Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT

ELSEVIER

Journal of Multivariate Analysis 89 (2004) 243–260

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

The limiting behavior of least absolute deviation estimators for threshold autoregressive models

Lihong Wang* and Jinde Wang

Department of Mathematics, Nanjing University, Nanjing, 210093, China

Received 13 September 2000

Abstract

The asymptotic behavior of the least squares (LS) estimators of the parameters in threshold autoregressive models has been completely studied in the literature. It is well known that in some cases the least absolute deviation (LAD) estimators are superior to the LS-estimators. This paper is devoted to studying the strong consistency and the asymptotic normality of the LAD-estimators in two cases where the threshold is known and/or unknown.

AMS 2000 subject classifications: primary 62M10; secondary 62F12

Keywords: Asymptotic normality; Least absolute deviation estimation; Nonlinear time series; Strong consistency; Threshold autoregressive models

1. Introduction

Nonlinear time series analysis is a field of growing popularity. Several classes of nonlinear time-series models have been proposed and illustrated in the literature. One particular class of models which has received a great deal of attention is the TAR model. This model is originally introduced by Tong [14]. In [15], numerous examples from diverse fields are described in which the notion of a threshold is dominant such as radio engineering, medical engineering, population biology, economics, ecology and so on. Specifically, in this paper we will treat the following

*Corresponding author.
E-mail address: lhwang@nju.edu.cn (L. Wang).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2004.02.006

TAR(p) model:

$$X_t = \sum_{j=1}^{m} \sum_{i=1}^{p} \theta_{ij} X_{t-i} H_j(X_{t-d}) + e_t, \quad t = 0, 1, \ldots, \tag{1}$$

where m, p are known positive integers and the delay d $(1 \leqslant d \leqslant p)$ is known; $H_j(X_{t-d}) = I(X_{t-d} \in F_j), I(.)$ is the indicator function and $F_1, \ldots, F_m$ are disjoint subsets of R such that $\sum_{j=1}^{m} F_j = R; \{e_t\}$ is an independent identically distributed (i.i.d.) random sequence with median zero, finite variance and e_t is independent of the past $X_{t-1}, X_{t-2}, \ldots$. $\theta = (\theta_1', \ldots, \theta_m')'$ (where $\theta_j = (\theta_{1j}, \ldots, \theta pj)', \ j = 1, \ldots, m)$ is the mp-dimensional parameter vector to be estimated and belongs to the stationary area S of model (1) (see [3] for necessary and sufficient conditions for stationarity).

When the TAR model is nonergodic, Pham et al. [12] studied the strong consistency of the LS-estimators of the parameters.

When the TAR model is stationary ergodic and the thresholds are known, the LS-estimators are strongly consistent and asymptotically normal [5,11]. However, in practice, the thresholds might be unknown and need to be estimated. In the case when the thresholds are unknown, Petruccelli [10] proved the strong consistency of the conditional LS-estimator of the threshold for the case $p = d = 1, \ m = 2$. Chan [2] derived the strong consistency and the limiting distribution of the conditional LS-estimator of model (1) for the case of arbitrary p and $m = 2$. In his paper the threshold parameter is unknown and it is shown that the estimator of the threshold parameter is N consistent and its limiting distribution is related to a compound Poisson process. But it is still not known what is the exact limiting distribution. Furthermore, in [2] it was assumed that the autoregressive function was discontinuous. Recently, Chan and Tsay [4] investigated the limiting properties of the conditional least squares estimator for a continuous TAR model.

It was pointed out in earlier papers that in the observed data of time-series models there may be some outlier points quite often and the LAD-estimation is more robust and efficient against the outliers than the LS-estimation. Thus it is of practical importance to study the asymptotic properties of LAD-estimation for time series models. In [18] we investigated the asymptotic normality of LAD-estimators of stationary linear autoregressive models. However, as far as we know, there are few works considering this problem for nonlinear TAR models. Koul [8] obtained the asymptotic normality of LAD-estimation for TAR(1) model with known threshold and $m = 2$.

In this exposition we first discuss the case when the thresholds of TAR(p) model (1) are known and obtain the strong consistency and the asymptotic normality of LAD-estimator of the parameter vector θ under some regularity conditions similar to that of LS-estimators. Then we assume that the threshold is unknown and the autoregressive function is discontinuous for model (1) with $m = 2$ and briefly establish the N consistency and the limiting distribution of the estimator of the threshold. We will not study the threshold-unknown case for arbitrary m. Generally speaking, the case $m > 2$ is rare in practice and much difficult to tackle (see [15,16]).

We can see from Section 3 that the $m = 2$ case is already very complicated. Anyway, it is beyond the scope of this paper.

Namely, we suppose that there are N observations $X_1, \ldots, X_N$ from model (1) and we use the method of LAD-estimation to estimate the true parameter θ^0. The estimator, denoted by θ^N, is the optimal solution of the following minimizing problem:

$$\min_{\theta \in S} \sum_{t=p+1}^{N} \left| X_t - \sum_{j=1}^{m} \sum_{i=1}^{p} \theta_{ij} X_{t-i} H_j(X_{t-d}) \right|. \tag{2}$$

Our main goal is to show that under some mild conditions θ^N converges to θ^0 almost surely and $N^{\frac{1}{2}}(\theta^N - \theta^0)$ converges in distribution to a normal random vector. These results are stated in Theorems 2.1.1 and 2.2.1. Essentially problem (2) is a stochastic optimization problem. It is natural to use optimization theory to get our desired results. In Section 3, we examine the asymptotic properties of the estimators in the case that $m = 2$ and the threshold is unknown. The proofs of the theorems in Section 2 are presented in the appendix.

2. Asymptotics of the least absolute deviation estimator

2.1. Strong consistency of θ^N

In this subsection we prove the strong consistency of the LAD-estimator θ^N. Note that

$$X_t = \sum_{j=1}^{m} \sum_{i=1}^{p} \theta_{ij}^0 X_{t-i} H_j(X_{t-d}) + e_t. \tag{3}$$

Substituting (3) into problem (2), we get an equivalent minimizing problem

$$\min_{\theta \in S} N^{-1} \sum_{t=p+1}^{N} \left\{ \left| e_t - \sum_{j=1}^{m} \sum_{i=1}^{p} (\theta_{ij} - \theta_{ij}^0) X_{t-i} H_j(X_{t-d}) \right| - |e_t| \right\}. \tag{4}$$

Obviously, the LAD-estimator, θ^N, is the optimal solution of (4). Denote the objective function of (4) by $\widetilde{F_N}(e, \theta)$, where $e = (e_1, \ldots, e_N)'$.

The reason why we choose this special form of problem (4) to replace program (2) can be completely figured out by the proof of Theorem 2.1.1 in the appendix. Here we briefly discuss the technique which is basically different from the classical one. We do not prove directly the strong consistency of θ^N. Actually, we convert the consistency problem into the convergence of the optimal solution of the mathematical programming. First, we show that θ^0 is the optimal solution of the mathematical program $\min_{\theta \in S} \tilde{G}(\theta)$, where

$$\tilde{G}(\theta) = \lim_{N \to \infty} N^{-1} \sum_{t=p+1}^{N} E\left\{ \left| e_t - \sum_{j=1}^{m} \sum_{i=1}^{p} (\theta_{ij} - \theta_{ij}^0) X_{t-i} H_j(X_{t-d}) \right| - |e_t| \right\}. \quad (5)$$

Then we prove that the objective function $\widetilde{F_N}(e, \theta)$ of program (4) converges almost surely to the objective function $\tilde{G}(\theta)$ of program (5). Hence the almost sure convergence of θ^N to θ^0 can be established along this line (see the proof of Theorem 2.1.1 in the appendix). That is why we substitute problem (4) for problem (2).

Denote the $mp \times mp$ matrix

$$C = \begin{pmatrix} C_1 & 0 & \dots & 0 \\ 0 & C_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & C_m \end{pmatrix},$$

$$C_j = \begin{pmatrix} EX_p^2 H_j(X_{p+1-d}) & EX_p X_{p-1} H_j(X_{p+1-d}) & \dots & EX_p X_1 H_j(X_{p+1-d}) \\ EX_p X_{p-1} H_j(X_{p+1-d}) & EX_{p-1}^2 H_j(X_{p+1-d}) & \dots & EX_{p-1} X_1 H_j(X_{p+1-d}) \\ \vdots & \vdots & \vdots & \vdots \\ EX_p X_1 H_j(X_{p+1-d}) & EX_{p-1} X_1 H_j(X_{p+1-d}) & \dots & EX_1^2 H_j(X_{p+1-d}) \end{pmatrix},$$

$j = 1, \dots, m.$

We assume the following regularity conditions

Condition 1. Model (1) is strictly stationary, having finite fourth moments and the stationary finite-dimensional distribution of $(X_1, \dots, X_p)$ admits a density function h and the matrix C is positive definite.

Condition 2. $\{e_t\}$ is an i.i.d. random sequence with median zero, finite variance and the density function $g(\cdot)$ of e_t is continuously differentiable at zero, $g(0) > 0$.

Condition 3. $\{Y_t\}$ has a unique invariant measure $\pi(\cdot)$ such that there exists $K, \rho < 1$, for any $x \in R^p$ and any $n, |P^n(x, \cdot) - \pi(\cdot)| \leqslant K(1 + ||x||)\rho^n$, where $Y_t = (X_{t-1}, \dots, X_{t-p})', P^n(x, A)$ ($x \in R^p$, A is a Borel set) stands for the n-step transition probability of the Markov chain $\{Y_t\}$ and $|\cdot|$ and $||\cdot||$ denote the total variation norm and the Euclidean norm, respectively.

Reviewing the literature on the LS-estimation for TAR models, we see that similar conditions as conditions 1–3 were also imposed (cf. [2]) to study the properties of LS-estimators in TAR models.

Lemma 2.1.1. *Assume that conditions* 1 *and* 2 *hold, then there exists an open neighborhood* U *of* θ^0 *such that* θ^0 *is the unique local optimal solution of the program* (5).

The first main result of this paper is the following

Theorem 2.1.1. *Assume that* X_1 *is a bounded random variable and conditions* 1–3 *hold, then* $\theta^N \to \theta^0$ *almost surely as* $n \to \infty$.

Theoretically speaking, the first assumption in Theorem 2.1.1 is very restrictive. However, in practical situation, we can always get the bounded observations.

2.2. Limiting distribution of $N^{1/2}(\theta^N - \theta^0)$

In this subsection we turn to establish the asymptotic normality of LAD-estimator of TAR model (1). Introducing a new optimization vector

$$v = (v_{11}, \dots, v_{p1}, \dots, v_{1m}, \dots, v_{pm})' = N^{1/2}(\theta - \theta^0).$$

Here the definition of v is rational, see Lemma 2.2.2. Then we can rewrite (2) as

$$\min_{v \in V} \sum_{t=p+1}^{N} \left\{ \left| e_t - N^{-1/2} \sum_{j=1}^{m} \sum_{i=1}^{p} v_{ij} X_{t-i} H_j(X_{t-d}) \right| - |e_t| \right\}, \tag{6}$$

where $V = \{v = N^{1/2}(\theta - \theta^0)\colon \theta \in S\}$. Denote by $F_N(e, v)$ and v^N the objective function and the optimal solution of (6). It is clear that $v^N = N^{1/2}(\theta^N - \theta^0)$. As stated in Section 2.1, in order to show the asymptotic normality of $N^{1/2}(\theta^N - \theta^0)$ by our special technique, we should use program (6) to replace program (2). The main idea for deriving the limiting distribution of v^N is as follows. First, we try to find a function $G(\eta, v)$ such that $F_N(e, v) \to G(\eta, v)$ in distribution. Thereafter we show that the optimal solution v^N of (6) converges weakly to the optimal solution v^* of $\min_{v \in V} G(\eta, v)$. As we will see from Lemma 2.2.1, $G(\eta, v)$ is a stochastic quadratic function. Thus the distribution of v^* can be found easily. This is the desired limiting distribution of $N^{1/2}(\theta^N - \theta^0)$.

The way adapted here was also used by Prakasa Rao [13], Wang [17] and Wang and Wang [18] to study the asymptotic behavior of LS-estimators and LAD-estimators for nonlinear regression models and linear time series models respectively.

The following lemma gives the limit function of $F_N(e, v)$.

Lemma 2.2.1. *Under conditions* 1 *and* 2, *for any fixed* $v, F_N(e, v)$ *converges in distribution to*

$$G(\eta, v) = -v'\eta + g(0)v'Cv,$$

where η *is a mp-dimensional normal random vector with zero mean and covariance matrix* C.

The following lemma is important for the proofs of Theorems 2.1.1 and 2.2.1 and may be of independent interest on its own.

Lemma 2.2.2. *Under conditions* 1 *and* 2, *we have that* $v^N = N^{1/2}(\theta^N - \theta^0)$ *is bounded in probability*.

With Lemmas 2.2.1 and 2.2.2 we can show

Theorem 2.2.1. *Under conditions* 1 *and* 2, *we have*

$$N^{1/2}(\theta^N - \theta^0) \to v^*$$

in distribution, *where* v^* *is mp-dimensional normal vector*, *centered at the origin*, *with covariance matrix* $\Sigma = (4g^2(0)C)^{-1}$.

3. Asymptotics of the estimator of the threshold

In this section we only consider the case $m = 2$. Then (1) becomes

$$X_t = \begin{cases} \theta_{11}X_{t-1} + \ldots + \theta_{p1}X_{t-p} + e_t, & \text{if} \quad X_{t-d} \leqslant r, \\ \theta_{12}X_{t-1} + \ldots + \theta_{p2}X_{t-p} + e_t, & \text{if} \quad X_{t-d} > r. \end{cases} \tag{7}$$

Here we suppose r is unknown and $H_1(X_{t-d}) = I(X_{t-d} \leqslant r), H_2(X_{t-d}) = I(X_{t-d} > r)$. The true value of r is r^0. The LAD-estimator θ^N can be derived by two steps. First for any fixed r, we minimize

$$\min_{\theta \in S} \overline{F_N}(\theta, r) = \sum_{t=p+1}^{N} \left| X_t - \sum_{i=1}^{p} \theta_{i1} X_{t-i} I(X_{t-d} \leqslant r) - \sum_{i=1}^{p} \theta_{i2} X_{t-i} I(X_{t-d} > r) \right|. \tag{8}$$

After we get the optimal solution $\theta^N(r)$ and the optimal value $\overline{F_N}(\theta^N(r), r)$, we solve the following problem:

$$\min_{r} \overline{F_N}(\theta^N(r), r) \tag{9}$$

and obtain the optimal solution r^N as the estimator of the threshold. Practically, when analyzing the data $X_1, \ldots, X_N$ observed from model (7), we usually rearrange them as $X_{(1)}, \ldots, X_{(N)}$, the order statistics of the sample. Then we choose the threshold r from the points of 25%, 35%, 50%, 75%, etc. significance levels in this order statistics sequence and compute the optimal values $\overline{F_N}(\theta^N(r), r)$ using these thresholds. Finally, the threshold involved in the smallest one among these optimal values is the estimator of the threshold r (see [16] for details).

We first verify the conclusion that r^N is a consistent estimator of r^0.

Theorem 3.1. *Under the conditions assumed in Theorem* 2.1.1 *and condition* 4 *in Chan* [2], $N(r^N - r^0)$ *is bounded in probability*.

Proof. From Theorem 2.1.1, θ^N is strongly consistent, then the parameter space can be restricted to a neighborhood of θ^0. First we consider the case $p = d = 1$. In order to show that $N(r^N - r^0)$ is bounded in probability, it suffices to prove that for any $\varepsilon > 0$, there exists some $0 < \delta < 1$ and $K > 0$ such that $|\theta_{1j} - \theta_{1j}^0| < \delta,\ j = 1, 2$, $|r - r^0| > K/N$ and with probability greater than $1 - \varepsilon$

$$\overline{F_N}(\theta_{11}, \theta_{12}, r) - \overline{F_N}(\theta_{11}, \theta_{12}, r^0) > 0.$$

First we consider the case $r \geqslant r^0$. In fact, from (7) and (8), we obtain

$$\begin{aligned}
&\overline{F_N}(\theta_{11}, \theta_{12}, r) - \overline{F_N}(\theta_{11}, \theta_{12}, r^0) \\
&\quad = \sum_{t=p+1}^{N} \{|e_t - (\theta_{11} - \theta_{11}^0)X_{t-1}| - |e_t|\} I(r^0 < X_{t-1} \leqslant r) \\
&\qquad + \sum_{t=p+1}^{N} \{|e_t| - |e_t - (\theta_{12} - \theta_{12}^0)X_{t-1}|\} I(r^0 < X_{t-1} \leqslant r).
\end{aligned}$$

Moreover, by the convexity of the absolute value function and the fact that $\text{sign}(u)$ is one of subgradients of $|\cdot|$ at u, we have

$$|e_t - (\theta_{11} - \theta_{11}^0)X_{t-1}| - |e_t| \geqslant -(\theta_{11} - \theta_{11}^0)X_{t-1}\,\text{sign}(e_t)$$

and

$$|e_t| - |e_t - (\theta_{12} - \theta_{12}^0)X_{t-1}| \geqslant (\theta_{12} - \theta_{12}^0)X_{t-1}\,\text{sign}\,(e_t - (\theta_{12} - \theta_{12}^0)X_{t-1}),$$

where

$$\text{sign}\,(u) = \begin{cases} 1, & u > 0, \\ 0, & u = 0, \\ -1, & u < 0. \end{cases}$$

It follows that

$$\begin{aligned}
&\overline{F_N}(\theta_{11}, \theta_{12}, r) - \overline{F_N}(\theta_{11}, \theta_{12}, r^0) \\
&\quad \geqslant \sum_{t=p+1}^{N} \{-(\theta_{11} - \theta_{11}^0)X_{t-1}\,\text{sign}\,(e_t) I(r^0 < X_{t-1} \leqslant r)\} \\
&\qquad + \sum_{t=p+1}^{N} \{(\theta_{12} - \theta_{12}^0)X_{t-1}\,\text{sign}\,(e_t - (\theta_{12} - \theta_{12}^0)X_{t-1}) I(r^0 < X_{t-1} \leqslant r)\}. \quad (10)
\end{aligned}$$

The RHS of (10) is bounded in absolute value by $c_1 \sum_{t=p+1}^{N} I(r^0 < X_{t-1} \leqslant r)$ for some constant c_1 independent of N and $r^0 < r \leqslant r^0 + \delta$.

For any $\varepsilon > 0, c_2 > 0$, there exists $K > 0$ such that

$$P\left\{ \sup_{K/N < r - r^0 \leqslant \delta} \left| \frac{\sum_{t=p+1}^{N} I(r^0 < X_{t-1} \leqslant r)}{NE\{I(r^0 < X_1 \leqslant r)\}} - 1 \right| < c_2 \right\} > 1 - \varepsilon.$$

(This result is proved in Claim 2 of Proposition 1 in [2].)

Let c_2 be chosen so that $c_1(1-c_2)>0$. Then for $K/N<r-r^0\leqslant\delta$,

$$P\{\overline{F_N}(\theta_{11},\theta_{12},r)-\overline{F_N}(\theta_{11},\theta_{12},r^0)\geqslant c_1(1-c_2)NE\{I(r^0<X_1\leqslant r)\}>0\}>1-\varepsilon.$$

That is, $r-r^0>K/N$ implies that $\overline{F_N}(\theta_1,\theta_2,r)>\overline{F_N}(\theta_1,\theta_2,r^0)$ with probability greater than $1-\varepsilon$.

The case of $r<r^0$ is similar. This is the proof for the case $p=d=1$.

For the general case, by Condition 4 in [2], there exists $Y^*=(r_{p-1},r_{p-2},\ldots,r_0)'$ such that $(\theta_1^0-\theta_2^0)'Y^*\neq 0$ and $r_{p-d}=r^0$. Then there exists $\varpi>0$ such that $(\theta_1^0-\theta_2^0)'Y$ is bounded away from 0 for all Y such that $||Y-Y^*||\leqslant\varpi$. Then the preceding proof would go through only if we replace each $I(r^0<X_1\leqslant r)$ by $I(r^0<X_{t-d}\leqslant r;||Y_t-Y^*||\leqslant\varpi)$, where $Y_t=(X_{t-1},\ldots,X_{t-p})'$. In detail, again we have

$$\begin{aligned}&\overline{F_N}(\theta_1,\theta_2,r)-\overline{F_N}(\theta_1,\theta_2,r^0)\\&\quad\geqslant\sum_{t=p+1}^{N}\{(\theta_2-\theta_2^0-\theta_1+\theta_1^0)'Y_tI(r^0<X_{t-1}\leqslant r;||Y_t-Y^*||\leqslant\varpi)\}.\end{aligned}$$

Then again for $K/N<r-r^0\leqslant\delta$, we get

$$P\{\overline{F_N}(\theta_1,\theta_2,r)-\overline{F_N}(\theta_1,\theta_2,r^0)>0\}>1-\varepsilon.$$

Hence we have

$$r^N=r^0+O(1/N)\tag{11}$$

in probability for arbitrary $p, 1\leqslant d\leqslant p$. This completes the proof of Theorem 3.1. □

Remark. From the results of Theorems 2.1.1 and 3.1, we know that $\theta^N(r)\to\theta^0$ almost surely for any r satisfying $|r-r^0|\leqslant K/N$. This means that

$$\theta^N(r^N)\to\theta^0$$

almost surely. On the other hand, let $C(r)$ denotes the covariance matrix C of η in Lemma 2.2.1. By Condition 1 and Schwartz inequality, it is easy to check that the matrix $C(r)$ is continuous in r. Therefore $r^N\to r^0$ in probability implies that $C(r^N)\to C(r^0)$ in probability. Now Theorem 2.2.1 yields that

$$N^{1/2}(\theta^N(r^N)-\theta^0)\to v^*$$

in distribution. Then we conclude that the asymptotic behaviors of the LAD-estimator for model (7) with unknown threshold are equal to that in the case when the threshold is known.

As to the limiting behavior of $N(r^N-r^0)$, applying the discussions in Kushner [9], it can be shown that the limiting law of $N(r^N-r^0)$ is related to a compound Poisson process, see the following relation (15).

Now we briefly prove this result. First we define a new variable $w=N(r-r^0)$ and then by (11) we change the objective function of problem (9) to get an equivalent problem

$$\min_{|w| \leqslant K} T_N(w) = \overline{F_N}(\theta^N(r^0 + w/N), r^0 + w/N) - \overline{F_N}(\theta^N(r^0), r^0). \tag{12}$$

Obviously $N(r^N - r^0)$ is the optimal solution of program (12). From Lemma 2.2.2, we obtain

$$\theta_{ij}^N(r^0 + w/N) = \theta_{ij}^0 + N^{-1/2} v_{ij}^N(r^0 + w/N), \quad i = 1, \ldots, p, \quad j = 1, 2$$

and

$$|v_{ij}^N(r^0 + w/N)| \leqslant c_3$$

in probability for some positive constant c_3. Moreover, note that if $w \geqslant 0$,

$$\begin{aligned}
&\overline{F_N}(\theta^N(r^0 + w/N), r^0 + w/N) - \overline{F_N}(\theta^N(r^0), r^0) \\
&\quad - (\overline{F_N}(\theta^0, r^0 + w/N) - \overline{F_N}(\theta^0, r^0)) \\
&\quad = \sum_{t=p+1}^{N} \left\{ \left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^N(r^0 + w/N) \right| - \left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^0 \right| \right. \\
&\qquad \left. - \left(\left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i2}^N(r^0) \right| - \left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i2}^0 \right| \right) \right\} \\
&\qquad \times I(r^0 < X_{t-d} \leqslant r^0 + w/N).
\end{aligned} \tag{13}$$

Using similar arguments as in the proof of Theorem 3.1, we have

$$\begin{aligned}
&\left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^N(r^0 + w/N) \right| - \left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^0 \right| \\
&\quad = \left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^0 - N^{-1/2}\sum_{i=1}^{p} X_{t-i} v_{i1}^N(r^0 + w/N) \right| - \left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^0 \right| \\
&\quad \leqslant -N^{-1/2}\sum_{i=1}^{p} X_{t-i} v_{i1}^N(r^0 + w/N)\, \mathrm{sign}\left(X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^0 \right. \\
&\qquad \left. - N^{-1/2}\sum_{i=1}^{p} X_{t-i} v_{i1}^N(r^0 + w/N) \right).
\end{aligned}$$

And

$$\begin{aligned}
&\left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i2}^N(r^0) \right| - \left| X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i2}^0 \right| \\
&\quad \geqslant -N^{-1/2}\sum_{i=1}^{p} X_{t-i} v_{i2}^N(r^0)\, \mathrm{sign}\left(X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i2}^0 \right).
\end{aligned}$$

Let $\Delta_i = \max_{1 \leqslant t \leqslant N} X_{t-i} I(r^0 < X_{t-d} \leqslant r^0 + w/N)$, then

$$\begin{aligned}\Delta_i = \max_{1 \leqslant t \leqslant N} \; & X_{t-i}\{I(r^0 < X_{t-d} \leqslant r^0 + w/N, ||Y_t - Y^*|| \leqslant \varpi) \\ & + I(r^0 < X_{t-d} \leqslant r^0 + w/N, ||Y_t - Y^*|| > \varpi)\}.\end{aligned}$$

If $||Y_t - Y^*|| > \varpi$, by Condition 4 in [2], we know that $(\theta_1^0 - \theta_2^0)' Y_t = 0$. Since $\theta_1^0 \neq \theta_2^0$, it holds that $||Y_t|| < \infty$ almost surely. This means that $\Delta_i < \infty$ almost surely.

Hence by (13), Schwartz inequality and the properties of "sign" function, when n big enough, we have

$$\begin{aligned}
&\operatorname{Var}\{T_N(w) - [\overline{F_N}(\theta^0, r^0 + w/N) - \overline{F_N}(\theta^0, r^0)]\} \\
&\leqslant E\Bigg\{ N^{-1/2} \sum_{t=p+1}^{N} \sum_{i=1}^{p} X_{t-i} \Bigg[v_{i2}^N(r^0)\, \operatorname{sign}\Bigg(X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i2}^0 \Bigg) \\
&\quad - v_{i1}^N(r^0 + w/N)\, \operatorname{sign}\Bigg(X_t - \sum_{i=1}^{p} X_{t-i}\theta_{i1}^0 - N^{-1/2} \sum_{i=1}^{p} X_{t-i} v_{i1}^N(r^0 + w/N) \Bigg) \Bigg] \\
&\quad \times I(r^0 < X_{t-d} \leqslant r^0 + w/N) \Bigg\}^2 \\
&\leqslant c_3^2 N^{-1} E\Bigg\{ \sum_{t=p+1}^{N} \sum_{i=1}^{p} X_{t-i} \Bigg[\operatorname{sign}(e_t) \\
&\quad - \operatorname{sign}\Bigg(e_t - N^{-1/2} \sum_{i=1}^{p} X_{t-i} v_{i1}^N(r^0 + w/N) \Bigg) \Bigg] \\
&\quad \times I(r^0 < X_{t-d} \leqslant r^0 + w/N) \Bigg\}^2 \\
&\leqslant c_3^2 N^{-1} E\Bigg\{ \sum_{i=1}^{p} \Delta_i \sum_{t=p+1}^{N} \Bigg[\operatorname{sign}(e_t) - \operatorname{sign}\Bigg(e_t - N^{-1/2} c_3 \sum_{i=1}^{p} \Delta_i \Bigg) \Bigg] \Bigg\}^2 \\
&\leqslant c_3^2 N^{-1} \Bigg\{ E\Bigg\{ \sum_{i=1}^{p} \Delta_i \Bigg\}^4 \Bigg\}^{1/2} \Bigg\{ E\Bigg\{ \sum_{t=p+1}^{N} o_P\Bigg(N^{-1/2} c_3 \sum_{i=1}^{p} \Delta_i \Bigg) \Bigg\}^4 \Bigg\}^{1/2} \\
&\leqslant c_3^2 N^{-1} \Bigg\{ E\Bigg[\sum_{i=1}^{p} \Delta_i \Bigg]^4 \Bigg\}^{1/2} o(N) \\
&\leqslant c_3^2 \Bigg\{ E\Bigg[\sum_{i=1}^{p} \Delta_i \Bigg]^4 \Bigg\}^{1/2} o(1) \\
&\to 0 \quad \text{as} \quad N \to \infty,
\end{aligned}$$

where $o_P(.)$ stands for convergence in probability as $N \to \infty$. We get similar result for the case $w < 0$. Hence Chebyshev's inequality yields

$$\sup_{|w| \leqslant K} |T_N(w) - \{\overline{F_N}(\theta^0, r^0 + w/N) - \overline{F_N}(\theta^0, r^0)\}| \to 0 \tag{14}$$

in probability. Owing to (14), we can approximately write $T_N(w)$ as

$$T_N(w) = \begin{cases} \sum \left\{ \left| e_t + \sum_{i=1}^{p} (\theta_{i2}^0 - \theta_{i1}^0) X_{t-i} \right| - |e_t| \right\} I(r^0 < X_{t-d} \leqslant r^0 + w/N), & w \geqslant 0, \\ \sum \left\{ \left| e_t + \sum_{i=1}^{p} (\theta_{i1}^0 - \theta_{i2}^0) X_{t-i} \right| - |e_t| \right\} I(r^0 + w/N < X_{t-d} \leqslant r^0), & w < 0. \end{cases}$$

Employing weak convergence theory for random processes in [9], it can be shown that under conditions 1–4 in Chan [2],

$$\{T_N(w), w < 0\} \to \{T^{(1)}(w), w < 0\},$$

$$\{T_N(w), w \geqslant 0\} \to \{T^{(2)}(w), w \geqslant 0\}$$

in distribution. Here $\{T^{(1)}(w), w < 0\}$ and $\{T^{(2)}(w), w \geqslant 0\}$ are compound Poisson processes with the distributions of jump being given by the conditional distribution $|e_{p+1} + \sum_{i=1}^{p} (\theta_{i1}^0 - \theta_{i2}^0) X_{p+1-i}| - |e_{p+1}|$ given $X_{p+1-d} = r_-^0$ and the conditional distribution $|e_{p+1} + \sum_{i=1}^{p} (\theta_{i2}^0 - \theta_{i1}^0) X_{p+1-i}| - |e_{p+1}|$ given $X_{p+1-d} = r_+^0$, respectively.

For completeness we outline the proof of the above statement. In fact, by Lemma 3.2 in Ibragimov and Has'minskii [7], it is easy to check the tightness of $\{T^{(1)}(w), w < 0\}$ and $\{T^{(2)}(w), w \geqslant 0\}$. Therefore, to complete the proof, it suffices to show the weak convergence of finite-dimensional distributions of $\{T_N(w), w < 0\}$ and $\{T_N(w), w \geqslant 0\}$. This can be established by following the arguments in the proof of Theorem 2 in [2]. The main difference is that we replace J_N^ε in Chan [2] by

$$J_N^\varepsilon = \left(\left| e_t + \sum_{i=1}^{p} (\theta_{i2}^0 - \theta_{i1}^0) X_{t-i} \right| - |e_t| \right) I(r^0 < X_{t-d} \leqslant r^0 + w/N).$$

Now by the weak convergence of $\{T_N(w), w < 0\}, \{T_N(w), w \geqslant 0\}$, and by adapting the similar method used in the proof of Theorem 2.2.1, we get the conclusion that

$$N(r^N - r^0) \to w^* \tag{15}$$

in distribution, where w^* is the global minimizer of the function $T^{(1)}(w)I(w < 0) + T^{(2)}(w)I(w \geqslant 0)$.

Acknowledgments

This paper is partly supported by NSFC and NSFJS. The authors express their sincere gratitude for the referee's helpful comments and suggestions that improved the early version of the paper.

Appendix

Proof of Lemma 2.1.1. What we should do is to compute

$$\sum_{t=p+1}^{N} E\left\{\left|e_t - \sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^0)X_{t-i}H_j(X_{t-d})\right| - |e_t|\right\}.$$

By condition 2, for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $|u| < \delta$,

$$g(u) = g(0) + O(\varepsilon).$$

Let $U = \{\theta\colon \|\theta - \theta^0\| \leqslant \delta\} \subset S$ denotes the open neighborhood of θ^0. Then for any $\theta \in U$, by the independence of e_t and $(X_{t-1}, \ldots, X_{t-p})$, we have

$$\begin{aligned}
&E\left\{\left|e_t - \sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^0)X_{t-i}H_j(X_{t-d})\right| - |e_t|\right\}\\
&= 2\int_{u>0,u-\tau<0}(\tau-u)g(u)h(v_1,\ldots,v_p)\,du\,dv_1\ldots dv_p\\
&\quad + 2\int_{u<0,u-\tau>0}(u-\tau)g(u)h(v_1,\ldots,v_p)\,du\,dv_1\ldots dv_p\\
&= 2\int_{-\infty}^{\infty}\left\{\int_0^{\tau}(\tau-u)g(u)du\right\}h(v_1,\ldots,v_p)dv_1\ldots dv_p\\
&= 2\int_{-\infty}^{\infty}\frac{1}{2}\tau^2[g(0)+O(\varepsilon)]h(v_1,\ldots,v_p)dv_1\ldots dv_p\\
&= g(0)(\theta-\theta^0)'C(\theta-\theta^0) + O(\varepsilon(\theta-\theta^0)'(\theta-\theta^0)),
\end{aligned}$$

where $\tau = \sum_{j=1}^{m}\sum_{i=1}^{p} v_iH_j(v_d)(\theta_{ij}-\theta_{ij}^0)$. Let $\varepsilon \to 0$, we obtain

$$\begin{aligned}
&\lim_{N\to\infty} N^{-1}\sum_{t=p+1}^{N} E\left\{\left|e_t - \sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^0)X_{t-i}H_j(X_{t-d})\right| - |e_t|\right\}\\
&= g(0)(\theta-\theta^0)'C(\theta-\theta^0). \qquad \text{(A.1)}
\end{aligned}$$

Since $g(0) > 0$ and C is positive definite, then θ^0 is the unique local optimal solution of $\min_{\theta\in S}\tilde{G}(\theta)$. □

Proof of Theorem 2.1.1. Set $\eta_t = \left|e_t - \sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^0)X_{t-i}H_j(X_{t-d})\right| - |e_t| + \sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^0)X_{t-i}H_j(X_{t-d})\,\mathrm{sign}(e_t)$. Then $\widetilde{F_N}(e,\theta) = N^{-1}\sum_{t=p+1}^{N}\left\{\eta_t - \sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^0)X_{t-i}H_j(X_{t-d})\,\mathrm{sign}(e_t)\right\}$.

At first step we prove that for any fixed θ,

$$\widetilde{F_N}(e,\theta) \to \tilde{G}(\theta)$$

almost surely. It suffices to show that

$$N^{-1}\sum_{t=p+1}^{N}\{\eta_t - E(\eta_t)\}\to 0, \tag{A.2}$$

$$N^{-1}\sum_{t=p+1}^{N}\sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\,\text{sign}\,(e_t)\to 0 \tag{A.3}$$

almost surely. Using the similar arguments as in the proof of Theorem 3.1, we have

$$\begin{aligned}&\left|e_t-\sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\right|-|e_t|\\&\quad\geqslant -\sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\,\text{sign}\,(e_t)\end{aligned}$$

and

$$\begin{aligned}&|e_t|-\left|e_t-\sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\right|\\&\quad\geqslant \sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\,\text{sign}\left(e_t-\sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\right)\end{aligned}$$

This means that

$$\begin{aligned}0\leqslant\eta_t\leqslant&\sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\\&\times\left\{\text{sign}(e_t)-\text{sign}\left(e_t-\sum_{j=1}^{m}\sum_{i=1}^{p}(\theta_{ij}-\theta_{ij}^{0})X_{t-i}H_j(X_{t-d})\right)\right\}.\end{aligned} \tag{A.4}$$

Then for any fixed θ and some constant c_4, we have

$$0\leqslant\eta_t\leqslant c_4\sum_{i=1}^{p}X_{t-i}. \tag{A.5}$$

Since $\{Y_t\}$ is a stationary Markov process, by condition 3, we have that $\left\{\sum_{i=1}^{p}X_{t-i}\right\}$ is ϕ-mixing with $\phi_n=c_5\rho^n$, where c_5 is a positive constant (cf. p. 167, Example 2 of Billingsley [1]).

Then following from a result in Billingsley [1](cf. p. 172, Lemma 4), we arrive at

$$E\left(\sum_{t=p+1}^{N}\sum_{i=1}^{p}X_{t-i}\right)^4\leqslant c_6N^2$$

with positive constant c_6.

This, combined with (A.5) and Chebyshev's inequality, now yields

$$\sum_{N=1}^{\infty} P\left\{\left|N^{-1}\sum_{t=p+1}^{N}\{\eta_t - E(\eta_t)\}\right| > \varepsilon\right\}$$
$$\leqslant c_4^4 \sum_{N=1}^{\infty} \frac{E\left(\sum_{t=p+1}^{N}\sum_{i=1}^{p} X_{t-i}\right)^4}{N^4\varepsilon^4} \leqslant c_7 \sum_{N=1}^{\infty} N^{-2} < \infty,$$

where c_7 is a positive constant. Then (A.2) follows from Borel-Cantelli's lemma. Similar arguments yield (A.3).

Now we have showed that for any fixed $\theta \in S$, the objective function $\widetilde{F_N}(e,\theta)$ of program (4) converges almost surely to the objective function $\tilde{G}(\theta)$ of program (5). We are in the position to prove the almost sure convergence of θ^N to θ^0.

Let Ω denotes the underlying probability space and let $\Omega - A$ be a measure-zero set in Ω. Assume that $\widetilde{F_N}(e,\theta)(\omega) \to \widetilde{G}(\theta)$ for every $\omega \in A$.

Note that the closure of the stationary region, cl S, is a compact subset of R^{mp}, hence for every $\omega \in A, \{\theta^N(\omega)\}$ always contains convergent subsequence in cl S. Without loss of generality we suppose that for every $\omega \in A, \{\theta^N(\omega)\}$ converges to $a_\omega \in \text{cl } S, a_\omega \neq \theta^0$. By the fact that $N^{1/2}(\theta^N - \theta^0)$ is bounded in probability (this result will be proved later, see Lemma 2.2.2 which is independent of Theorem 2.1.1), we see that for each $\omega \in A$, there exists a $\varepsilon > 0$ such that $\omega \in B_\varepsilon$ and $\theta^N(\omega) \in U$ when N is large enough, where $B_\varepsilon \subset \Omega$ and $P(B_\varepsilon) \geqslant 1-\varepsilon$, U is an open neighborhood of θ^0 defined in the proof of Lemma 2.1.1. This implies that $a_\omega \in U$.

By Lemma 2.1.1, $\{\theta^0\}$ is the unique local minimizer of $\tilde{G}(.)$, then $\tilde{G}(a_\omega) > \tilde{G}(\theta^0) = 0$. On the other hand, since $\{\theta^N\}$ is the optimal solution of $\widetilde{F_N}(e,\theta)$, by the continuity of the function $\tilde{G}$, we have

$$\begin{aligned}\tilde{G}(a_\omega) &= g(0)(a_\omega - \theta^0)' C(a_\omega - \theta^0)\\ &= \lim_{N\to\infty} \tilde{G}(\theta^N(\omega)) = \lim_{N\to\infty} \widetilde{F_N}(e,\theta^N(\omega))\\ &\leqslant \lim_{N\to\infty} \widetilde{F_N}(e,\theta^0) = \tilde{G}(\theta^0) = 0.\end{aligned}$$

Here $\lim_{N\to\infty}$ means almost sure convergence. This contradiction proves that $\theta^N \to \theta^0$ almost surely as $N \to \infty$. □

Proof of Lemma 2.2.1. Since $\eta_t = \left|e_t - N^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{p} v_{ij}X_{t-i}H_j(X_{t-d})\right| - |e_t| + N^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{p} v_{ij}X_{t-i}H_j(X_{t-d})\,\text{sign}\,(e_t)$. From (A.4), we have

$$0 \leqslant \eta_t \leqslant N^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{p} v_{ij}X_{t-i}H_j(X_{t-d})$$
$$\times\left\{\text{sign}\,(e_t) - \text{sign}\left(e_t - N^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{p} v_{ij}X_{t-i}H_j(X_{t-d})\right)\right\}.$$

Then by Schwartz inequality and the properties of "sign" function, we get that, for any fixed v,

$$\begin{aligned}
&\mathrm{Var}(\eta_t)\\
&\leqslant N^{-1}E\left\{\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}H_j(X_{t-d})o_P\left(N^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}H_j(X_{t-d})\right)\right\}^2\\
&=o(1)N^{-2}E\left\{\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}H_j(X_{t-d})\right\}^4\\
&\leqslant o(1)N^{-2}E\left\{\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}\right\}^4\\
&=o(1)N^{-2}EX_1^4.
\end{aligned}$$

Again by Schwartz inequality, we obtain

$$\mathrm{Var}\left(\sum_{t=p+1}^{N}\eta_t\right)\leqslant N^2\,\mathrm{Var}\,(\eta_t)\leqslant N^2N^{-2}EX_1^4o(1)\to 0$$

as $N\to\infty$. Now Chebyshev's inequality yields

$$F_N(e,v)-F_N^*(e,v)\to 0$$

in probability as $N\to\infty$, where

$$\begin{aligned}
F_N^*(e,v)=&-N^{-1/2}\sum_{t=p+1}^{N}\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}H_j(X_{t-d})\,\mathrm{sign}\,(e_t)\\
&+\sum_{t=p+1}^{N}E\left[\left|e_t-N^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}H_j(X_{t-d})\right|-|e_t|\right].
\end{aligned}$$

Hence for our purpose it suffices to derive the limit of $F_N^*(e,v)$. First, by the independence of e_t and X_s $(s<t)$ and $E\{\mathrm{sign}(e_t)\}=0$, the central limit theorem for martingales as given in Corollary 3.1 in [6] yields

$$N^{-1/2}\sum_{t=p+1}^{N}\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}H_j(X_{t-d})\,\mathrm{sign}\,(e_t)\to v'\eta \tag{A.6}$$

in distribution.

Moreover, (A.1) can be rewritten as

$$\lim_{N\to\infty}\sum_{t=p+1}^{N}E\left\{\left|e_t-N^{-1/2}\sum_{j=1}^{m}\sum_{i=1}^{p}v_{ij}X_{t-i}H_j(X_{t-d})\right|-|e_t|\right\}=g(0)v'Cv.$$

This, together with (A.6), completes the proof. □

Proof of Lemma 2.2.2. Note that $v = 0$ is a feasible solution of program (6) and v^N is the optimal solution of (6), that is,

$$0 \geqslant F_N(e, v^N) - F_N(e, 0) = F_N^*(e, v^N) + o_P(1)$$

Then it suffices to verify the following

Claim. *For any* $\varepsilon > 0$, *there exist positive constants* M_ε, δ *such that with probability greater than* $1 - \varepsilon$, *when* $||v|| > M_\varepsilon$, *we have* $F_N^*(e, v) \geqslant \delta$ *for N large enough. Clearly in this case* $F_N(e, v) > 0$ *for N large enough.*

First, by Lemma 2.2.1, we know that

$$F_{N1}^*(e, v) \to -v'\eta$$

in distribution and

$$F_{N2}^*(v) = g(0)v'Cv + o(1),$$

where

$$F_{N1}^*(e, v) = -N^{-1/2} \sum_{t=p+1}^{N} \sum_{j=1}^{m} \sum_{i=1}^{p} v_{ij} X_{t-i} H_j(X_{t-d}) \operatorname{sign}(e_t),$$

$$F_{N2}^*(e, v) = \sum_{t=p+1}^{N} E\left[\left| e_t - N^{-1/2} \sum_{j=1}^{m} \sum_{i=1}^{p} v_{ij} X_{t-i} H_j(X_{t-d}) \right| - |e_t| \right].$$

By condition 1, we know that $C = \mathrm{Var}(\eta)$ is positive definite, then let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{mp} > 0$ be its characteristic roots and $l_1, \ldots, l_{mp}$ be its standardized characteristic vectors, we have

$$\mathrm{Var}\,(F_N^*(e, v)) \to v'Cv = \sum_{i=1}^{mp} \lambda_i a_i^2 \leqslant \lambda_1 ||v||^2$$

here the vector v can be written as $v = \sum_{i=1}^{mp} a_i l_i$. Hence for N large enough and any $\varepsilon > 0$, by Chebyshev's inequality, we have

$$P\{|F_N^*(e, v) - E\{F_N^*(e, v)\}| \leqslant ||v|| b_\varepsilon\} \geqslant 1 - \varepsilon \tag{A.7}$$

by choosing $b_\varepsilon = (2\lambda_1/\varepsilon)^{1/2}$. (A.7) implies that

$$P\{F_N^*(e, v) \geqslant -||v|| b_\varepsilon + g(0)v'Cv + o(1)\} \geqslant 1 - \varepsilon. \tag{A.8}$$

Again since $g(0)C$ is positive definite, still define $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{mp} > 0$ and $l_1, \ldots, l_{mp}$ be characteristic roots and standardized characteristic vectors of $g(0)C$, we have

$$-||v|| b_\varepsilon + g(0)v'Cv = -||v|| b_\varepsilon + \sum_{i=1}^{mp} \lambda_i a_i^2 \geqslant \lambda_{mp} ||v||^2 - ||v|| b_\varepsilon.$$

For N large enough, let M_ε, δ be chosen so that

$$\inf_{||v|| > M_\varepsilon} \{-||v|| b_\varepsilon + g(0)v'Cv + o(1)\} \geqslant \delta.$$

This, in turn, when using (A.8), means that

$$\begin{aligned} &P\left\{\inf_{||v||>M_\varepsilon}\{F_N^*(e,v)\}\geqslant\delta\right\}\\ &\geqslant P\left\{\inf_{||v||>M_\varepsilon}\{F_N^*(e,v)\}\geqslant\inf_{||v||>M_\varepsilon}\{-||v||b_\varepsilon+g(0)v'Cv+o(1)\}\right\}\\ &\geqslant 1-\varepsilon \end{aligned}$$

and hence the validity of Lemma 2.2.2. □

Proof of Theorem 2.2.1. Here we give only the sketch of the proof, since the details are quite similar to that of Theorem 2 in [17].

The main idea is that we view v^N as an image of the stochastic processes $\{F_N(e,v), v\in V\}$ under the minimizing operator H and use continuous mapping theorem in weak convergence theory of probability measures (see [1]). Here H is a minimizing operator on C_V (where C_V is the space of all continuous functions defined on V) such that the image of f in C_V under H is the optimal solution of

$$\begin{aligned} &\min \quad f(v)\\ &\text{s.t.} \quad v\in V. \end{aligned}$$

By Lemmas 2.2.1 and 2.2.2, the optimization theory, and the theory of stochastic processes, we can show that the sequence of stochastic processes $\{F_N(e,v), v\in V\}$ converges weakly to the stochastic process $\{G(\eta,v), v\in V\}$ and that $H(\cdot)$ is a continuous mapping in the sense that $H(f_n)\to H(f)$ for $f_n\to f$ if $H(f)$ is unique. Then by continuous mapping theorem (see Theorem 5.1 in [1]), we obtain

$$H\{F_N(e,v), v\in V\}\to H\{G(\eta,v), v\in V\}$$

in distribution. This implies that

$$v^N\to v^*$$

in distribution, where v^* is the unique solution of the equation $\frac{\partial}{\partial v}G(\eta,v)=0$, then $v^*=(2g(0)C)^{-1}\eta$. Thus we get the desired conclusion. □

References

[1] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.

[2] K.S. Chan, Consistency and limiting distribution of the least squares estimation of a threshold autoregressive model, Ann. Statist. 21 (1993) 520–533.

[3] K.S. Chan, H. Tong, On the use of the determining Lyapunov function for the ergodicity of stochastic difference equations, Adv. in Appl. Probab. 17 (1985) 666–678.

[4] K.S. Chan, R.S. Tsay, Limiting properties of the least squares estimator of a continuous threshold autoregressive model, Biometrika 85 (1998) 413–426.

[5] W. Fan, Asymptotic properties of estimators of parameters in TAR models, Chinese J. Appl. Probab. 2 (1986) 259–266.

[6] P. Hall, C.C. Heyde, Martingale Limit Theory and its Applications, Academic Press, New York, 1980.

[7] I.A. Ibragimov, R.Z. Has'minskii, Statistical Estimation: Asymptotic Theory, Springer, New York, 1981.
[8] H.L. Koul, Asymptotics of some estimators and sequential residual empiricals in nonlinear time series, Ann. Statist. 24 (1996) 380–404.
[9] H.J. Kushner, Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory, MIT Press, Cambridge, MA, 1984.
[10] J.D. Petruccelli, On the consistency of least squares estimator of a threshold AR(1) model, J. Time Ser. Anal. 7 (1986) 269–278.
[11] J.D. Petruccelli, S.W. Woolford, A threshold AR(1) model, J. Appl. Probab. 21 (1984) 270–286.
[12] D.T. Pham, K.S. Chan, H. Tong, Strong consistency of the least squares estimator for a nonergodic threshold autoregressive model, Statist. Sinica 1 (1991) 361–369.
[13] B.L.S. Prakasa Rao, Asymptotic Theory of Statistical Inference, Wiley, New York, 1987.
[14] H. Tong, Discuss of a paper by A.J. Lawrance and N.T. Kottegoda, J. Roy. Statist. Soc. A 140 (1977) 34–35.
[15] H. Tong, in: Threshold Models in Nonlinear Time Series Analysis, Lecture Notes in Statistics, Vol. 21, Springer, Berlin, 1983.
[16] H. Tong, Non-linear Time Series: A Dynamical System Approach, Oxford Statistical Science Series, Vol. 6, Oxford, 1990.
[17] J. Wang, Asymptotic normality of least absolute deviationestimators in nonlinear regression, J. Multivariate Anal. 54 (1995) 227–238.
[18] L. Wang, J. Wang, Asymptotic distribution of least absolute deviation estimators of stationary AR(p) models, Acta Math. Sinica 42 (1999) 259–267.