МАТЕМАТИЧКИ ВЕСНИК
54 (2002), 31–36

UDK 519.218
оригинални научни рад
research paper

M-ESTIMATES OF SETAR MODEL PARAMETERS

Dragan Đorić

Abstract. For a stationary ergodic self-exiting threshold autoregressive model with single threshold parameter Chan (1993) obtained the consistency of the least-squares estimator and Qian (1998) proved it for the maximum likelihood estimator. The aim of this paper is to derive the similar results for the M-estimates of the same model under some regularity conditions.

1. Introduction

Nonlinear time series has drawn much attention in recent years and many classes of models have been proposed. One of the most popular is a class of threshold models characterized by piecewise linear processes separated according to the magnitude of threshold variable. When each linear regime follows an autoregressive process we have the well known threshold autoregressive (TAR) models. The major features of this class of models include limit cycles, jump resonance and interesting asymmetric features observed in economic and financial time series. Tong (1990) provides an excellent review of properties of that models. One of the TAR models is self-exciting threshold autoregressive (SETAR) model where threshold variable is one of the past observations. SETAR model can also exhibit nonlinear phenomena and Petruccelli (1992) shown that processes generated by this model may be viewed as an approximation to a more general class of nonlinear processes.

There are mainly two methods of parameter estimation which have been used in the literature of nonlinear time series analysis. The general theory of least squares estimates and maximum-likelihood estimators of nonlinear smooth autoregressive models is due to Klimko and Nelson (1978). Tjostheim (1986) extended these results to a general class of estimators. As in the threshold models the autoregression function is not differentiable at some points, these results are not applicable.

There are also results for estimators of nonlinear models with non-smooth, but Lipschitz continuous, autoregression function. Liebscher (2000) considered the model

$$X_{t+1} = g(X_t, \dots, X_{t-p+1}; \theta_0) + \varepsilon_{t+1}, \quad t = p, p+1, \dots$$

AMS Subject Classification: 62 M 10
Keywords and phrases: Autoregressive threshold model, M-estimates, strong consistency.

where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables, $\theta_0 \in \Theta \subset R^p$ is the vector of the true parameters and $g : R^p \times \Theta \to R$ is a continuous function. He utilized a variational principle from stochastic optimization theory and proved strong convergence for some classes of robust estimators. Suppose that

$$Q_n(\theta) = \sum_{t=p}^{n-1} \rho(X_{t+1} - g(\mathbf{X}_t; \theta)),$$

where $\mathbf{X}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})^T$. Then M-estimator is determined by the function $\rho \colon R \to R$. Liebscher (2000) assumed that ρ satisfies a nonuniform Lipschitz condition

$$|\rho(x) - \rho(y)| \le L_\rho(|x|^\alpha + |y|^\alpha + 1)|x - y|$$

for $\alpha > 0$. It is possible to apply his result to continuous SETAR, but not for a general SETAR models.

Chan (1993) developed the strong consistency and limiting distribution of the conditional least-squares estimators in a SETAR$(2; p, p)$ model. Qian (1998) obtained it for the maximum-likelihood estimates. In this paper we derive the similar results for the M-estimates. Section 2 presents the model and parameter estimates, while Section 3 describes assumptions and obtains the strong consistency of the M-estimates of the true parameters.

2. Model and parameter estimates

The process $\{X_t\}$ is said to be a self-exciting threshold autoregressive process with threshold variable X_{t-d} if it is generated by SETAR$(k; p_1, \dots, p_k)$ model defined by

$$X_t = a_{i,0} + a_{i,1}X_{t-1} + \cdots + a_{i,p_i}X_{t-p_i} + \varepsilon_t$$

for $r_{i-1} < X_{t-d} \le r_i, \quad i = 1, 2, \dots, k$, where real numbers r_i satisfy

$$-\infty = r_0 < r_1 < \cdots < r_k = +\infty$$

and form a partition of the space of X_{t-d}. The innovation $\{\varepsilon_t\}$ is an i.i.d. sequence of random variables independent of the past $X_{t-1}, X_{t-2}, \dots$. The parameters r_i and the positive integer d are called the thresholds and the delay, respectively. The coefficients $a_{i,j}$ and $b_{i,j}$ are some real constants. Here, we only consider the case $k = 2$ and $p_1 = p_2 = p$,

$$X_t = \begin{cases} a_{1,0} + a_{1,1}X_{t-1} + \cdots + a_{1,p}X_{t-p} + \varepsilon_t, & X_{t-d} \le r \\ a_{2,0} + a_{2,1}X_{t-1} + \cdots + a_{2,p}X_{t-p} + \varepsilon_t, & X_{t-d} > r. \end{cases} \tag{2.1}$$

Given the data $X_0, X_1, \dots, X_n$ generated from (2.1), the true parameter θ_0,

$$\theta_0 = (A_1^T, A_2^T, r, d)^T \tag{2.2}$$

where

$$A_i = (a_{i,0}, a_{i,1}, \dots, a_{i,p})^T \quad i = 1, 2 \tag{2.3}$$

should be estimated. We assume that θ_0 is an interior point of the parameter space $R^{2p+2} \times \overline{R} \times \{1, 2, \dots, p\}$ where $\overline{R} = R \cup \{-\infty, \infty\}$. There exists a compact subset

K of R^{2p+2} such that θ_0 is an interior point of $\Theta = K \times \overline{R} \times \{1, 2, \ldots, p\}$. We define the estimator $\widehat{\theta}_n$ as a global minimizer of a criterion $Q_n(\theta)$ on a compact subset Θ

$$\widehat{\theta}_n \in \operatorname*{argmin}_{\theta \in \Theta} Q_n(\theta), \tag{2.4}$$

where $Q_n(\theta) = \sum_{i=p}^{n} \rho_k(e_k(\theta), \theta)$ and $e_k(\theta) = X_k - E_\theta\left(X_k \mid \mathcal{F}_{k-1}\right)$. Least-squares estimators represent a special case of $\widehat{\theta}_n$ for $\rho_k(x) = \rho(x) = x^2$. The minimization can be done in two steps. Assume that $\theta = (B_1^T, B_2^T, s, c)^T$ is a general parameter in Θ and let $A = (A_1^T, A_2^T)^T$ and $B = (B_1^T, B_2^T)^T$.

1. For fixed $s \in \overline{R}$ and $c \in \{1, 2, \ldots, p\}$ let $B_n(s, c)$ be defined by

$$B_n(s, c) \in \operatorname*{argmin}_{B \in K} Q_{nsc}(B),$$

where $Q_{nsc}(B) = Q_n(B, s, c) = Q_n(\theta)$.

2. Let $\widehat{r}_n$ i $\widehat{d}_n$ be the values of s and c for which

$$Q_n(B_n(\widehat{r}_n, \widehat{d}_n), \widehat{r}_n, \widehat{d}_n) = \min_{s \in \overline{R}, c \in \{1, 2, \ldots, p\}} Q_n(B_n(s, c), s, c)$$

and let $\widehat{A}_n = B_n(\widehat{r}_n, \widehat{d}_n)$.

For any $\theta \in \Theta$, by definition of B_n, $\widehat{r}_n$ and $\widehat{d}_n$, we have

$$\begin{aligned} Q_n(\theta) &= Q_n(B, s, c) \geq Q_n(B_n(s, c), s, c) \\ &\geq Q_n(B_n(\widehat{r}_n, \widehat{d}_n), \widehat{r}_n, \widehat{d}_n) = Q_n(\widehat{A}_n^T, \widehat{r}_n, \widehat{d}_n) = Q_n(\widehat{\theta}_n) \end{aligned}$$

and, hence, $\widehat{\theta}_n = \left(\widehat{A}_n^T, \widehat{r}_n, \widehat{d}_n\right)^T$.

3. Assumptions and strong consistency

Assume that $\rho_k = \rho(e_k)$. The model (2.1) can be represented as

$$X_t = h(\mathbf{X}_t, \theta_0) + \varepsilon_t, \quad t \geq 1$$

where, for $\mathbf{x} = (x_1, \ldots, x_p) \in R^p$, $\theta \in \Theta$,

$$h(\mathbf{x}, \theta) = \left(a_{1,0} + \sum_{i=1}^{p} a_{1,i} x_i\right) I(x_d \leq r) + \left(a_{2,0} + \sum_{i=1}^{p} a_{2,i} x_i\right) I(x_d > r).$$

Let $h(x, \theta) = h_{sc}(x, B) = h(x, B, s, c)$,

$$\dot{h}_{sc}(x) = \frac{\partial}{\partial B} h_{sc}(x, B) \tag{3.1}$$

$$z = (1, x^T)^T. \tag{3.2}$$

Then,

$$\dot{h}_{sc}(x, B) = \left(z^T I(x_c \leq s), z^T I(x_c > s)\right)^T \tag{3.3}$$

and

$$h(x,\theta) = B^T \dot{h}_{sc}(x) \tag{3.4}$$

for $s \in \overline{R}$, $x \in R^p$ and $c \in \{1,\dots,p\}$. Equations (3.1), (3.2) and (3.3) imply that

$$|\dot{h}_{sc}(x)| \le \sqrt{1+\|x\|^2}. \tag{3.5}$$

From the last inequality and (3.4) we have

$$|h(x,\theta)| \le \|B\|\sqrt{1+\|x\|^2}. \tag{3.6}$$

Furthermore, for any $s_1 \in \overline{R}$, $s_2 \in \overline{R}$ i $c \in \{1,2,\dots,p\}$

$$\begin{aligned} |\dot{h}_{s_1 c}(x) - \dot{h}_{s_2 c}(x)| &\le \sqrt{2(1+\|x\|^2)} \cdot I(\min\{s_1,s_2\} < x_c \le \max\{s_1,s_2\}) \\ &\le \sqrt{2(1+\|x\|^2)} \cdot I(|x_c - s_2| \le |s_1 - s_2|) \end{aligned} \tag{3.7}$$

To derive the consistency of the estimates we shall need the following assumptions on the class of function ρ.

CONDITION 1. *For $x \in R$, $y \in R$ and some $\alpha > 0$, $\beta \ge 1$, $\gamma > 0$ function ρ satisfies*

$$|\rho(x) - \rho(y)| \le \alpha\left(|x|^\beta + |y|^\beta + 1\right) \cdot |x-y|^\gamma.$$

CONDITION 2. *$\delta E\rho(\varepsilon_1 + a) > E\rho(\varepsilon_1)$ for $a \ne 0$.*

Moreover, suppose the existence of some moments of the process $\{X_t\}$.

CONDITION 3. *$E\,|X_t|^{2\beta} < \infty$ and $E\,|X_t|^{2\gamma} < \infty$.*

First, we need to prove the following statement.

LEMMA 1. *Suppose that $\{X_t\}$ in model (2.1) is stationary and ergodic and the Conditions 1 and 3 hold. Then, for any $\theta \in \Theta$*

$$E \sup_{\theta^* \in U_\theta(\nu)} |\rho(e_1(\theta^*)) - \rho(e_1(\theta))| \;\to\; 0 \quad (\nu \to 0),$$

where $U_\theta(\nu)$ is a ν neighborhood for θ given by

$$U_\theta(\nu) = \left\{\theta^* \in \Theta \;:\; \theta^* = ({B^*}^T, s^*, c)^T,\ |B^* - B| < \nu,\ |s^* - s| < \nu,\ \nu > 0\right\}.$$

Proof. Let $\varepsilon_1(\theta) = X_1 - h(\mathbf{X}_0,\theta)$ and $\delta(x,\theta^*) = h(x,\theta) - h(x,\theta^*)$. Then,

$$\delta(x,\theta^*) = h_{s,c}(x,B) - h_{s,c}(x,B^*).$$

From (3.4) and (3.5) we conclude

$$|h_{sc}(x,B) - h_{sc}(x,B^*)| \le \|B - B^*\|\sqrt{1+\|x\|^2}.$$

By (3.3) and (3.7) we have

$$\begin{aligned} |h_{sc}(x,B^*) - h_{s^*c}(x,B^*)| &\le \|B^*\|\sqrt{2(1+\|x\|^2)} I(\min\{s,s^*\} < s_c \le \max\{s,s^*\}) \\ &\le \|B^*\|\sqrt{2(1+\|x\|^2)} I(|x_c - s| \le |s^* - s|). \end{aligned} \tag{3.8}$$

Thus for $\theta^* \in U_\theta(\nu)$ and $s \in R$, $x \in R^p$

$$
\begin{aligned}
|\delta(x,\theta^*)| &\le |h_{sc}(x,B) - h_{s^*c}(x,B)| + |h_{s^*c}(x,B) - h_{s^*c}(x,B^*)| \\
&\le \left(\sqrt{2}\|B\| I(|x_c - s| \le |s^* - s|) + \|B - B^*\|\right)\sqrt{1+\|x\|^2} \\
&\le \left(\sqrt{2}\|B\| I(|x_c - s| \le \nu) + \nu\right)\sqrt{1+\|x\|^2}. \qquad (3.9)
\end{aligned}
$$

It follows from Conditions 1. and inequality

$$
(a+b)^\beta \le 2^{\beta-1}(a^\beta + b^\beta), \quad a>0,\ b>0,\ \beta \ge 1
$$

that for $\alpha > 0$, $\beta \ge 1$ i $\gamma > 0$

$$
\begin{aligned}
\Delta\rho(\theta,\theta^*) &= |\rho(e_1(\theta^*)) - \rho(e_1(\theta))| \\
&= |\rho(X_1 - h(\mathbf{X}_0,\theta^*)) - \rho(X_1 - h(\mathbf{X}_0,\theta))| \\
&\le \alpha\left(|X_1 - h(\mathbf{X}_0,\theta^*)|^\beta + |X_1 - h(\mathbf{X}_0,\theta)|^\beta + 1\right)|h(\mathbf{X}_0,\theta) - h(\mathbf{X}_0,\theta^*|^\gamma \\
&= \alpha\left(|\varepsilon_1 + \delta(\mathbf{X}_0,\theta_0,\theta^*)|^\beta + |\varepsilon_1 + \delta(\mathbf{X}_0,\theta_0,\theta)|^\beta + 1\right)|\delta(\mathbf{X}_0,\theta,\theta^*)|^\gamma \\
&\le \alpha(2^\beta|\varepsilon_1|^\beta + 2^{\beta-1}|\delta(\mathbf{X}_0,\theta_0,\theta^*)|^\beta + 2^{\beta-1}|\delta(\mathbf{X}_0,\theta_0,\theta)|^\beta + 1)|\delta(\mathbf{X}_0,\theta,\theta^*)|^\gamma \\
&\le \alpha\left(2^\beta|\varepsilon_1|^\beta + 2^{\beta-1}\|A - B^*\|^\beta\right. \\
&\quad \left.\cdot\,(1+\|\mathbf{X}_0\|^2)^{\beta/2} + 2^{\beta-1}\|A-B\|^\beta(1+\|\mathbf{X}_0\|^2)^{\beta/2} + 1\right) \\
&\quad \cdot\,|\delta(\mathbf{X}_0,\theta,\theta^*)|^\gamma \\
&= \alpha\left(2^\beta|\varepsilon_1|^\beta + 1\right)|\delta(\mathbf{X}_0,\theta,\theta^*)|^\gamma + \\
&\quad + \alpha 2^{\beta-1}\left(\|A-B^*\|^\beta + \|A-B\|^\beta\right)(1+|\mathbf{X}_0|^2)^{\beta/2}|\delta(\mathbf{X}_0,\theta,\theta^*)|^\gamma
\end{aligned}
$$

If $\overline{\delta}(x,\nu) = \left(\sqrt{2}\|B\| I(|x_c - s| \le \nu) + \nu\right)\sqrt{1+\|x\|^2}$, then by (3.9)

$$
\Delta\rho(\theta,\theta^*) \le K_1|\overline{\delta}(\mathbf{X}_0,\nu)|^\gamma + K_2|\varepsilon_1|^\beta|\overline{\delta}(\mathbf{X}_0,\nu)|^\gamma + K_3(1+\|\mathbf{X}_0\|^2)^{\beta/2}|\overline{\delta}(\mathbf{X}_0,\nu)|^\gamma.
$$

By the inequalities

$$
E\ |\varepsilon_1|^\beta|\overline{\delta}|^\gamma \le \left(E\ \varepsilon_1^{2\beta}\right)^{1/2}\cdot\left(E\ \overline{\delta}^{2\gamma}\right)^{1/2}
$$

$$
E\ \left(1+\|\mathbf{X}_0\|^2\right)^{\beta/2}|\overline{\delta}|^\gamma \le \left(E\ \left(1+\|\mathbf{X}_0\|^2\right)^\beta\right)^{1/2}\cdot\left(E\ \overline{\delta}^{2\gamma}\right)^{1/2}
$$

and the fact that $E\overline{\delta}^{2\gamma}(\mathbf{X}_0,\nu) \to 0, \quad (\nu \to 0)$ we conclude the statement for $s \in R$.

In the case $s = \infty$, for $x \in R^p$,

$$
\begin{aligned}
|\delta(x,\theta,\theta^*)| &\le \left(\sqrt{2}\|B\| I(x_c > s^*) + \nu\right)\sqrt{1+\|x\|^2} \\
&\le \left(\sqrt{2}\|B\| I(d(x_c,\infty) < \nu) + \nu\right)\sqrt{1+\|x\|^2} = \overline{\delta}_\infty(x,\nu)
\end{aligned}
$$

and again, $E\overline{\delta}_\infty^{2\gamma}(\mathbf{X}_0,\nu) \to 0, \quad (\nu \to 0)$.

The proof is similar in the case $s = -\infty$. Thus, the lemma is proved. ■

Now, we are ready to state the main result.

THEOREM 1. *Under the assumptions of Lemma 1 and Condition 2*

$$\widehat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0, \quad n \to \infty.$$

Proof. Let the function $f : \Theta \to R$ be defined by $f(\theta) = E\,\rho\,(e_1(\theta))$. Condition 2. implies

$$f(\theta) = \int_{R^p} E\,(\varepsilon_1 + h(y, \theta_0) - h(y, \theta))\,dF_{\mathbf{X}_0}(y) > \; E\,\rho(\varepsilon_1) = \; f(\theta_0), \tag{3.10}$$

where $F_{\mathbf{X}_0}$ is distribution function for $\mathbf{X}_0$.

Define, then, the sequence of functions $f_n : \Theta \to R$ by

$$f_n(\theta) = \frac{1}{n-p+1} \sum_{k=p}^{n} \rho(X_{k+1} - h(\mathbf{X}_0, \theta)).$$

Liebscher (2000) proved that, under the stationarity and ergodicity of the process $\{X_t\}$,

$$\sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| \xrightarrow{\text{a.s.}} 0, \quad n \to \infty$$

As, by definition of the functions f_n and the estimates $\widehat{\theta}_n$, $\widehat{\theta}_n \in \operatorname{argmin}_{\theta\in\Theta} f_n(\theta)$ and, by (3.10), $\theta_0 = \operatorname{argmin}_{\theta\in\Theta} f(\theta)$, then the result of Lemma 1 gives sufficient conditions for the convergence of $\widehat{\theta}_n$ to θ_0 as shown by Vogel (1994). This completes the proof. ∎

REFERENCES

[1] K. S. Chan, *Consistency and limiting distributions of the least squares estimator of a threshold autoregressive model*, Annals of Statistics, **21** (1993), 520–533.

[2] L. A. Klimko and P. L. Nelson, *On conditional least-squares estimation for stochastic processes*, Annals of Statistics, **6** (1978), 629–642.

[3] E. Liebscher, *Strong convergence of estimators in nonlinear autoregressive models*, Working paper, (2000), Technical University Ilmenau.

[4] D. J. Petruccelli, *On the approximation of the time series by threshold autoregressive models*, Sankhya, Series B, **54** (1992), 106–113.

[5] L. Qian, *On maximum likelihood estimators for a threshold autoregression*, J. Statistical Planning and Inference, **75** (1998), 21–46.

[6] D. Tjostheim, *Estimation in nonlinear time series models*, Stochastic Processes Appl., **21** (1986), 251–273.

[7] H. Tong, *Nonlinear Tome Series: A Dynamical Approach*, Oxford University Press.

[8] S. Vogel, *A stochastic approach to stability in stochastic programming*, J. Comp. Appl. Math., **56** (1994), 65–96.

(received 06.03.2002)

University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, Belgrade, Yugoslavia
E-mail: djoricd@fon.bg.ac.yu