Efficient Estimation in Smooth Threshold Autoregressive(1) Models

D. Nur, *School of Mathematical and Physical Sciences,*
University of Newcastle, NSW, Australia.
Email: Darfiana.Nur@newcastle.edu.au

G.M. Nair, *Department of Mathematics and Statistics,*
Curtin University of Technology, Perth, WA, Australia.
Email: G.Nair@curtin.edu.au

N.D. Yatawara, *Department of Mathematics and Statistics,*
Curtin University of Technology, Perth, WA, Australia.
Email: N.Yatawara@curtin.edu.au

Abstract

Verifiable conditions are given for the existence of efficient estimation in Smooth Threshold Autoregressive models of order 1. The paper establishes local asymptotic normality in the semi-parametric setting which is then used to discuss adaptive and efficient estimates of the models. It is found that the adaptation is satisfied if the error densities are symmetric. Simulation results are presented to compare the conditional least squares estimate with the adaptive and efficient estimates for the models.

AMS Subject Classification: 62M10, 62F10

Key-words: Adaptive estimation, efficient estimation, locally asymptotically normal, non-linear time series, smooth threshold autoregressive, stationarity.

1. Introduction

Consider the first-order *Smooth Threshold Autoregressive* (STAR) semiparametric model, with delay parameter one, defined by

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-1} G\left(\frac{X_{t-1} - r}{z}\right) + \varepsilon_t, \tag{1.1}$$

where $r \in \Re$ is a threshold parameter, $z \in \Re^+$ is a smoothing parameter, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with ε_t independent of X_s, $s < t$, and G is a distribution function. Let the distribution of ε_t belong to $\mathcal{F}$, a class of Lebesgue

densities. We assume that G, r and z are known.

This model was introduced by Chan and Tong (1986) as an extension of the SETAR model. One of the applications of this model is in economics as discussed in a survey of recent developments of this models by Dijk, *et al* (2002). This model allows for regime-switching behaviour which is very beneficial for modelling economic time series.

In this paper we consider the problem of construction of an adaptive and efficient estimator of $\theta = (\theta_1, \theta_2)^T$ in the presence of the nuisance parameter ϕ, the unknown density of the innovations, ε_t, in the above model.

General theory on the construction of estimators of an Euclidean parameter θ that are asymptotically efficient, in the presence of infinite dimensional nuisance parameters ϕ, has been the focus of numerous researchers in the last three decades. To study what is best possible asymptotically, one needs a bound on the asymptotic performance of estimators of θ (Hájek (1970), Fabian and Hannan (1982)). Hájek(1970) established a lower bound for the local asymptotic minimax risk of a sequence of estimators for Locally Asymptotically Normal(LAN) models. A large majority of models are LAN, and when this holds the Hájek-Le Cam convolution theorem yields an appropriate lower bound for the variance of the estimator. On an ad hoc basis, it is often possible to find estimators of θ that have the right rate of consistency. Typically, such estimators may be used to construct efficient estimators which attain the bound of the convolution theorem. If this bound is the same as in the parametric model with ϕ known, then such estimators are called *adaptive.*

Adaptive estimation in nonlinear time series models is considered by Linton (1993), Koul and Schick (1996), Drost and Klaassen (1997) and Drost, Klaassen and Werker (1997). Koul and Schick (1997) also discussed efficient estimation for a class of nonlinear time series models with unknown error densities. They gave several methods for constructing efficient estimates and these results were then applied for SETAR, EXPAR and ARMA models. It is shown that adaptation is not possible in SETAR models with asymmetric error densities. They also considered the construction of adaptive estimators in time series models with time varying location. Eventhough this is less general than the models considered by Drost, *et.al* (1997), they obtained efficient estimates, with general error models, that are automatically adaptive under the symmetry conditions.

This paper addresses, as in Koul and Schick (1997), adaptive and efficient estimation problems for the STAR model. STAR model is a more general than SETAR model, and the adaptive and efficiency results for SETAR model do not imply the corresponding results for STAR model. The conditions given in Koul and Schick (1997) were used to derive our main results as presented in Propositions 2.1 and 2.2, Lemma 2.1, Theorems 2.1, 3.1 and 3.2 respectively.

The paper is organised as follows. Section 2 discusses local asymptotic normality of the above STAR semiparametric model. Section 3 addresses the question of adaptive estimation of θ, where the necessary condition for adaptive estimation, given in Koul and Schick (1997) is used to prove that the STAR model is adaptive followed by the construction of an efficient estimator. In Section 4, two examples are presented to compare the conditional least squares estimate with the adaptive and efficient estimates for the STAR model. Finally, the conclusion is given in Section 5. This paper uses much of the notations, definitions and results given in Koul and Schick (1997).

2. Local Asymptotic Normality

By redefining G, the model (1.1) can be rewritten as

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-1} G(X_{t-1}) + \varepsilon_t. \tag{2.1}$$

In this section we show that the above model is LAN using the sufficient conditions given in Theorem 2.4 of Koul and Schick (1997). First we set up the required terminology. Let $\mathcal{F} = \mathcal{F}_0^+$ be the set of all positive Lebesgue densities with zero-mean, finite variances and finite Fisher information for location, and let

$$\Theta = \{\theta \in \Re^2 : \theta_1 < 1, \theta_1 + \theta_2 < 1, \theta_1(\theta_1 + \theta_2) < 1\}. \tag{2.2}$$

By Theorem 2.2 (Nur, 1998), (2.2) is the sufficient condition for (2.1) to be ergodic and there is a unique stationary process satisfying the model (see Proposition 2.1 of Chan and Tong (1986)).

For $j \geq 0$, let

$$H_j(\theta) = \theta_1 X_{j-1} + \theta_2 X_{j-1} G(X_{j-1}), \tag{2.3}$$

and define the partial derivative vector, $\dot{H}_j(\theta)$, by

$$\dot{H}_j(\theta) = \left(\frac{\partial}{\partial \theta} H_j(\theta) \right) = \left(\begin{array}{c} X_{j-1} \\ X_{j-1} G(X_{j-1}) \end{array} \right). \tag{2.4}$$

Let $\mathcal{P} = \{P_{\theta,\phi} : (\theta, \phi) \in \Theta \times \mathcal{F}\}$ be a family of probability measures. Under each $P_{\theta,\phi} \in \mathcal{P}$, we assume that the random vector $\mathbf{X}_0$ has a Lebesgue density $g_{\theta,\phi}$, and the random variables

$$\varepsilon_j(\theta) = X_j - H_j(\theta), \quad j = 1, 2, \ldots,$$

are independent with common density ϕ and are independent of $\mathbf{X}_0$.

Let $\theta_0 = (\theta_{01}, \theta_{02})$ and ϕ_0 be the true parameter values of θ and ϕ respectively. We fix these values throughout the paper. A sequence $\{\theta_n\}$ in Θ such that $\sqrt{n}(\theta_n - \theta_0)$ is bounded, is called a local sequence. For a local sequence $\{\theta_n\}$, a sequence $\{a_n\}$ of positive numbers and a sequence of random variables $\{Y_n\}$ we write $Y_n = o_{\theta_n}(a_n)(O_{\theta_n}(a_n))$ denoting a sequence of random variables $\{Y_n\}$ such that $a_n^{-1} Y_n$ converges to 0 (is bounded) in P_{θ_n}-probability.

The following provides sufficient assumptions, given by Koul and Schick (1997), for a nonlinear time series to be LAN.

Assumption 2.1. *The density ϕ_0 has finite Fisher information for location, i.e. ϕ_0 is absolutely continuous with a.e-derivative $\phi_0^{(1)}$ and*

$$J = \int l^2 dF < \infty, \quad \text{where} \quad l = -\frac{\phi_0^{(1)}}{\phi_0}. \tag{2.5}$$

Moreover,

$$\int |g_{\theta,\phi}(\mathbf{x}) - g_{\theta_0,\phi}(\mathbf{x})| d\mathbf{x} \to 0, \quad \text{as } \theta \to \theta_0, \tag{2.6}$$

where $g_{\theta,\phi}$ is a Lebesgue density of $\mathbf{X}_0$ under $P_{\theta,\phi} \in \mathcal{P}$.

Assumption 2.2. *There exists a $v \in \Re^2$, a positive definite 2×2 matrix M and measurable functions h_j from $\Re^j \times \Theta$ to $\Re^2$, $j = 1, 2, \ldots$ such that for all local sequences $\langle \vartheta_n \rangle$ and $\langle \theta_n \rangle$*

$$\sum_{j=1}^{n} |H_j(\vartheta_n) - H_j(\theta_n) - (\vartheta_n - \theta_n)^T \dot{H}_j(\theta_n)|^2 = o_{\theta_n}(1), \tag{2.7}$$

$$\max_{1 \le j \le n} \frac{1}{\sqrt{n}} \|\dot{H}_j(\theta_n)\| = o_{\theta_n}(1), \tag{2.8}$$

$$\frac{1}{n} \sum_{j=1}^{n} \dot{H}_j(\theta_n) = v + o_{\theta_n}(1), \tag{2.9}$$

$$\frac{1}{n} \sum_{j=1}^{n} \dot{H}_j(\theta_n)\dot{H}_j^T(\theta_n) = M + o_{\theta_n}(1), \tag{2.10}$$

where $\dot{H}_j(\theta) = \dot{h}_j(X_j, \theta)$ for $j = 1, 2, \ldots$ and $\theta \in \Theta$.

The following definition, taken from Koul and Schick (1997), gives a parametrization of the error density.

Definition 2.1. *A mapping $\eta \mapsto f_\eta$ from a neighbourhood Δ of the origin in $\Re^s$ into $\mathcal{F}$ such that $f_0 = \phi_0$ is called an s-dimensional path. Let ξ be a measurable function from $\Re$ to $\Re^s$ such that $\int \|\xi\|^2 dF < \infty$, $\int \xi \xi^T dF$ is nonsingular. The path $\eta \mapsto f_\eta$ is said to be ξ-smooth if*

$$\int \left(\sqrt{f_\eta(x)} - \sqrt{\phi_0(x)} - \frac{1}{2}\eta^T \xi(x)\sqrt{\phi_0(x)} \right)^2 dx = o(\|\eta\|^2). \tag{2.11}$$

The path $\eta \mapsto f_\eta$ is said to be ξ-regular if it is ξ-smooth and if

$$\int |g_{\theta,f_\eta}(\mathbf{x}) - g_{\theta_0,\phi_0}(\mathbf{x})| d\mathbf{x} \to 0, \quad \text{as} \quad \theta \to \theta_0 \text{ and } \eta \to 0. \tag{2.12}$$

By Theorem 2.4 in Koul and Schick (1997), in order to prove that the model (2.1) is LAN, we need to show that Assumptions 2.1 and 2.2 hold, the smooth path $\eta \mapsto f_\eta$ is ξ-regular and the existence of a positive-definite variance covariance matrix $V(\xi)$. In Proposition 2.1 we prove that Assumption 2.2 holds. The Assumption 2.1 and the regularity of the path are proved using some sufficient conditions given by the above authors. These sufficient conditions are given in the Proposition 2.2. Verification of these sufficient conditions, in terms of a measurable function ψ, is detailed in Lemma 2.1.

Proposition 2.1. *Consider the STAR model (2.1) with Θ given in (2.2) and let $\mathcal{F} = \mathcal{F}_0^+$. Then there exists a $v \in \Re^2$, a positive definite 2×2 matrix M and measurable functions h_j from $\Re^j \times \Theta$ to $\Re^2$, $j = 1, 2, \ldots$ such that for all local sequences $\langle \vartheta_n \rangle$ and $\langle \theta_n \rangle$ Assumption 2.2 is satisfied.*

Proof. By Remark 2.6 of Koul and Schick (1997), under uniformity condition, it is enough to verify that conditions (2.7)-(2.10) hold with $\theta_n = \theta_0$ and that

$$\frac{1}{n} \sum_{j=1}^{n} \| \dot{H}_j(\vartheta_n) - \dot{H}_j(\theta_0) \|^2 = o_{\theta_0}(1).$$

As $\dot{H}_j(\theta)$, given in (2.4), does not depend on the parameter θ the above condition is obviously satisfied.

The linearity of $H_j(.)$ in θ also shows (2.7), trivially.

To prove (2.8), one needs to show that

$$P_{\theta_n} \left[\max_{1 \le j \le n} \frac{1}{\sqrt{n}} \sqrt{X_{j-1}^2 \left(1 + G^2(X_{j-1})\right)} > \varepsilon \right] \to 0.$$

Let $|Y_{j-1}| = |X_{j-1} \sqrt{(1 + G^2(X_{j-1}))}|$. As $G(x) \le 1$ for any $x \in \Re$, it follows that $|Y_{j-1}| \le 2|X_{j-1}|, \forall \, j \ge 1$. Under the assumption that $\{X_j\}$ is stationary and ergodic, and that $EX_{j-1}^2 < \infty$, we have $EY_{j-1}^2 < \infty$. Then by using a conditional Chebyshev inequality and the stationarity of the process we have,

$$
\begin{aligned}
P_{\theta_n} \left(\max_{1 \le j \le n} \frac{|Y_{j-1}|}{\sqrt{n}} > \varepsilon \right) &\le P_{\theta_n} \left(\max_{1 \le j \le n} \frac{|X_{j-1}|}{\sqrt{n}} > \frac{\varepsilon}{2} \right) \\
&\le \frac{4}{\varepsilon^2 n} \sum_{j=1}^{n} E \left[X_{j-1}^2 I \left(|X_{j-1}| > \frac{\varepsilon \sqrt{n}}{2} \right) \right] \\
&= \frac{4}{\varepsilon^2} E \left[X_0^2 I \left(|X_0| > \frac{\varepsilon \sqrt{n}}{2} \right) \right].
\end{aligned}
$$

The above expression tends to 0 as $n \to \infty$ due to the fact that $EX_0^2 < \infty$.

To prove condition (2.9) and (2.10), we need to prove that, as $n \to \infty$,

$$\begin{pmatrix} \frac{1}{n} \sum_{j=1}^{n} X_{j-1} \\ \frac{1}{n} \sum_{j=1}^{n} X_{j-1} G(X_{j-1}) \end{pmatrix} \xrightarrow{p} v,$$

and

$$\begin{pmatrix} \frac{1}{n} \sum_{j=1}^{n} X_{j-1}^2 & \frac{1}{n} \sum_{j=1}^{n} X_{j-1}^2 G(X_{j-1}) \\ \frac{1}{n} \sum_{j=1}^{n} X_{j-1}^2 G(X_{j-1}) & \frac{1}{n} \sum_{j=1}^{n} X_{j-1}^2 G^2(X_{j-1}) \end{pmatrix} \xrightarrow{p} M,$$

where

$$v = \begin{pmatrix} E_{\theta_0} X_0 \\ E_{\theta_0}(X_0 G(X_0)) \end{pmatrix}, \quad M = \begin{pmatrix} E_{\theta_0} X_0^2 & E_{\theta_0}(X_0^2 G(X_0)) \\ E_{\theta_0}(X_0^2 G(X_0)) & E_{\theta_0}(X_0^2 G^2(X_0)) \end{pmatrix}. \qquad (2.13)$$

By the ergodicity assumption these results follow from Theorem 5.5 and Theorem 5.6 of Karlin and Taylor (1975). $\qquad \square$

Koul and Schick (1997), in Remark 2.7 gave a sufficient condition for a stationary and ergodic NLAR(1) process to satisfy Assumption 2.1 and for a smooth path to be regular in the following proposition. These results are summarised for our case in the following Proposition 2.2.

Proposition 2.2. *Consider the STAR model* (2.1) *with Θ given in* (2.2) *and let $\mathcal{F} = \mathcal{F}_0^+$. For $\theta = (\theta_1, \theta_2) \in \Theta$, let $h(x, \theta) = \theta_1 x + \theta_2 x G(x)$, $x \in \Re$. Assume that there exists a measurable non-negative function ψ, positive constants A, C and δ, for all θ close to θ_0, such that*

$$|h(x, \theta)| \leq A\psi(x), \quad x \in \Re, \tag{2.14}$$

$$|h(x, \theta) - h(x, \theta_0)| \leq \|\theta - \theta_0\| A\psi(x), \quad x \in \Re, \tag{2.15}$$

$$\int \psi(y + h(x, \theta))\phi(y)dy \leq C + (1 - 2\delta)\psi(x), \quad x \in \Re, \tag{2.16}$$

Then the condition (2.6) *in Assumption* 2.1 *is satisfied. Also, if $\eta \mapsto f_\eta$, is a ξ-smooth path then*

$$\limsup_{\theta \to \theta_0, \eta \to 0} E_{\theta, f_\eta} \psi(X_0) < \infty,$$

and the path is regular.

By the above Proposition to prove that STAR is LAN it is enough to show that there exists a function ψ satisfying (2.14)-(2.16). The choice of the function $\psi(x)$ depends on the value of θ. From Nur (1998), it is possible to choose positive constants c, d, e with $e < 1$, satisfying

$$-\frac{d}{c} < \theta_1 < e, \quad -\frac{c}{d} < \theta_1 + \theta_2 < e, \tag{2.17}$$

for all $\theta \in \Theta$.

Lemma 2.1. *Define the non-negative measurable function ψ as*

$$\psi(x) = \begin{cases} cx, & x > 0 \\ -dx, & x \leq 0. \end{cases}$$

Then ψ satisfies the conditions (2.14)-(2.16).

Proof. Proof of (2.14) follows from the fact that

$$|h(x, \theta)| = |\theta_1 x + \theta_2 x G(x)| = |\theta_1 + \theta_2 G(x)||x| \leq A\psi(x),$$

for some positive constant A. Similarly,

$$\begin{aligned}
|h(x, \theta) - h(x, \theta_0)| &= |(\theta_1 - \theta_{01})x + (\theta_2 - \theta_{02})x G(x)| \\
&= |(\theta_1 - \theta_{01} \ \theta_2 - \theta_{02})(x \ x G(x))^T| \\
&\leq \|\theta - \theta_0\| A\psi(x), \quad x \in \Re,
\end{aligned}$$

for some constant A, and (2.15) follows. To prove (2.16) (see the proof of ergodicity in Nur (1998)) we have, for $x \in \mathfrak{R}$,

$$E(\psi(X_t)|X_{t-1} = x)$$

$$= \begin{cases} cx(\theta_1 + \theta_2 G(x)) + (c+d)E[h(x,\theta) + \varepsilon_t]^-, & x > 0 \\ -dx(\theta_1 + \theta_2 G(x)) + (c+d)E[h(x,\theta) + \varepsilon_t]^+, & x \le 0 \end{cases}$$

$$\le \psi(x)(\theta_1 + \theta_2 G(x)) + (c+d)E|h(x,\theta) + \varepsilon_t|,$$

$$\le e\psi(x) + (c+d)E|h(x,\theta) + \varepsilon_t|,$$

as $\theta_1 + \theta_2 G(x) < e$ from (2.17). From the condition (2.14) and that $E(\varepsilon_t^2) < \infty$, it is clear that

$$E|h(x,\theta) + \varepsilon_t| \le |h(x,\theta)| + E|\varepsilon_t| \le A\psi(x) + E|\varepsilon_t| < \infty, \quad x \in \mathfrak{R}.$$

Therefore condition (2.16) is satisfied with $(c+d)E|h(x,\theta) + \varepsilon_t| \le C$, $x \in \mathfrak{R}$ and $\delta = \frac{1-e}{2}$.
$\square$

By taking $\mathcal{F} = \mathcal{F}_0^+$, the condition (2.6) in the Assumption 2.1 holds. Hence by combining the Propositions 2.1 and 2.2, and the Lemma 2.1 we have proved the following.

Theorem 2.1. *Let* $\mathcal{F} = \mathcal{F}_0^+$. *Then the STAR model* (2.1) *is LAN.*

3. Adaptivity

In this section we show that the STAR is adaptive with respect to a suitable class of functions $\mathcal{F}$. By Koul and Schick (1997) Stein's necessary condition for adaptivity is

$$v \int l\xi_q dF = 0, \quad \forall q \in \mathfrak{Q}, \tag{3.1}$$

where v is given in (2.13) and $\mathfrak{Q}$ is the set of regular paths and, for $q \in \mathfrak{Q}$, ξ_q is the smoothness parameter, that is, q is ξ_q-regular. For the STAR model,

$$v = \begin{pmatrix} E_{\theta_0} X_0 \\ E_{\theta_0}(X_0 G(X_0)) \end{pmatrix} \ne 0.$$

Hence we need to choose our $\mathcal{F}$ such that $\int l\xi_q dF = 0$ is satisfied. An appropriate class for this is $\mathcal{F} = \mathcal{F}_S^+$, the set of all positive Lebesgue densities which are symmetric, having zero means, finite variances and finite Fisher information for location.

In order to show that the STAR model is adaptive, we use the sufficient conditions given in Remark 5.4 of Koul and Schick (1997). The conditions require that there exists a function ψ such that

$$E_{\theta_0} \psi(X_0) < \infty \tag{3.2}$$

$$\psi(x) \ge \sup_{\|\theta - \theta_0\| < \delta} \|\dot{h}(x,\theta)\|^2, \quad \text{for all } x \in \mathfrak{R} \text{ and some } \delta > 0 \tag{3.3}$$

Following Koul and Schick (1997) we construct an adaptive estimator $\hat{\theta}_n$ for the STAR(1) model as follows. Let $\langle c_n \rangle$ be a sequence of positive numbers tending to infinity and $\langle \theta_n \rangle$ be a local sequence for θ. The truncated function $\dot{H}_j(\tilde{\theta}_n))$ is defined by

$$\dot{H}_{n,j} = \dot{H}_j(\theta_n) I \left[\|\dot{H}_j(\theta_n)\| \le c_n \right] + c_n \frac{\dot{H}_j(\theta_n)}{\|\dot{H}_j(\theta_n)\|} I \left[\dot{H}_j(\theta_n) > c_n \right], \qquad (3.4)$$

where $\dot{H}_j(\theta_n)$ as defined in (2.4). Let $\varepsilon_{n,j} = \varepsilon_j(\tilde{\theta}_n), \hat{l}_n$ the estimate of the score function, $\hat{J}_n = \frac{1}{n} \sum_{j=d_n}^{n} \hat{l}_n^2(\varepsilon_{n,j})$, and $M_n = \frac{1}{n} \sum_{j=d_n}^{n} \dot{H}_{n,j} \dot{H}_{n,j}^T$. Let $\tilde{\theta}_n$ be a discretized $\sqrt{n}$-consistent estimator of θ_0. With this as the initial estimator we define the estimator $\hat{\theta}_n$ as

$$\hat{\theta}_n = \tilde{\theta}_n + (\hat{J}_n M_n)^{-1} \frac{1}{N_n} \sum_{j=d_n}^{n} \dot{H}_{n,j} \hat{l}_n(\varepsilon_{n,j}). \qquad (3.5)$$

Theorem 3.1. *Consider the STAR(1) model given in (2.1), and let $\mathcal{F} = \mathcal{F}_S^+$.*

(a) *For $A > 2$, let $\psi(x) = Ax^2$, $x \in \Re$. Then the conditions (3.2) and (3.3) are satisfied.*

(b) *The estimator $\hat{\theta}_n$ is adaptive.*

Proof. (3.2) follows from the fact that $E_{\theta_0} X_0^2 < \infty$. To prove (3.3) note that

$$\sup_{\|\theta - \theta_0\| < \delta} \|\dot{h}(x, \theta)\|^2 = \sup_{\|\theta - \theta_0\| < \delta} x^2 (1 + G^2(x)) \le Ax^2$$

for $x \in \Re$ and for some $\delta > 0$, $A > 2$. Part(b) follows from Remark 5.4 and Theorem 5.2 of Koul and Schick (1997). $\qquad\qquad\square$

The construction of an efficient estimator for STAR(1) models is similar to the one for SETAR(2;1,1) given in Example 3.5 of Koul and Schick (1997) by taking $\mathcal{F} = \mathcal{F}_0^+$. The construction of an efficient estimate using a sample splitting technique as follows. Let $\langle d_n \rangle$ and $\langle m_n \rangle$ be sequences of positive integers such that $d_n \le m_n \le n, d_n/n \to 0$ and $m_n/n \to 1/2$ and choose $\langle a_n \rangle$ and $\langle b_n \rangle$ to be sequences of positive numbers converging to 0 such that $n^{-1} a_n^{-3} b_n^{-1} \to 0$. Let $\langle \tilde{\theta}_n \rangle$ be a preliminary estimate of θ which is discretized and $\sqrt{n}$-consistent. Set $\varepsilon_{n,j} = \varepsilon_j(\tilde{\theta}_n)$, $j = 1, \ldots, n$, $N_n' = m_n - d_n + 1$, $N_n'' = n - m_n$, $\mathbf{e}_{n,1} = (\varepsilon_{n,d_n}, \ldots, \varepsilon_{n,m_n})$ and $\mathbf{e}_{n,2} = (\varepsilon_{n,m_n+1}, \ldots, \varepsilon_{n,n})$.

Let

$$\hat{\psi}_{n,j} = \begin{cases} \dot{H}_j(\tilde{\theta}_n) L_{N_n''}(\varepsilon_{n,j}, \mathbf{e}_{n,2}) - \hat{v}_{2,n} L_{*,N_n''}(\varepsilon_{n,j}, \mathbf{e}_{n,2}) & \text{if } j = d_n, \ldots m_n, \\ \dot{H}_j(\tilde{\theta}_n) L_{N_n'}(\varepsilon_{n,j}, \mathbf{e}_{n,1}) - \hat{v}_{1,n} L_{*,N_n'}(\varepsilon_{n,j}, \mathbf{e}_{n,1}) & j = m_n + 1, \ldots n. \end{cases}$$

The functions $L_n(.)$ and $L_{*,n}(.)$ are defined in (4.4) and (4.5) of Koul and Schick (1997) and

$$\hat{v}_{1,n} = \frac{1}{N_n'} \begin{pmatrix} \sum_{j=d_n}^{m_n} X_{j-1} \\ \sum_{j=d_n}^{m_n} X_{j-1} G(X_{j-1}) \end{pmatrix} \text{ and } \hat{v}_{2,n} = \frac{1}{N_n''} \begin{pmatrix} \sum_{j=m_n+1}^{n} X_{j-1} \\ \sum_{j=m_n+1}^{n} X_{j-1} G(X_{j-1}) \end{pmatrix}$$

whereas $\dot{H}_j(\tilde{\theta}_n)$ is defined in (2.4). Define the efficient estimates as

$$\hat{\theta}_n = \tilde{\theta}_n + \left(\frac{1}{n}\sum_{j=d_n}^{n}\hat{\psi}_{n,j}\hat{\psi}_{n,j}^T\right)^{-1}\frac{1}{n}\sum_{j=d_n}^{n}\hat{\psi}_{n,j}. \tag{3.6}$$

Theorem 3.2. *Consider the STAR*(1) *model given in* (2.1)*, and let* $\mathcal{F} = \mathcal{F}_0^+$*. Then* $\hat{\theta}_n$ *given in* (3.6) *is efficient.*

Remark. For the STAR model, if $\mathcal{F} = \mathcal{F}_S^+$, then Stein's condition for adaptivity is satisfied. Hence the efficient estimator is also adaptive. Whereas, if $\mathcal{F} = \mathcal{F}_0^+$ then Stein's condition is not satisfied. Hence θ_n is efficient but not adaptive. In other words, in the class of symmetric error densities, an adaptive estimator is always efficient, and an efficient estimator is also adaptive. But for a general class of error densities, an efficient estimator is optimal but not adaptive.

4. Examples

This section presents two examples to complement the theoretical results developed in the previous sections. The necessary computations were written in Fortran 77, with a few subroutines included from the Nag-library. We first present a simulation study in which the adaptive and efficient estimators for the STAR(1) are constructed and compared to conditional least squares estimator. The adaptive estimator is constructed without the truncation of $\dot{H}_j(\theta_n)$. The efficient estimators are constructed using a sample splitting technique as explained in the previous section and non-splitting method as in Koul and Schick (1997).

4.1. Simulation

The data come from a first-order STAR model

$$X_t = -2.0X_{t-1} + 1.6X_{t-1}G\left(\frac{X_{t-1}-5.0}{0.5}\right) + \varepsilon_t,$$

where $\theta_0 = (\theta_{01}, \theta_{02}) = (-2.0, 1.6)$, $(r, d, z) = (5.0, 1, 0.5)$ and $G(.)$ is a standard Normal distribution function. We generate 2500 independent replications each with sample size 200. The following five densities of the errors ε_t were chosen:

$$\phi_1(x) = (0.5/\sqrt{2\pi})\exp(-(x-3)^2/2) + (0.5/\sqrt{2\pi})\exp(-(x+3)^2/2),$$

$$\phi_2(x) = (0.05/\sqrt{50\pi})\exp(-x^2/50) + (0.95/\sqrt{2\pi})\exp(-x^2/2),$$

$$\phi_3(x) \sim t_5, \quad \phi_4(x) \sim t_7, \quad \phi_5(x) \sim t_9.$$

Here ϕ_1, ϕ_2 are mixture normal densities and ϕ_3, ϕ_4 and ϕ_5 are t-densities functions. Note that all these densities belong to $\mathcal{F} = \mathcal{F}_S^+$. The densities ϕ_1, ϕ_2 are commonly used for studying the behaviour of estimators in autoregressive models (Kreiss, 1987) and t-densities are used for GARCH models (Drost, *et al*, 1997).

For each series, conditional on $(r,d,z) = (5.0, 1, 0.5)$, different estimates $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$, of the parameter coefficients $\theta_0 = (\theta_{01}, \theta_{02})$ are obtained using three methods, viz, conditional least squares, adaptive and efficient estimation. The preliminary estimate $\tilde{\theta}$, used as the initial value for these estimators, is a discretized estimate obtained by the conditional least squares method. Note that $\tilde{\theta}$ is a $\sqrt{n}$-consistent estimator by the results given in Nur (1998).

The adaptive estimates, $\hat{\theta}_{Ad}$, is constructed by using (3.4) and (3.5) without truncation, i.e. $c_n \to \infty$. The efficient estimate $\hat{\theta}_{Ef1}$, as defined in (3.6), is constructed using a sample splitting technique (Theorem 4.1 of Koul and Schick (1997)) whereas $\hat{\theta}_{Ef2}$, is constructed using a non-splitting technique given in Koul and Schick (1997). The value $(\hat{\theta}_1(C), \hat{\theta}_2(C))$ given in Table 1 is the usual conditional least squares estimator. For all estimates, we used standardised logistic kernels with a bandwidth in the interval $0.5 \leq a_n \leq 0.9$.

The simulation results are constructed for $a_n = 0.5, 0.6, 0.7, 0.8, 0.9$. For each bandwidth, the average values of estimates from 2500 independent replications are given for different error densities. The sample mean squared error (MSE) is given in the ( ) parentheses. For convenience, the numerical results are presented only for $a_n = 0.7$ in Table 1 whereas we briefly summarise the results for others which was given in details in Nur (1998).

Generally for all values of a_n the adaptive estimates give the smallest sample MSE compared to others when the error densities are t_5, t_7, t_9. The second smallest sample MSE were given either by the CLS estimator or the efficient estimates without splitting the sample. As it is expected from a moderate sample size of 200, for all densities, the efficient estimate using splitting technique gives greater sample MSE compare to others. On the other hand, the smallest bias is obtained by the adaptive estimates followed by the CLS and the efficient estimates.

Table 1. Average values of estimates and their sample mean squared error for $a_n = 0.7$. The indexes C stands for conditional least squares, Ad for adaptive, $Ef1$ for efficient estimates with splitting the sample and $Ef2$ for efficient estimates without splitting the sample

Estimate	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
$\hat{\theta}_1(C)$	-1.99943	-1.99972	-2.00096	-1.99937	-1.99964
	(0.000225)	(0.000403)	(0.0006699)	(0.000563)	(0.000505)
$\hat{\theta}_2(C)$	1.59947	1.59985	1.60206	1.60055	1.60016
	(0.000291)	(0.000515)	(0.000879)	(0.000717)	(0.000668)
$\hat{\theta}_{1,Ad}$	-1.99950	-1.99992	-2.00099	-2.00001	-2.00015
	(0.000241)	(0.000423)	(0.000529)	(0.000527)	(0.000491)
$\tilde{\theta}_{2,Ad}$	1.59843	1.59902	1.60088	1.60013	1.59967
	(0.000318)	(0.000548)	(0.000699)	(0.000672)	(0.000652)
$\hat{\theta}_{1,Ef1}$	-1.99921	-1.99989	-2.00104	-1.99913	-1.99918
	(0.000466)	(0.000702)	(0.000968)	(0.000879)	(0.000857)
$\tilde{\theta}_{2,Ef1}$	1.59914	1.60020	1.60245	1.60068	1.60033
	(0.001005)	(0.001326)	(0.001757)	(0.001569)	(0.001556)
$\tilde{\theta}_{1,Ef2}$	-2.00107	-2.00126	-2.00248	-2.00118	-2.00142
	(0.000275)	(0.000446)	(0.000576)	(0.000561)	(0.000536)
$\tilde{\theta}_{2,Ef2}$	1.60182	1.60188	1.60394	1.60271	1.60243
	(0.000393)	(0.000596)	(0.000840)	(0.000779)	(0.000756)

4.2. Canadian lynx data

The second study is based on the well-known Canadian lynx data which consists of the annual record of the numbers of the lynx trapped in MacKenzie River, Canada for the period 1821 to 1934 inclusively. Tong (1990) presented a complete review of analysis of lynx data using different models which includes AR(2) model and SETAR(2;7,2) model. The best fitted STAR model for the data is of order 1 with $d = 2$ (Nur, 1998). In order to be consistent with the theoretical results given previously, a STAR model order 1 with $d = 1$ was fitted to the logarithmic transformation of the original data as in the previous studies of the data set. Here we take the error distribution ε_t as t_5. We fitted the STAR order 1 model with $d = 1$, $\hat{r} = 3.83$ and $\hat{z} = 0.721$, taken from Tong (1983, 1990), with the cumulative distribution function G as the Gaussian distribution. The estimated variances obtained by each estimation method are given in Table 2 below. It shows that the CLS method performs as well as the adaptive estimates followed by the efficient estimates without splitting the sample and with splitting the sample. Similar results were obtained for other densities, with the CLS and the adaptive estimates methods outperforming others, and are omitted here.

Table 2. Estimated variances for the the lynx data with $a_n = 0.5$. *CLS* stands for Conditional Least Squares, $Ef1$ stands for efficient estimates with splitting the sample and $Ef2$ stands for efficient estimates without splitting the sample

Method	CLS	Adaptive	$Ef1$	$Ef2$
$v\hat{a}r$	0.2250	0.2255	0.2272	0.2260

5. Conclusion

In conclusion, we have shown that a first-order STAR model with delay parameter one is locally asymptotically normal by Theorem 2.1. The adaptive estimator of this model only exists for a class of symmetric and positive Lebesgue densities with zero-mean, finite variances and finite Fisher information for location as given in Theorem 3.1. The efficient estimates of this model exist for a class of all positive Lebesgue densities with zero-mean, finite variances and finite Fisher information for location as shown in Theorem 4.1. The simulation results were presented to compare conditional least squares, adaptive and efficient estimators, employing a sample splitting and a non splitting methods. The error densities chosen were mixture normals, t_5, t_7 and t_9. The simulation study leads to the conclusion that, in general, the adaptive estimates or efficient estimates without splitting the sample perform better in terms of MSE than the CLS for the t-densities. The CLS estimates, however, perform better in MSE for mixture normals. For the Canadian lynx data, the CLS and the adaptive estimates perform better than other methods.

References

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J., 1993. *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD; John Hopkins Univ. Press.

Chan, K.S., Tong, H., 1986. On Estimating Threshold in Autoregressive models. *Journ. of Time Series Anal.*, 7, 179–190.

Dijk, van D., Terasvirta, T., Werker, Franses, P.H., 2002. Smooth transition autoregressive models – a survey of recent developments. *Econometric Reviews*, 21 (1), 1–47.

Drost, F.C., Klaassen, C.A.J., Werker, B.J.M., 1993. Adaptiveness in time series models. In P. Mandl and M. Huskova (eds.), *Asymptotic Statistics, Proceedings of the Fifth Prague Symposium,* pp. 203–211. Heidelberg: Physica-Verlag.

Drost, F.C., Klaassen, C.A.J., Werker, B.J.M., 1994. Adaptive estimation in time series models. Discussion Paper Series, No. 9488, CentER, Tilburg University, The Netherlands.

Drost, F.C., Klaassen, C.A.J., Werker, B.J.M., 1997. Adaptive estimation in time series models. *Annals of Statistics* 25, 786–817.

Drost, F.C., Klaassen, C.A.J., 1997. Efficient estimation in semiparametric GARCH models. *Journal of Econometrics*, 81, 193–221.

Fabian, V., Hannan, J., 1982. On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsvheinlichkeitstheorie Verw. Geb.,* 59, 459–478.

Fabian, V., Hannan, J., 1987. Local asymptotic behaviour of densities. *Statist. Decisions*, 5, 105–138. Correction: 6, 1988, p. 195.

Hájek, J., 1970. A characterisation of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlich keitstheorie und Verwandte Gebiete*, 14, 323–330.

Karlin, S., Taylor, H.M., 1975. *A First Course in Stochastic Processes*, 2nd edition. Academic Press, New York.

Koul, H.L., Schick, A., 1996. Adaptive estimation in a random coefficient auto regressive model. *Ann. Statist.*, 24, 1025–1052.

Koul, H.L., Schick, A., 1997. Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli*, 3, 247–277.

LeCam, L., 1960. Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.*, 3, 37–98.

LeCam, L., Yang, G.L., 1990. *Asymptotics in Statistics: Some Basic Concepts.* Springer, New York.

Linton, O., 1993. Adaptive estimation in ARCH models. *Econometric Theory*, 9, 539–569.

Nur, D., 1998. *Parameter Estimation of Smooth Threshold Autoregressive(1) models.* Ph.D. thesis, Curtin University of Technology, Australia.

Stein, C., 1956. Efficient nonparametric estimation and testing. In J. Neyman (ed.) *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, 187–195. Berkeley, University of California Press.

Tong, H., 1983. *Threshold Models in Nonlinear Time Series Analysis.* New York: Springer-Verlaag.

Tong, H., 1990. *Nonlinear Time Series Analysis: a Dynamical System Approach.* London: Oxford Univ. Press.