On Adaptive Estimation in the Semiparametric Threshold AR (1) Model

by

Beong-Soo So
Department of Statistics
Purdue University

Technical Report # 90–44

Department of Statistics
Purdue University

August 1990
Revised May 1991

△

ON ADAPTIVE ESTIMATION IN THE SEMIPARAMETRIC THRESHOLD AR (1) MODEL

BY BEONG-SOO SO

Department of Statistics
Purdue University

Abstract. We consider the problem of efficient estimation for the parameter θ of a non-linear threshold AR (1) model with unknown error density $f(\cdot)$. On the basis of the locally asymptotic normality (LAN) of the model, we introduce two optimality criteria which are based on the locally asymptotic minimax (LAM) bound and the convolution theorem respectively. Then we construct a sequence of adaptive estimators which is LAM and achieves the smallest possible covariance matrix among all regular estimators for a wide class of *symmetric* error density $f(\cdot)$.

Keywords. Non-linear threshold AR (1) model; locally asymptotic normal; locally asymptotic minimax; adaptive estimator; kernel estimator; convolution theorem.

1. INTRODUCTION

It seems generally agreed that the class of threshold time series models forms one useful class of non-linear time series models. The practical relevance of non-linear analysis of time series data seems to be widely recognized. See for example Tong (1983).

Recently Chan and Tong (1985) have discussed a simple first-order threshold AR (1) model which is defined by the following difference equation;

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

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

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

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

and $\{\epsilon_t\}$ is a sequence of white noises. They have obtained the necessary and sufficient condition for the *ergodicity* of the process $\{X_t\}$ and shown that least squares estimators of $\theta = \{\theta(i,k) : i = 0, 1, 1 \leq k \leq p\}$ are strongly consistent and asymptotically normal.

One relevant issue in this model is the efficient estimation of the parameter θ when the distribution of the error is regarded as unknown nuisance parameter. Most of previous works on this subject assume either linearity of the process $\{X_t\}$ as in Beran (1976), Kreiss (1986, 1987) or require normality of the error distribution for the efficiency of the least square estimator even though asymptotic results hold for quite general distributional shapes as is shown by Chan and Tong (1985).

In this paper we will introduce two alternative definitions of the asymptotic efficiency of the sequence of estimators of θ which are based on the *locally asymptotic minimax* (LAM) bound and the *convolution theorem* for the *regular* estimators respectively. Then we construct the sequence of adaptive estimator which is locally asymptotic minimax and achieve the smallest possible covariance matrix among all regular sequence of estimators for a wide class of *symmetric* error density $f(\cdot)$.

This paper is organized as follows:

In section 2 we prove the *local asymptotic normality* (LAN) of the process $\{X_t\}$ and introduce two asymptotic optimality criteria for the sequence of estimators of θ which are based on LAM bound and convolution theorem respectively.

In section 3 we construct a sequence of adaptive estimators which satisfies two optimality criteria simultaneously for a wide class of *symmetric* error density $f(\cdot)$.

Finally, in section 4 we discuss the possible generalization of the main results to other class of non-linear time series models. All technical proofs are given in Section 5.

2. LOCAL ASYMPTOTIC NORMALITY

First we assume the following regularity conditions on the process $\{X_t\}$.

A1: $\theta(1,1) < 1, \theta(p,1) < 1, \theta(1,1)\theta(p,1) < 1$.

A2: f is absolute continuous with finite Fisher information $I(f) = \int (f'/f)^2 f dx$.

A3: $f(x) > 0, \ \ \forall\, x \in R, \int x f(x) dx = 0, \int x^2 f(x) dx < \infty$.

REMARK 2.1. A1 and A3 are sufficient conditions for the *geometric ergodicity* of the process $\{X_t\}$ and A1 is almost necessary for the ergodicity. See Chan and Tong (1985) for the details.

We also need the following condition in order to ensure LAN of the process $\{X_t\}$.

A4: the density of the stationary initial distribution $g(x;\theta)$ satisfies

$$g(x;\theta_n) \to g(x;\theta) \text{ in } P_{\theta_0} \text{ prob.} \quad (2.1)$$
$$\text{if} \quad \theta_n \to \theta \text{ as } n \to \infty.$$

REMARK 2.2. In fact condition A4 seems to be implied by the condition A1, A2 and A3. See Chan and Tong (1986) for the proof of the similar continuity property of the initial density.

Now the density of the distribution of $(X_0, \ldots, X_n)$ can be expressed in the form

$$g(X_0;\theta) \prod_{i=1}^{n} f(e_i(\theta)), \quad (2.2)$$

where $e_t(\theta)$ is the residual calculated from (1.1). Thus we obtain the likelihood ratio

$$dP_{n,\theta}/dP_{n,\theta_0} = g(X_0;\theta_n)/g(X_0;\theta_0) \prod_{t=1}^{n} f(e_t - (\theta-\theta_0)^T X(t-1))/f(e_t), \tag{2.3}$$

where the abbreviations $X^T(i) = (I(X_i \in R_k), I(X_i \in R_k)X_i; k = 1,\ldots,p)$ and $e_t = e_t(\theta_0) = X_t - \theta_0^T X(t-1)$ are used. After these preliminaries we are now ready to establish *local asymptotic normality* (LAN) for the above likelihood ratio.

THEOREM 2.1. (Local Asymptotic Normality). *Let $h_n \in R^{2p}$ be a bounded sequence and $\theta_n = \theta_0 + h_n/\sqrt{n}$. Let conditions $A1, A2, A3$ and $A4$ be satisfied and let*

$$\Delta_n(\theta) = 2\sum_{j=1}^{n} \dot{\varphi}(e_j(\theta))X(j-1)/\sqrt{n}, \dot{\varphi} = f'/f. \tag{2.4}$$

Then we have

$$\log\left[dP_{n,\theta_n}/dP_{n,\theta_0}\right] - h_n^T\Delta_n(\theta_0) + \frac{1}{2}h_n^T I(f)\Gamma(\theta_0)h_n \to 0 \tag{2.5}$$

in P_{n,θ_0} probability,

and

$$\mathcal{L}(\Delta_n(\theta)|P_{n,\theta_0}) \Longrightarrow N(0, I(f)\Gamma(\theta_0)), \tag{2.6}$$

where $\Gamma(\theta_0) = E\left[X(j)X^T(j)\right]$ and "$\Rightarrow$" denotes weak convergence. $\square$

From the above theorem we can obtain the following result immediately.

COROLLARY 2.2. *Under the same assumptions as above*

$$\{P_{n,\theta_n}\} \text{ and } \{P_{n,\theta_0}\} \text{ are contiguous}$$

if h_n is bounded. $\square$

One immediate consequence of the above theorem is the following standard result on the *locally asymptotic minimax* (LAM) lower bound for the risk of estimators of θ.

THEOREM 2.3. (LAM bound for fixed $f(\cdot)$). *Suppose that $\ell(\cdot)$ is lower semicontinuous and subconvex. Then*

$$\lim_{k\to\infty} \liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in B(\theta_0,k/\sqrt{n})} \int \ell(\sqrt{n}(\hat{\theta}_n - \theta))dP_{n,\theta} \geq E\ell(Z), \tag{2.7}$$

where $B(\theta_0, k/\sqrt{n}) = \{\theta \in R^{2p} : |\theta - \theta_0| \leq k/\sqrt{n}\}, Z \sim N(0, (I(f)\Gamma)^{-1})$ *and* $|\cdot|$ *is any standard norm in* R^{2p}.

□

Above theorem suggests the following definition of the asymptotic efficiency of the sequence of estimators, namely

DEFINITION 2.1. Any sequence of estimators $\hat{\theta}_n$ which achieves the LAM lower bound (2.7) for any bounded continuous subconvex loss function $\ell(\cdot)$ will be called LAM-*efficient.*

In order to define alternative concept of efficiency we need to define the concept of *regular* sequence of estimators of θ.

DEFINITION 2.2. $\{T_n\}$ is said to be *regular* sequence of estimators if, for any sequence $\theta_n = \theta_0 + h_n/\sqrt{n},\ h_n = h + 0(1)$, we have

$$\mathcal{L}(\sqrt{n}(T_n - \theta_n)|P_{n,\theta_n}) \Longrightarrow \mathcal{L}(V) \tag{2.8}$$

and the limit distribution $\mathcal{L}(V)$ does not depend on the choice of sequences θ_n.

For regular sequence of estimators we have the following standard convolution theorem.

THEOREM 2.4 (Convolution theorem). *Let* $\{T_n\}$ *be a regular sequence of estimators. Then*

$$\mathcal{L}(\sqrt{n}(T_n - \theta_0)|P_{n,\theta_0}) \Longrightarrow \mathcal{L}(Z + V), \tag{2.9}$$

where $Z \sim N(0, (I(f)\Gamma)^{-1})$ *and* V *is a random variable on* R^{2p} *which is independent of* Z.

□

The above result suggests the following alternative definition of the asymptotic efficiency of the regular sequence of estimators.

DEFINITION 2.3. Regular sequence of estimators $\{T_n\}$ whose asymptotic distribution in (2.9) is $N(\underline{0}, (I(f)\Gamma)^{-1})$ is said to be *regular-efficient.*

Typically in order to show that a sequence of estimators $\{T_n\}$ is efficient in either sense, it is enough to show that they are *asymptotically linear*; namely

$$T_n = \theta_0 + [\Gamma^{-1}/I(f)]\Delta_n(\theta_0)/\sqrt{n} + 0(1/\sqrt{n}) \text{ in } P_{n,0} \text{ prob.,} \tag{2.10}$$

where $\Delta_n(\theta_0) = (1/\sqrt{n}) \sum_{j=1}^{n} \dot{\varphi}(e_j(\theta))X(j-1)$ is the *efficient* score function for the parameter θ.

REMARK 2.3. Regularity of the asymptotically linear sequence of estimators is an immediate consequence of the LAN property and the standard contiguity argument. See (3.3) of Kreiss (1987) for the similar result in the linear AR (p) model.

Standard method of constructing efficient estimators start with the existence of the preliminary $\sqrt{n}$-consistent estimators $\{\tilde{\theta}_n\}$ of θ. For technical reason, we also use discretized version $\overline{\theta}_n$ of the $\tilde{\theta}_n$ which is defined as a point in $n^{-\frac{1}{2}}Z^{2p}$ closest to $\tilde{\theta}_n$. In this paper we choose the usual least squares estimators of θ as an initial estimator $\tilde{\theta}_n$ which was shown to be $\sqrt{n}$-consistent by Chan and Tong (1985) under the conditions A1 and A3. Furthermore, we also assume the following regularity conditions for $\dot{\varphi}(\cdot)$:

A5: $\lim_{h\to 0} \int [\dot{\varphi}(x+h) - \dot{\varphi}(x)]^2 f(x)dx = 0$

A6: $\lim_{h\to 0} \int [\dot{\varphi}(x+h) - \dot{\varphi}(x))/h] f(x)dx = -I(f)/2$

Then we have

PROPOSITION 2.5 (Efficient estimator for fixed $f(\cdot)$). *Let $\{\overline{\theta}_n\}$ be a sequence of discrete $\sqrt{n}$-consistent estimators. Then $\hat{\theta}_n$ defined below is asymptotically linear and hence efficient.*

$$\hat{\theta}_n = \overline{\theta}_n + [\hat{\Gamma}_n^{-1}/I(f)]\Delta_n(\overline{\theta}_n)/\sqrt{n}, \tag{2.11}$$

where $\hat{\Gamma}_n = \sum_{j=1}^{n} X(j-1)X^T(j-1)/n.$ □

Since above estimator $\hat{\theta}_n$ depends on the unknown density $f(\cdot)$ of the white noise, one natural question is whether it is possible to construct estimators which are independent of the density $f(\cdot)$ of the white noise but are asymptotically linear simultaneously for a wide class of $f(\cdot)$. Such an estimator, if it exists, will be called *adaptive* estimator of θ for the given class of densities.

3. CONSTRUCTION OF ADAPTIVE ESTIMATES

In order to find the adaptive estimator, we first construct appropriate estimates of the score function $\dot{\varphi}(\cdot)$ and the Fisher information $I(f)$ and then show that the corresponding sequence of estimators is asymptotically linear for each density $f(\cdot)$ in the class. Our method of construction follows closely that of Schick (1986) in the semiparametric linear model. See Kreiss (1986) for similar but more complicated version. First we introduce following notations;

$$\text{(i)}\ k(x) = e^{-x}/(1+e^{-x})^2, x \in R \tag{3.1}$$

$$\text{(ii)}\ f_n(x) = \int f(x - a_n t)k(t)dt$$

$$\text{(iii)}\ \hat{f}_{nj}(x;\theta) = a_n + \sum_{i \neq j}^{n} k(x - e_i(\theta))/a_n)/na_n,$$

where $a_n = 0(1)$ and $e_i(\theta) = X_i - \theta^T X(i-1)$. Then we define $\hat{q}_{nj}$ to be the following estimator of $\dot{\varphi}(\cdot)$

$$\hat{q}_{nj}(x;\theta) = [\hat{f}'_{nj}(x)/\hat{f}_{nj}(x) - \hat{f}'_{nj}(-x)/\hat{f}_{nj}(-x)]/2, \tag{3.2}$$

where $\hat{f}_{nj}(x) = \hat{f}_{nj}(x;\theta)$ and $j = 1, \ldots, n$.

Let

$$\hat{\Delta}_n(\theta) = 2 \sum_{j=1}^{n} \hat{q}_{nj}(e_j(\theta);\theta)X(j-1)/\sqrt{n} \tag{3.3}$$

be the estimator of $\Delta_n(\theta)$ and let

$$\hat{\theta}_n = \overline{\theta}_n + [\hat{\Gamma}_n^{-1}/\hat{I}(\overline{\theta}_n)]\hat{\Delta}(\overline{\theta}_n)/\sqrt{n}, \tag{3.4}$$

where

$$\hat{I}_n = 4 \sum_{j=1}^{n} \hat{q}_{nj}(e_j(\overline{\theta}_n);\overline{\theta}_n)/n \tag{3.5}$$

and $\{\overline{\theta}_n\}$ is a sequence of discrete $\sqrt{n}$-consistent estimators.

We will prove that this estimator is adaptive for a wide class of *symmetric* densities $f(\cdot)$.

REMARK 3.1. As is noted by Kreiss (1986) in the ARMA model, our proof of the asymptotic linearity depends on the symmetry of $f(\cdot)$ crucially. The corresponding result for the non-symmetric but zero mean densities will be considered in a separate paper.

Now we can establish the following lemma.

LEMMA 3.1. *Let* $\{\overline{\theta}_n\}$ *be a sequence of discrete* $\sqrt{n}$*-consistent estimators of* θ*. Let* $a_n = 0(1)$ *and* $a_n^{-6}n^{-1} = 0(1)$ *as* $n \to \infty$*. Then*

$$\hat{\Delta}_n(\overline{\theta}_n) - \Delta_n(\overline{\theta}_n) = 0(1) in P_{n,\theta_0} prob. \tag{3.6}$$

□

REMARK 3.2. While Kreiss (1986) considered similar adaptive estimator in the ARMA model, his proof cannot be carried over to the *non-linear* time series model considered here. Our proof of (3.6) depends on the non-linear property of the process $\{X_t\}$ crucially.

As an immediate consequence of the above lemma, we have

THEOREM 3.2 (Existence of Adaptive estimators). *Let conditions* A1, A2, A3, A4, *and* A5 *be satisfied and the assumptions of Lemma 3.1 be satisfied. Then we have that the sequence of estimators*

$$\hat{\theta}_n = \overline{\theta}_n + \left[\hat{\Gamma}_n^{-1}/\hat{I}_n(\overline{\theta}_n)\right] \hat{\Delta}_n(\overline{\theta}_n)/\sqrt{n} \tag{3.7}$$

is asymptotically linear and

$$\mathcal{L}(\sqrt{n}(\hat{\theta}_n - \theta)|P_{n,\theta_0}) \Longrightarrow N(0, (\Gamma I(f))^{-1}) \tag{3.8}$$

for any symmetric $f(\cdot)$*. Therefore it is LAM and regular-efficient.* □

4. CONCLUDING REMARKS

There are several possible generalizations of the results in this paper to other class of semiparametric non-linear AR models. As is noted in Section 3, we may relax the symmetry requirement: $f(x) = f(-x)$ in order to obtain adaptive estimator for a wide class of densities $f(\cdot)$ with mean zero and finite variance. For technical reasons, we will pursue this possibility in a separate paper.

As a second important extension, we note that the proofs of the main results depend on some general non-linear properties of the process $\{X_t\}$ which can be satisfied in a variety of other non-linear time series models too. For example, we can prove the same optimality results for the process $\{X_t\}$ defined by

$$X_t = T(X_{t-1}) + \theta^T X(t-1) + \epsilon_T, t \in Z, \tag{4.1}$$

where $T(\cdot)$ is a bounded function on R. Similarly fussy extension of the model (4.1) considered by Chan and Tong (1986) provides another example of non-linear time series model which allows the same type of adaptive estimation. Both of these possibilities will be considered in a subsequent paper.

5. PROOFS

We start with the following auxiliary lemma.

LEMMA 5.1. *Under the conditions* A1 *and* A3, *there exist* $0 < \rho < 1, C_1, C_2 > 0$ *such that*

$$|X_i| \leq C_1 + C_2 \sum_{j=1}^{i} |e_j|\rho^{i-j} + \rho^i |X_0| \text{ for } i \in Z^+, \tag{5.1}$$

where $X_i = X_i(\theta_n) = \theta_n^T X(i-1) + \epsilon_i, \theta_n = \theta_0 + h_n/\sqrt{n}, h_n = 0(1)$ *as* $n \to \infty$. □

Proof of the Lemma. From the proof of the geometric ergodicity of the process $\{X_t\}$ in Chan and Tong (1985), we note that there exist a function $g(\cdot)$ such that

$$\text{(i) } g(T(x;\theta_n)) \leq (1-\delta)g(x), x \in R, \tag{5.2}$$

$$\text{(ii) } g(x+y) \leq g(x) + g(y), x, y \in R$$

$$\text{(iii) } a|x| \leq g(x) \leq b(1+|x|), x \in R,$$

where $T(x;\theta) = \theta^T X, 0 < \delta < 1, a, b > 0$. Then we note

$$\begin{aligned} g(X_i) &= g(T(X_i;\theta_n) + e_i) \\ &\leq (1-\delta)g(X_i) + g(e_i) \\ &\leq \sum_{j=0}^{i-1} (1-\delta)^j g(e_{i-j}) + g(X_0)(1-\delta)^i. \end{aligned} \tag{5.3}$$

This together with (iii) completes the proof.

□

Proof of Lemma 3.1. Let θ_n be any sequence such that $\theta_n = \theta_0 + h_n/\sqrt{n}, h_n = 0(1)$. Then by the discreteness of $\overline{\theta}_n$, it suffices to prove

$$\hat{\Delta}_n(\theta_n) - \Delta_n(\theta_n) = 0(1), \ in\ P_{n,\theta_n}\ prob. \tag{5.4}$$

We note that

$$\begin{aligned} & E_n||\hat{\Delta}_n(\theta_n) - \Delta_n(\theta_n)||^2 \\ & = 4 \sum_{j=1}^{n} E_n \left[||X(i-1)||^2 \int \left[\hat{q}_{nj}(x|\theta_n) - \dot{\varphi}(x)\right]^2 f(x)dx \right] /n \end{aligned} \tag{5.5}$$

by the symmetry of $\hat{q}_{nj}(\cdot)$ and $\dot{\varphi}(\cdot)$. Because of (5.1), (5.5) is bounded above by

$$1/n \sum_{i=1}^{n} \left[\sum_{j=1}^{i} \sum_{j=1}^{i} \rho^{i-j-1} \rho^{i-k-1} |\epsilon_j||\epsilon_k| \int (\hat{q}_{ni} - \dot{\varphi})^2 f(x)dx \right]. \tag{5.6}$$

Next we note that

$$\begin{aligned} & \sup_{j \le i} E_n|\epsilon_j| \int \left[\hat{q}_{ni}(x) - \dot{\varphi}(x)\right]^2 f(x)dx \\ & \le 2E_n|\epsilon_j| \int \left[\hat{q}_{n+j}(x) - \dot{\varphi}(x)\right]^2 f(x)dx \\ & + 2E_n|\epsilon_j| \int \left[\hat{q}_{ni}(x) - \hat{q}_{n-1}(x)\right]^2 f(x)dx \\ & \le 0(1)E_n \int \left[\hat{q}_{n-2}(x) - \dot{\varphi}(x)\right]^2 f(x)dx + 2\ 0(1)a_n^{-3} n^{-1} \\ & = 0(1) \end{aligned} \tag{5.7}$$

by using inequalities (3.15) and (3.16) of Schick (1987). Similarly we obtain

$$\sup_{k,j \le i} \ E_n \left[|\epsilon_j \epsilon_k| \int (\hat{q}_{ni}(x) - \dot{\varphi}(x))^2 f(x)dx \right] = 0(1). \tag{5.8}$$

(5.7) and (5.8) together with (5.6) show that

$$\text{l.h.s. of (5.4)} = 0(1) \text{ as } n \to \infty.$$

This completes the proof.

□

Proof of Theorem 2.1. The proof follows closely the corresponding proof of Theorem 3.1 in Kreiss (1986) with obvious modification. □

Proofs of Theorems 2.3 and 2.4. They follow from essentially the same arguments given in the proofs of Theorems 3.1 and 3.2 of Begun et.al. (1983) together with the LAN of the process $\{X_t\}$. □

Proofs of Proposition 2.5. By using the arguments given in the proof of Theorem 2.4 in Beran (1976), we can show that

$$\Delta_n(\theta_n) - \Delta_n(\theta) - (\Gamma I(f))^{-1}\sqrt{n}(\theta_n - \theta) = 0(1) \text{ } in \text{ } P_{n,\theta_0} \text{ } prob. \tag{5.9}$$

This fact together with discreteness of $\overline{\theta}_n$ completes the proof. □

Proof of Theorem 3.2. The assertion following directly from Theorem 2.4 and Lemma 3.1 by the discreteness of $\overline{\theta}_n$. □

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.

——— and ——— (1986) On estimating thresholds in autoregressive models. *J. Time Series Anal.* **7**(3), 179–190.

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

——— (1987) On adaptive Estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5**, 59–76.

Schick, A. (1987) A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference* **16**, 89–105.

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

On Adaptive Estimation in the
Semiparametric Threshold AR (1) Model

by

Beong-Soo So
Department of Statistics
Purdue University

Technical Report # 90–44

Department of Statistics
Purdue University

August 1990
Revised May 1991

△

ON ADAPTIVE ESTIMATION IN THE SEMIPARAMETRIC THRESHOLD AR (1) MODEL

BY BEONG-SOO SO

Department of Statistics
Purdue University

Abstract. We consider the problem of efficient estimation for the parameter θ of a non-linear threshold AR (1) model with unknown error density $f(\cdot)$. On the basis of the locally asymptotic normality (LAN) of the model, we introduce two optimality criteria which are based on the locally asymptotic minimax (LAM) bound and the convolution theorem respectively. Then we construct a sequence of adaptive estimators which is LAM and achieves the smallest possible covariance matrix among all regular estimators for a wide class of *symmetric* error density $f(\cdot)$.

Keywords. Non-linear threshold AR (1) model; locally asymptotic normal; locally asymptotic minimax; adaptive estimator; kernel estimator; convolution theorem.

1. INTRODUCTION

It seems generally agreed that the class of threshold time series models forms one useful class of non-linear time series models. The practical relevance of non-linear analysis of time series data seems to be widely recognized. See for example Tong (1983).

Recently Chan and Tong (1985) have discussed a simple first-order threshold AR (1) model which is defined by the following difference equation;

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

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

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

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

and $\{\epsilon_t\}$ is a sequence of white noises. They have obtained the necessary and sufficient condition for the *ergodicity* of the process $\{X_t\}$ and shown that least squares estimators of $\theta = \{\theta(i,k) : i = 0, 1, 1 \leq k \leq p\}$ are strongly consistent and asymptotically normal.

One relevant issue in this model is the efficient estimation of the parameter θ when the distribution of the error is regarded as unknown nuisance parameter. Most of previous works on this subject assume either linearity of the process $\{X_t\}$ as in Beran (1976), Kreiss (1986, 1987) or require normality of the error distribution for the efficiency of the least square estimator even though asymptotic results hold for quite general distributional shapes as is shown by Chan and Tong (1985).

In this paper we will introduce two alternative definitions of the asymptotic efficiency of the sequence of estimators of θ which are based on the *locally asymptotic minimax* (LAM) bound and the *convolution theorem* for the *regular* estimators respectively. Then we construct the sequence of adaptive estimator which is locally asymptotic minimax and achieve the smallest possible covariance matrix among all regular sequence of estimators for a wide class of *symmetric* error density $f(\cdot)$.

This paper is organized as follows:

In section 2 we prove the *local asymptotic normality* (LAN) of the process $\{X_t\}$ and introduce two asymptotic optimality criteria for the sequence of estimators of θ which are based on LAM bound and convolution theorem respectively.

In section 3 we construct a sequence of adaptive estimators which satisfies two optimality criteria simultaneously for a wide class of *symmetric* error density $f(\cdot)$.

Finally, in section 4 we discuss the possible generalization of the main results to other class of non-linear time series models. All technical proofs are given in Section 5.

2. LOCAL ASYMPTOTIC NORMALITY

First we assume the following regularity conditions on the process $\{X_t\}$.

A1: $\theta(1,1) < 1, \theta(p,1) < 1, \theta(1,1)\theta(p,1) < 1$.

A2: f is absolute continuous with finite Fisher information $I(f) = \int (f'/f)^2 f dx$.

A3: $f(x) > 0,\ \ \forall\, x \in R, \int x f(x) dx = 0, \int x^2 f(x) dx < \infty$.

REMARK 2.1. A1 and A3 are sufficient conditions for the *geometric ergodicity* of the process $\{X_t\}$ and A1 is almost necessary for the ergodicity. See Chan and Tong (1985) for the details.

We also need the following condition in order to ensure LAN of the process $\{X_t\}$.

A4: the density of the stationary initial distribution $g(x;\theta)$ satisfies

$$g(x;\theta_n) \to g(x;\theta) \text{ in } P_{\theta_0} \text{ prob.} \qquad (2.1)$$
$$\text{if} \quad \theta_n \to \theta \text{ as } n \to \infty.$$

REMARK 2.2. In fact condition A4 seems to be implied by the condition A1, A2 and A3. See Chan and Tong (1986) for the proof of the similar continuity property of the initial density.

Now the density of the distribution of $(X_0, \ldots, X_n)$ can be expressed in the form

$$g(X_0;\theta) \prod_{i=1}^{n} f(e_i(\theta)), \qquad (2.2)$$

where $e_t(\theta)$ is the residual calculated from (1.1). Thus we obtain the likelihood ratio

$$dP_{n,\theta}/dP_{n,\theta_0} = g(X_0;\theta_n)/g(X_0;\theta_0) \prod_{t=1}^{n} f(e_t - (\theta-\theta_0)^T X(t-1))/f(e_t), \tag{2.3}$$

where the abbreviations $X^T(i) = (I(X_i \in R_k), I(X_i \in R_k)X_i; k = 1,\ldots,p)$ and $e_t = e_t(\theta_0) = X_t - \theta_0^T X(t-1)$ are used. After these preliminaries we are now ready to establish *local asymptotic normality* (LAN) for the above likelihood ratio.

THEOREM 2.1. (Local Asymptotic Normality). *Let $h_n \in R^{2p}$ be a bounded sequence and $\theta_n = \theta_0 + h_n/\sqrt{n}$. Let conditions $A1, A2, A3$ and $A4$ be satisfied and let*

$$\Delta_n(\theta) = 2\sum_{j=1}^{n} \dot{\varphi}(e_j(\theta))X(j-1)/\sqrt{n}, \dot{\varphi} = f'/f. \tag{2.4}$$

Then we have

$$\log\left[dP_{n,\theta_n}/dP_{n,\theta_0}\right] - h_n^T\Delta_n(\theta_0) + \frac{1}{2}h_n^T I(f)\Gamma(\theta_0)h_n \to 0 \tag{2.5}$$

in P_{n,θ_0} probability,

and

$$\mathcal{L}(\Delta_n(\theta)|P_{n,\theta_0}) \Longrightarrow N(0, I(f)\Gamma(\theta_0)), \tag{2.6}$$

where $\Gamma(\theta_0) = E\left[X(j)X^T(j)\right]$ and "$\Rightarrow$" denotes weak convergence. □

From the above theorem we can obtain the following result immediately.

COROLLARY 2.2. *Under the same assumptions as above*

$$\{P_{n,\theta_n}\} \textit{and } \{P_{n,\theta_0}\} \textit{are contiguous}$$

if h_n is bounded. □

One immediate consequence of the above theorem is the following standard result on the *locally asymptotic minimax* (LAM) lower bound for the risk of estimators of θ.

THEOREM 2.3. (LAM bound for fixed $f(\cdot)$). *Suppose that $\ell(\cdot)$ is lower semicontinuous and subconvex. Then*

$$\lim_{k\to\infty} \liminf_{n\to\infty} \inf_{\hat{\theta}_n} \sup_{\theta\in B(\theta_0,k/\sqrt{n})} \int \ell(\sqrt{n}(\hat{\theta}_n - \theta))dP_{n,\theta} \geq E\ell(Z), \tag{2.7}$$

where $B(\theta_0, k/\sqrt{n}) = \{\theta \in R^{2p} : |\theta - \theta_0| \leq k/\sqrt{n}\}, Z \sim N(0, (I(f)\Gamma)^{-1})$ *and* $|\cdot|$ *is any standard norm in* R^{2p}. $\square$

Above theorem suggests the following definition of the asymptotic efficiency of the sequence of estimators, namely

DEFINITION 2.1. Any sequence of estimators $\hat{\theta}_n$ which achieves the LAM lower bound (2.7) for any bounded continuous subconvex loss function $\ell(\cdot)$ will be called LAM-*efficient.*

In order to define alternative concept of efficiency we need to define the concept of *regular* sequence of estimators of θ.

DEFINITION 2.2. $\{T_n\}$ is said to be *regular* sequence of estimators if, for any sequence $\theta_n = \theta_0 + h_n/\sqrt{n},\ h_n = h + 0(1)$, we have

$$\mathcal{L}(\sqrt{n}(T_n - \theta_n)|P_{n,\theta_n}) \Longrightarrow \mathcal{L}(V) \tag{2.8}$$

and the limit distribution $\mathcal{L}(V)$ does not depend on the choice of sequences θ_n.

For regular sequence of estimators we have the following standard convolution theorem.

THEOREM 2.4 (Convolution theorem). *Let* $\{T_n\}$ *be a regular sequence of estimators. Then*

$$\mathcal{L}(\sqrt{n}(T_n - \theta_0)|P_{n,\theta_0}) \Longrightarrow \mathcal{L}(Z + V), \tag{2.9}$$

where $Z \sim N(0, (I(f)\Gamma)^{-1})$ *and* V *is a random variable on* R^{2p} *which is independent of* Z. $\square$

The above result suggests the following alternative definition of the asymptotic efficiency of the regular sequence of estimators.

DEFINITION 2.3. Regular sequence of estimators $\{T_n\}$ whose asymptotic distribution in (2.9) is $N(\underline{0}, (I(f)\Gamma)^{-1})$ is said to be *regular-efficient.*

Typically in order to show that a sequence of estimators $\{T_n\}$ is efficient in either sense, it is enough to show that they are *asymptotically linear*; namely

$$T_n = \theta_0 + [\Gamma^{-1}/I(f)]\Delta_n(\theta_0)/\sqrt{n} + 0(1/\sqrt{n}) \text{ in } P_{n,0} \text{ prob.}, \tag{2.10}$$

where $\Delta_n(\theta_0) = (1/\sqrt{n}) \sum_{j=1}^{n} \dot{\varphi}(e_j(\theta))X(j-1)$ is the *efficient* score function for the parameter θ.

REMARK 2.3. Regularity of the asymptotically linear sequence of estimators is an immediate consequence of the LAN property and the standard contiguity argument. See (3.3) of Kreiss (1987) for the similar result in the linear AR (p) model.

Standard method of constructing efficient estimators start with the existence of the preliminary $\sqrt{n}$-consistent estimators $\{\tilde{\theta}_n\}$ of θ. For technical reason, we also use discretized version $\overline{\theta}_n$ of the $\tilde{\theta}_n$ which is defined as a point in $n^{-\frac{1}{2}}Z^{2p}$ closest to $\tilde{\theta}_n$. In this paper we choose the usual least squares estimators of θ as an initial estimator $\tilde{\theta}_n$ which was shown to be $\sqrt{n}$-consistent by Chan and Tong (1985) under the conditions A1 and A3. Furthermore, we also assume the following regularity conditions for $\dot{\varphi}(\cdot)$:

A5: $\lim_{h\to 0} \int[\dot{\varphi}(x+h) - \dot{\varphi}(x)]^2 f(x)dx = 0$

A6: $\lim_{h\to 0} \int[\dot{\varphi}(x+h) - \dot{\varphi}(x))/h]f(x)dx = -I(f)/2$

Then we have

PROPOSITION 2.5 (Efficient estimator for fixed $f(\cdot)$). *Let $\{\overline{\theta}_n\}$ be a sequence of discrete $\sqrt{n}$-consistent estimators. Then $\hat{\theta}_n$ defined below is asymptotically linear and hence efficient.*

$$\hat{\theta}_n = \overline{\theta}_n + [\hat{\Gamma}_n^{-1}/I(f)]\Delta_n(\overline{\theta}_n)/\sqrt{n}, \tag{2.11}$$

where $\hat{\Gamma}_n = \sum_{j=1}^{n} X(j-1)X^T(j-1)/n.$ □

Since above estimator $\hat{\theta}_n$ depends on the unknown density $f(\cdot)$ of the white noise, one natural question is whether it is possible to construct estimators which are independent of the density $f(\cdot)$ of the white noise but are asymptotically linear simultaneously for a wide class of $f(\cdot)$. Such an estimator, if it exists, will be called *adaptive* estimator of θ for the given class of densities.

3. CONSTRUCTION OF ADAPTIVE ESTIMATES

In order to find the adaptive estimator, we first construct appropriate estimates of the score function $\dot{\varphi}(\cdot)$ and the Fisher information $I(f)$ and then show that the corresponding sequence of estimators is asymptotically linear for each density $f(\cdot)$ in the class. Our method of construction follows closely that of Schick (1986) in the semiparametric linear model. See Kreiss (1986) for similar but more complicated version. First we introduce following notations;

$$\text{(i)}\ k(x) = e^{-x}/(1+e^{-x})^2, x \in R \tag{3.1}$$

$$\text{(ii)}\ f_n(x) = \int f(x - a_n t)k(t)dt$$

$$\text{(iii)}\ \hat{f}_{nj}(x;\theta) = a_n + \sum_{i \neq j}^{n} k(x - e_i(\theta))/a_n)/na_n,$$

where $a_n = 0(1)$ and $e_i(\theta) = X_i - \theta^T X(i-1)$. Then we define $\hat{q}_{nj}$ to be the following estimator of $\dot{\varphi}(\cdot)$

$$\hat{q}_{nj}(x;\theta) = [\hat{f}'_{nj}(x)/\hat{f}_{nj}(x) - \hat{f}'_{nj}(-x)/\hat{f}_{nj}(-x)]/2, \tag{3.2}$$

where $\hat{f}_{nj}(x) = \hat{f}_{nj}(x;\theta)$ and $j = 1, \ldots, n$.

Let

$$\hat{\Delta}_n(\theta) = 2 \sum_{j=1}^{n} \hat{q}_{nj}(e_j(\theta);\theta)X(j-1)/\sqrt{n} \tag{3.3}$$

be the estimator of $\Delta_n(\theta)$ and let

$$\hat{\theta}_n = \overline{\theta}_n + [\hat{\Gamma}_n^{-1}/\hat{I}(\overline{\theta}_n)]\hat{\Delta}(\overline{\theta}_n)/\sqrt{n}, \tag{3.4}$$

where

$$\hat{I}_n = 4 \sum_{j=1}^{n} \hat{q}_{nj}(e_j(\overline{\theta}_n);\overline{\theta}_n)/n \tag{3.5}$$

and $\{\overline{\theta}_n\}$ is a sequence of discrete $\sqrt{n}$-consistent estimators.

We will prove that this estimator is adaptive for a wide class of *symmetric* densities $f(\cdot)$.

REMARK 3.1. As is noted by Kreiss (1986) in the ARMA model, our proof of the asymptotic linearity depends on the symmetry of $f(\cdot)$ crucially. The corresponding result for the non-symmetric but zero mean densities will be considered in a separate paper.

Now we can establish the following lemma.

LEMMA 3.1. *Let* $\{\overline{\theta}_n\}$ *be a sequence of discrete* $\sqrt{n}$*-consistent estimators of* θ. *Let* $a_n = 0(1)$ *and* $a_n^{-6} n^{-1} = 0(1)$ *as* $n \to \infty$. *Then*

$$\hat{\Delta}_n(\overline{\theta}_n) - \Delta_n(\overline{\theta}_n) = 0(1) in P_{n,\theta_0} prob. \tag{3.6}$$

□

REMARK 3.2. While Kreiss (1986) considered similar adaptive estimator in the ARMA model, his proof cannot be carried over to the *non-linear* time series model considered here. Our proof of (3.6) depends on the non-linear property of the process $\{X_t\}$ crucially.

As an immediate consequence of the above lemma, we have

THEOREM 3.2 (Existence of Adaptive estimators). *Let conditions* A1, A2, A3, A4, *and* A5 *be satisfied and the assumptions of Lemma 3.1 be satisfied. Then we have that the sequence of estimators*

$$\hat{\theta}_n = \overline{\theta}_n + \left[\hat{\Gamma}_n^{-1} / \hat{I}_n(\overline{\theta}_n)\right] \hat{\Delta}_n(\overline{\theta}_n)/\sqrt{n} \tag{3.7}$$

is asymptotically linear and

$$\mathcal{L}(\sqrt{n}(\hat{\theta}_n - \theta) | P_{n,\theta_0}) \Longrightarrow N(0, (\Gamma I(f))^{-1}) \tag{3.8}$$

for any symmetric $f(\cdot)$. *Therefore it is LAM and regular-efficient.* □

4. CONCLUDING REMARKS

There are several possible generalizations of the results in this paper to other class of semiparametric non-linear AR models. As is noted in Section 3, we may relax the symmetry requirement: $f(x) = f(-x)$ in order to obtain adaptive estimator for a wide class of densities $f(\cdot)$ with mean zero and finite variance. For technical reasons, we will pursue this possibility in a separate paper.

As a second important extension, we note that the proofs of the main results depend on some general non-linear properties of the process $\{X_t\}$ which can be satisfied in a variety of other non-linear time series models too. For example, we can prove the same optimality results for the process $\{X_t\}$ defined by

$$X_t = T(X_{t-1}) + \theta^T X(t-1) + \epsilon_T, t \in Z, \tag{4.1}$$

where $T(\cdot)$ is a bounded function on R. Similarly fussy extension of the model (4.1) considered by Chan and Tong (1986) provides another example of non-linear time series model which allows the same type of adaptive estimation. Both of these possibilities will be considered in a subsequent paper.

5. PROOFS

We start with the following auxiliary lemma.

LEMMA 5.1. *Under the conditions* A1 *and* A3, *there exist* $0 < \rho < 1, C_1, C_2 > 0$ *such that*

$$|X_i| \leq C_1 + C_2 \sum_{j=1}^{i} |e_j|\rho^{i-j} + \rho^i |X_0| \text{ for } i \in Z^+, \tag{5.1}$$

where $X_i = X_i(\theta_n) = \theta_n^T X(i-1) + \epsilon_i, \theta_n = \theta_0 + h_n/\sqrt{n}, h_n = 0(1)$ *as* $n \to \infty$. □

Proof of the Lemma. From the proof of the geometric ergodicity of the process $\{X_t\}$ in Chan and Tong (1985), we note that there exist a function $g(\cdot)$ such that

$$\text{(i)}\ \ g(T(x;\theta_n)) \leq (1-\delta)g(x), x \in R, \tag{5.2}$$

$$\text{(ii)}\ \ g(x+y) \leq g(x) + g(y), x, y \in R$$

$$\text{(iii)}\ \ a|x| \leq g(x) \leq b(1+|x|), x \in R,$$

where $T(x;\theta) = \theta^T X, 0 < \delta < 1, a, b > 0$. Then we note

$$\begin{aligned} g(X_i) &= g(T(X_i;\theta_n) + e_i) \\ &\leq (1-\delta)g(X_i) + g(e_i) \\ &\leq \sum_{j=0}^{i-1} (1-\delta)^j g(e_{i-j}) + g(X_0)(1-\delta)^i. \end{aligned} \tag{5.3}$$

This together with (iii) completes the proof. $\square$

Proof of Lemma 3.1. Let θ_n be any sequence such that $\theta_n = \theta_0 + h_n/\sqrt{n}, h_n = 0(1)$. Then by the discreteness of $\overline{\theta}_n$, it suffices to prove

$$\hat{\Delta}_n(\theta_n) - \Delta_n(\theta_n) = 0(1), \ \textit{in } P_{n,\theta_n} \textit{ prob.} \tag{5.4}$$

We note that

$$\begin{aligned} &E_n||\hat{\Delta}_n(\theta_n) - \Delta_n(\theta_n)||^2 \\ &= 4\sum_{j=1}^{n} E_n\left[||X(i-1)||^2 \int \left[\hat{q}_{nj}(x|\theta_n) - \dot{\varphi}(x)\right]^2 f(x)dx\right]/n \end{aligned} \tag{5.5}$$

by the symmetry of $\hat{q}_{nj}(\cdot)$ and $\dot{\varphi}(\cdot)$. Because of (5.1), (5.5) is bounded above by

$$1/n \sum_{i=1}^{n}\left[\sum_{j=1}^{i}\sum_{j=1}^{i} \rho^{i-j-1}\rho^{i-k-1}|\epsilon_j||\epsilon_k| \int (\hat{q}_{ni} - \dot{\varphi})^2 f(x)dx\right]. \tag{5.6}$$

Next we note that

$$\begin{aligned} &\sup_{j\le i} E_n|\epsilon_j| \int \left[\hat{q}_{ni}(x) - \dot{\varphi}(x)\right]^2 f(x)dx \\ &\le 2E_n|\epsilon_j| \int \left[\hat{q}_{n+j}(x) - \dot{\varphi}(x)\right]^2 f(x)dx \\ &+ 2E_n|\epsilon_j| \int \left[\hat{q}_{ni}(x) - \hat{q}_{n-1}(x)\right]^2 f(x)dx \\ &\le 0(1)E_n \int \left[\hat{q}_{n-2}(x) - \dot{\varphi}(x)\right]^2 f(x)dx + 2\ 0(1)a_n^{-3}n^{-1} \\ &= 0(1) \end{aligned} \tag{5.7}$$

by using inequalities (3.15) and (3.16) of Schick (1987). Similarly we obtain

$$\sup_{k,j\le i} \ E_n\left[|\epsilon_j\epsilon_k| \int (\hat{q}_{ni}(x) - \dot{\varphi}(x))^2 f(x)dx\right] = 0(1). \tag{5.8}$$

(5.7) and (5.8) together with (5.6) show that

$$\text{l.h.s. of (5.4)} = 0(1) \text{ as } n \to \infty.$$

This completes the proof. $\square$

Proof of Theorem 2.1. The proof follows closely the corresponding proof of Theorem 3.1 in Kreiss (1986) with obvious modification. □

Proofs of Theorems 2.3 and 2.4. They follow from essentially the same arguments given in the proofs of Theorems 3.1 and 3.2 of Begun et.al. (1983) together with the LAN of the process $\{X_t\}$. □

Proofs of Proposition 2.5. By using the arguments given in the proof of Theorem 2.4 in Beran (1976), we can show that

$$\Delta_n(\theta_n) - \Delta_n(\theta) - (\Gamma I(f))^{-1}\sqrt{n}(\theta_n - \theta) = 0(1) \ in \ P_{n,\theta_0} \ prob. \tag{5.9}$$

This fact together with discreteness of $\overline{\theta}_n$ completes the proof. □

Proof of Theorem 3.2. The assertion following directly from Theorem 2.4 and Lemma 3.1 by the discreteness of $\overline{\theta}_n$. □

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.

______ and ______ (1986) On estimating thresholds in autoregressive models. *J. Time Series Anal.* **7**(3), 179–190.

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

______ (1987) On adaptive Estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5**, 59–76.

Schick, A. (1987) A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference* **16**, 89–105.

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