))\}/n = o_{p_{n,0}}(1) \tag{5.12}$$

because $\sqrt{n}(\overline{X}_1 - \mu) = Op_{n,0}(1)$ and

$$E_0 \sum_{j=1}^{n} \{\dot\varphi(e_j(\theta_n)) - \dot\varphi(e_j(\theta))/n\} \le E \int |\varphi(x + \Delta^T X(j)/\sqrt{n}) - \varphi(X)| f(x) dx = o(1). \tag{5.13}$$

Now (5.12), (5.13) together with discreteness of $\{\overline{\theta}_n\}$ finish the proof of the LEMMA. □

PROOF OF THEOREMS 4.1 and 4.2

We note that the proofs of theorem 4.1 and 4.2 in Kreiss (1987) depend only on the validity of the key lemmas 5.2, 5.3 and 5.4. This completes the proofs. □

REFERENCES

Begun, J. M., Hall, W. J., Huang, W. M., Wellner, J. A. (1983). Information and asymptotic efficiency in parametric – nonparametric models. *Ann. Statist.* **11**, 432–452.

Beran, R. (1976). Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **28**, 77–89.

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10**, 647–671.

Chan, K. S. and Tong, H. (1985). A multiple threshold AR(1) model. *J. Appl. Prob.* **22**, 267–279.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes; characterization and convergence.* New York, John Wiley.

Kreiss, J. P. (1986). On adaptive estimation in stationary ARMA-models. *Ann. Statist.* **15**, 112–134.

Kreiss, J. P. (1987). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5**, 59–76.

Nummelin, E., Tuominen, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with application to renewal theory. *Stoch. Proc. Appl.* **12**, 187–202.

So, B. S. (1990). On adaptive estimation in the semiparametric threshold AR(1) model. Tech. Rept. No. 90–44, Dept. of Statistics, Purdue University.

Stone, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3**, 267–284.

Tong, H. (1983). Threshold models in non-linear time-series analysis. Lecture notes in statistics. 21 Springer-Verlag, New York.