Journal of the Korean
Statistical Society
Vol. 24, No. 1, 1995

On the Optimal Adaptive Estimation in the Semiparametric Non-linear Autoregressive Time Series Model †

Beong-Soo So[1]

ABSTRACT

We consider the problem of optimal adaptive estimation of the euclidean parameter vector θ of the univariate *non-linear* autoregressive time series model $\{X_t\}$ which is defined by the following system of stochastic difference equations ; $X_t = \sum_{i=1}^{p} \theta_i \cdot T_i(X_{t-1}) + e_t$, $t = 1, \cdots, n$, where θ is the unknown parameter vector which describes the deterministic dynamics of the stochastic process $\{X_t\}$ and $\{e_t\}$ is the sequence of white noises with *unknown* density $f(\cdot)$. Under some general growth conditions on $T_i(\cdot)$ which guarantee ergodicity of the process, we construct a sequence of *adaptive* estimators which is locally asymptotic minimax (LAM) efficient and also attains the least possible covariance matrix among all regular estimators for arbitrary *symmetric* density.

KEYWORDS: Semiparametric adaptive estimation, Locally asymptotic minimax, Non-linear AR (1) model

† This research was supported by Korea Science Foundation Grant 923-0100-001-1.
[1] Dept. of Statistics, Ewha Womans University, Seodaemun, Seoul, 120-750, Korea

1. INTRODUCTION

Recently, motivated by the current interest in the non-linear phenomena in science and engineering, non-linear time series models are generally regarded to form one useful class of tools in the time series analysis. The practical relevance of non-linear analysis of time series data is now widely recognized in various applications. For example Tong (1983) provides many interesting examples of the non-linear time series models including well-known threshold time series models. See also Tong (1990) for more recent review on the topic.

In this paper we consider univariate *non-linear* time series model $\{X_t\}$ defined by the following system of 1-st order autoregressive stochastic difference equations :

$$X_t = \sum_{i=1}^{p} \theta_i \cdot T_i(X_{t-1}) + e_t, \; t = 1, 2, \cdots \tag{1.1}$$

where $\{T_i(\cdot)\}$ is the vector of basis functions and $\theta^T = (\theta_1, \cdots, \theta_p)$ is the vector of parameters identifying systematic component of the model and random variables $\{e_t\}$ are white noise series with *unknown* density function $f(\cdot)$. Above type of models is called *semiparametric* model in the statistical literature because it incorporates both the usual parametric component θ and the nonparametric component $f(\cdot)$ in the model. See Begun et. al.(1983) for more examples of semiparametric models. Such general classes of non-linear autoregressive (AR) processes include many types of non-linear time series models considered in the literature by the appropriate choice of basis functions $T_i(\cdot)$. For example we may choose *spline* functions as basis functions in order to get the natural generalization of the threshold models of Tong (1983). See chap 3 of Tong (1990) for more examples.

One of the important issues in this *semiparametric* model is the efficient estimation of the parameter vector θ when the density $f(\cdot)$ of the noise $\{e_t\}$ is regarded as *unknown* nuisance function . Previous works in this area were mostly concerned with inferences in the *linear* ARMA processes as in Beran(1976) and Kreiss (1986).

In this paper we will identify new kind of general growth conditions on $T_i(\cdot)$ which will guarantee not only the ergodicity of the process $\{X_t\}$ but also the existence of optimal adaptive estimators. Then we construct a sequence of optimal adaptive estimators which is LAM-efficient and attains the least possible covarince matrix among all regular estimators with respect to arbitrary *symmetric* density $f(\cdot)$.

This paper is organized as follows : In section 2 we first establish the local asymptotic normality (LAN) of the process $\{X_t\}$ and then introduce two asymptotic optimality criteria for the sequence of estimators of θ based on LAM bound and convolution theorem respectively. In section 3 we construct a sequence of semiparametric adaptive estimators which satisfies two optimality criteria simultaneously for a wide class of *symmetric* density functions $f(\cdot)$. Finally, in section 4 we discuss the possibility of extending 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 : For each $\theta \in \Theta$, Θ open parameter set in R^p, we have

$$\begin{aligned} &(\text{i})\quad |T(x;\theta)| \le (1-\delta)|x| + c,\ x \in R, \\ &(\text{ii})\quad \sup_{1\le i\le p} |T_i(x)| \le a + b|x|,\ x \in R \end{aligned} \tag{2.1}$$

where $T(x;\theta) = \sum_{i=1}^{p} \theta_i \cdot T_i(x)$, $0 < \delta < 1,\ a, b, c, > 0$.

A2 : $f(x)$ is an absolutely continuous function with finite positive Fisher information $I(f) = \int_{-\infty}^{+\infty} (f'/f)^2 f(x)dx > 0$.

A3 : $f(x) > 0, x \in R$, and $\int_{-\infty}^{+\infty} x^2 f(x)dx < \infty$.

A4 : The density of the stationary distribution $f(x;\theta)$ satisfies :

$$f(x;\theta_n) \to f(x;\theta),\ x \in R,\ \text{if } \theta_n \to \theta \text{ as } n \to \infty. \tag{2.2}$$

Remark 1. Conditions A1 and A3 are sufficient conditions for the geometric ergodicity of the process $\{X_t\}$ and the function $|\cdot|$ is often called *Lyapunov function* for the Markov process $\{X_t\}$. See Tong (1990) for more details on the application of the Lyapunov function in the non-linear time series model.

Remark 2. Simple sufficient conditions for A1 is the following :

$$|T_i(x)| \leq a_i + b_i|x|, x \in R \ , \ i = 1, \cdots, p \ \text{ and } \sum_{i=1}^{p} |\theta_i| b_i \leq (1 - \delta)$$

where $a_i, b_i > 0, \ 0 < \delta < 1$.

Now we note that the joint density of the simultaneous distribution of the data vector $(X_0, \cdots, X_n)$ can be expressed in the form :

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

where $e_t(\theta) = X_t - T(X_{t-1}; \theta)$ is the residual calculated from (1.1) . Therefore we can write the likelihood ratio of the random vector $(X_0, \cdots, X_n)$ as follows :

$$\frac{dP_n(\theta)}{dP_n(\theta_0)} = \frac{f(X_0; \theta)}{f(X_0; \theta_0)} \prod_{i=1}^{n} \frac{f(e_i - (\theta - \theta_0)^T X(i-1))}{f(e_i)} \tag{2.3}$$

where the abbreviations $X^T(t) = (T_1(X_t), \cdots, T_p(X_t))$ and $e_t = e_t(\theta_0)$ have been used and $P_n(\theta)$ denotes the probability measure of the random vector $(X_0, \cdots, X_n)$ on R^{n+1} when the true parameter vector is given by θ.

After these preliminaries we can now establish local asymptotic normality (LAN) property of the likelihood ratio of the process $\{X_t\}$.

Theorem 1. (Local Asymptotic Normality) Let $h_n \in R^p$ be a bounded sequence and $\theta_n = \theta_0 + h_n/\sqrt{n}$. Let the conditions A1, A2, A3 and A4 be satisfied and let

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

Then we have , as $n \to \infty$

$$log[dP_{n,\theta_n}/dP_{n,\theta_0}] = h_n^T \Delta_n(\theta_0) + (1/2) h_n^T I(f) \Gamma(\theta_0) h_n + o(1) \ \text{ in } P_{n,\theta_0} \text{ probability} \tag{2.5}$$

and

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

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

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

Corollary 1. Under the same assumptions as in Theorem 1,

$\{\mathrm{P}_{\mathrm{n},\theta_n}\}$ and $\{\mathrm{P}_{\mathrm{n},\theta_0}\}$ are contiguous in the sense of Roussas (1972)

if h_n is bounded.

Remark 3. See Roussas (1972) for more details on the applications of contiguity in statiatics.

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

Theorem 2. (LAM lower bound) Suppose that the loss fuction $l(\cdot)$ is lower semicontinuous and subconvex. Then

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

where $B(\theta_0, k/\sqrt{n}) = \{\theta \in R^p : |\theta - \theta_0| \leq k/\sqrt{n}\}, Z \sim N(0, (I(f)\Gamma)^{-1})$ and $|\cdot|$ is any standard norm in R^p . Here infimum is taken with respect to arbitrary estimators $\hat{\theta}_n$ of θ.

Above theorem suggests the following definition of the asymptotic efficiency of the sequence of the estimators.

Definition 1. A sequence of estimators $\{\hat{\theta}_n\}$ is called LAM-*efficient* if it attains the LAM lower bound (2.7) for any bounded continuous subconvex loss function $l(\cdot)$.

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

Definition 2. $\{T_n\}$ is said to be a *regular* sequence of estimators if , for any sequence $\theta_n = \theta_0 + h_n/\sqrt{n}, h_n = h + o(1)$, we have

$$\mathcal{L}(\sqrt{n}(T_n - \theta_n)|P_{n,\theta_n}) \Longrightarrow \mathcal{L}(U) \quad as\ n \to \infty \tag{2.8}$$

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

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

Theorem 3. (Convolution Theorem) Let $\{T_n\}$ be a regular sequence of estimators. Then

$$\mathcal{L}(\sqrt{n}(T_n - \theta_0)|P_{n.\theta_0}) \Longrightarrow \mathcal{L}(Z+V) \quad as\ n \to \infty \tag{2.9}$$

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

Now we introduce the following alternative definition of the asymptotic efficiency of the regular sequence of estimators.

Definition 3. Regular sequence of estimators $\{T_n\}$ is said to be *regular efficient* if the asymptotic distribution in (2.9) is $N(0, (I(f)\Gamma)^{-1})$.

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* ;

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

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

Remark 4. Regularity of the asymptotically linear sequence of estimators is immediate consequence of the LAN property and the standard contiguity argument. See (3.3) of Kreiss (1986) for the similar result in the linear AR(p) model.

Standard method of constructing efficient estimators begins with the existence of the preliminary $\sqrt{n}$-consistent estimator $\{\tilde{\theta}_n\}$ of θ. For technical reason, we also use discretized version $\bar{\theta}_n$ of the $\tilde{\theta}_n$ which is formally defined as a point in $n^{-1/2}Z^p$ closest to $\tilde{\theta}_n$. Here Z denotes the set of all integers. In this paper we choose the usual *least squares* estimator of θ as an initial estimator $\{\tilde{\theta}_n\}$ which was shown to be $\sqrt{n}$-consistent by Klimko and Nelson (1978) under the general conditions including A1, A2 and A3. Furthermore, we also assume the following additional regularity conditions for the score function $\psi(\cdot) = f'/f(\cdot)$:

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

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

Then we have the following result on the efficient estimator for fixed $f(\cdot)$.

Proposition 1. (Efficient estimator for fixed $f(\cdot)$) Let $\{\bar{\theta}_n\}$ be a sequence of discrete $\sqrt{n}$-consistent estimators. Then the sequence of estimators $\{\hat{\theta}_n\}$ defined below is asymptotically linear and efficient.

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

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 a sequence of estimators which is independent of the density $f(\cdot)$ of the noise but is asymptotically linear simultaneously for a wide class of densities $f(\cdot)$. Such an estimator, if it exists will be called optimal *adaptive* estimator of θ for the given class of densities.

3. OPTIMAL ADAPTIVE ESTIMATES

In order to find the optimal *adaptive* estimator, we first construct appropriate estimates of the score function $\psi(\cdot)$ and the Fisher information $I(f)$ and then show that the corresponding sequence of estimators is asymptotically linear for each of the *symmetric* density $f(\cdot)$. Our method of construction follows closely that of Schick (1987) in the semiparametric linear regression model. See Kreiss (1986) for the similar but more complicated version. First we introduce following notations ;

$$\begin{aligned}
&\text{(i)}\ \ k(x) = e^{-x}/(1+e^{-x})^2, x \in R \\
&\text{(ii)}\ \ f_n(x) = \int f(x - a_n t)k(t)dt \\
&\text{(iii)}\ \ \hat{f}_n(x;\theta) = a_n + \sum_{i=1}^{n} k((x - e_i(\theta))/a_n)/na_n \\
&\text{(iv)}\ \ \hat{f}_{nj}(x;\theta) = \hat{f}_n - k((x - e_j(\theta))/a_n)/na_n \\
&\text{(v)}\ \ \hat{f}_{njk}(x;\theta) = \hat{f}_{nj} - k((x - e_k(\theta))/a_n)/na_n
\end{aligned} \tag{3.1}$$

where $a_n = o(1)$ and $e_i(\theta) = X_i - \theta^T X(i-1)$. Then we define $\hat{q}_{nj}$ to be the *anti-symmetrized* esimator of $\psi(\cdot)$ given by ;

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

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

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 then let us define the estimator ;

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

with

$$\hat{I}_n(\bar{\theta}_n) = 4\sum_{j=1}^{n} \hat{q}_{jn}(e_j(\bar{\theta}_n);\bar{\theta}_n)/n, \quad \hat{\Gamma}_n = \sum_{j=1}^{n} X(j)X^T(j)/n \tag{3.5}$$

where $\{\bar{\theta}_n\}$ is a sequence of discrete $\sqrt{n}$-consistent estimators of θ.

We will prove that estimator (3.4) is optimal *adaptive* for a wide class of *symmetric* densities $f(\cdot)$. Now we first establish the following auxiliary lemma.

Lemma 1. Under the condition (i) of A1, there exist constants $0 < \rho < 1, C_1, C_2 > 0$ such that

$$|X_i| \leq C_1 + C_2 \sum_{j=1}^{i} \rho^{i-j}|e_j| + \rho^i |X_0| \text{ for } i \in Z^+. \tag{3.6}$$

Now we are ready to establish following important proposition.

Proposition 2. Let $\{\bar{\theta}_n\}$ be a sequence of discrete $\sqrt{n}$-consistent estimators of θ. Let $a_n = o(1)$ and $a_n^{-6}n^{-1} = o(1)$ as $n \to \infty$. Then

$$\hat{\Delta}_n(\bar{\theta}_n) - \Delta_n(\bar{\theta}_n) = o(1) \text{ in } P_{n,\theta_0} \text{ probability.} \tag{3.7}$$

Remark 5. While Kreiss (1986) considered similar adaptive estimator in the ARMA model, his proof cannot be carried over to the *non-linear* time series models considered in this paper. Our proof of (3.7) depends heavily on the existence of the appropriate *Lyapunov function* which reflects intrinsically non-linear property of the process $\{X_t\}$.

As an immediate consequence of the above proposition, we can establish the optimality of the adaptive estimator $\{\hat{\theta}_n^*\}$.

Theorem 4. (Optimal Adaptive Estimator) Let the conditions A1, A2, A3, A4 A5 and A6 be satisfied and the assumptions of Proposition 2 be satisfied. Then the sequence of estimators

$$\hat{\theta}_n^* = \bar{\theta}_n + [\hat{\Gamma}_n^{-1}/\hat{I}_n(\bar{\theta}_n)]\hat{\Delta}_n(\bar{\theta}_n)/\sqrt{n} \tag{3.8}$$

is asymptotically linear and as $n \to \infty$

$$\mathcal{L}(\sqrt{n}(\hat{\theta}_n^* - \theta)|P_{n,\theta_0}) \Longrightarrow N(0, [\Gamma I(f)]^{-1}) \tag{3.9}$$

for any symmetric density $f(\cdot)$. Therefore $\{\hat{\theta}_n^*\}$ is LAM and regular-efficient.

4. DISCUSSIONS

There are several possible generalizations of the results of this paper to other class of semiparametric non-linear AR models. First we may relax the symmetry requirement $f(x) = f(-x)$ and obtain adaptive estimator for a much wider class of non-symmetric densities $f(\cdot)$ with zero mean and finite variance with additional technical tools . As a second important extension, we note that the proofs of the main results of this paper does depend on some general non-linear properties of the process $\{X_t\}$ which can be satisfied in a variety of non-linear time series models. For example, we can obtain essentially the same optimality results for the process $\{X_t\}$ defined by ;

$$X_t = \sum_{i=1}^{p} \theta_i T_i(X_{t-1}; \phi) + e_t, \quad t \in Z \tag{4.1}$$

where $T_i(\cdot; \phi)$ is a function on R which depends on the extra parameters ϕ and satisfies similar growth condition as A1. Fussy extension of the familiar threshold model considered by Chan and Tong (1986) provides typical example of non-linear model of the type (4.1) which allows the same type of adaptive estimation. Both of these possibilities will be treated in a subsequent paper.

5. PROOFS

Proof of Theorem 1. The proof follows closely the similar proof of Theorem 3.1 of Kreiss (1986) with minor modifications.

Proofs of Theorem 2. and 3. Both follow directly from LAN property of the process $\{X_t\}$ by the standard arguments as are given in the proofs of Theorems 3.1 and 3.2 of Begun et. al. (1983).

Proof of Proposition 1. Following essentially the same argument as is given in the proof of Theorem 2.4 of Beran (1976), we have

$$\Delta_n(\theta_n) = \Delta_n(\theta) + [\Gamma I(f)]^{-1}\sqrt{n}(\theta_n - \theta) + o(1) \text{ in } P_{n,\theta_0} \text{ probability.} \tag{5.1}$$

Above identity together with discreteness of $\bar{\theta}_n$ completes the proof.

Proof of Lemma 1. By the condition (i) of A1, we have the inequality :

$$|T(x; \theta_n)| \leq (1-\delta)|x| + c, \; x \in R \tag{5.2}$$

where $T(x;\theta) = \sum_{i=1}^{p} \theta_i T_i(x),\ 0 < \delta < 1, a, b > 0$. Then we note

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

Above inequality together with $\rho = 1 - \delta$ finishes the proof.

Proof of Proposition 2. Let θ_n be any sequence such that $\theta_n = \theta_0 + h_n/\sqrt{n}$, $h_n = o(1)$. Then from the discreteness of $\bar{\theta}_n$, it suffices to prove

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

First we note that

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

by the symmetry of $\hat{q}_{ni}(\cdot)$ and $\psi(\cdot)$. Now Lemma 1 and the condition (ii) of A1 imply that (5.5) is bounded above by

$$1/n\sum_{i=1}^{n} E_n\left[\sum_{j=1}^{i}\sum_{k=1}^{i}\rho^{i-j-1}\rho^{i-k-1}|e_j||e_k|\int(\hat{q}_{ni} - \psi)^2 f(x)dx\right] \tag{5.6}$$

where $\rho = 1 - \delta$. Next we note that

$$\begin{aligned} &\sup_{j\leq i} E_n\left[|e_j|^2 \int [\hat{q}_{ni}(x) - \psi(x)]^2 f(x)dx\right] \\ &\leq 2E_n\left[|e_j|^2 \int [\hat{q}_{nij}(x) - \psi(x)]^2 f(x)dx\right] \\ &+2E_n\left[|e_j|^2 \int [\hat{q}_{ni}(x) - \hat{q}_{nij}(x)]^2 f(x)dx\right] \\ &\leq O(1)E_n\left[\int [\hat{q}_{n-2}(x) - \psi(x)]^2 f(x)dx\right] + 2O(1)a_n^{-3}n^{-1} \\ &= o(1) \end{aligned} \tag{5.7}$$

from the inequalities (3.15) and (3.16) of Schick (1987) where $\hat{q}_{nij}(\cdot)$ and $\hat{q}_n(\cdot)$ are the estimators of $\psi(\cdot)$ based on $\hat{f}_{nij}$ and $\hat{f}_n$ respectively. By the similar argument, we obtain

$$\sup_{k,j\leq i} E_n \left[|e_j e_k| \int [\hat{q}_{ni}(x) - \psi(x)]^2 f(x) dx \right] = o(1). \tag{5.8}$$

Now (5.7) and (5.8) together with (5.6) imply that left hand side of the (5.4) is o(1) as $n \to \infty$. This completes the proof.

Proof of Theorem 4. The proof follows directly from Proposition 1 and Proposition 2 by the discreteness of $\bar{\theta}_n$.

REFERENCES

(1) Begun,J.M., Hall, W.J., Huang,W.M., Wellner,J.A. (1983). Information and asymptotic efficiency in parametric-nonparametric models, *The Annals of Statistics*, **11**, 432-452.

(2) Beran, R. (1976). Adaptive estimates for autoregressive processes, *The Annals of Institute of Statistical Mathematics*, **28**, 77-89.

(3) Chan, K.S. and Tong, H. (1986). On estimating thresholds in autoregressive models, *Journal of Time Series Analysis*, **7(3)**, 179-190.

(4) Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes, *The Annals of Statistics*, **6**, 629-642.

(5) Kreiss, J.P. (1986). On adaptive estimation in stationary ARMA-models, *The Annals of statistics*, **15**, 112- 134.

(6) Roussas, G. (1972). *Contiguity of Probability Measures ; Some Applications in Statistics* , Cambridge University Press, Cambridge.

(7) Schick, A. (1987). A note on the construction of asymptotically linear estimators, *Journal of Statistical Planning and Inference*, **16**, 89-105.

(8) Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*, Lecture notes in Statistics. 21 Springer-Verlag, New York.

(9) Tong, H. (1990). *Non-linear Time Series*, Clarendon Press Oxford, New York.