Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 84 (2003) 247–261
http://www.elsevier.com/locate/jmva

Strong convergence of estimators in nonlinear autoregressive models

Eckhard Liebscher

Technical University Ilmenau, Institute of Mathematics, Ilmenau/Thür D-98684, Germany

Received 10 November 1999

Abstract

In the paper we prove rates of strong convergence of M-estimators for the parameters in a general nonlinear autoregressive model. In the proofs we utilize a variational principle from stochastic optimization theory which was proved by Shapiro (Ann. Oper. Res. 30 (1991) 169). The application of the general theory is illustrated in the case of continuous threshold models.

AMS 1991 subject classifications: primary 62M10; secondary 62F12

Keywords: Nonlinear autoregressive model; M-estimators; Strong convergence; Threshold models

1. Introduction

The goal of this paper is to prove rates of strong convergence of M-estimators in stationary autoregressive models with an autoregression function which is not necessarily smooth but Lipschitz-continuous. Here we treat least squares estimators as well as robust estimators (M-estimators). We endeavour after giving statements with mild assumptions which are easy to verify and allow an immediate application.

Section 2 contains the main results of the paper which are proved in Section 4. Assuming the differentiability of the autoregression function with respect to the parameters (except for a set of measure zero which may depend on the parameters), we prove that the rate of strong convergence of the estimator coincides with that of the law of the iterated logarithm. Moreover, we obtain the convergence rate $O((\ln n/n)^{1/2})$ of the estimator if the autoregression function is assumed to satisfy

E-mail address: lieb@mathematik.tu-ilmenau.de.

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(02)00022-2

only a Lipschitz-condition. The author is not aware of literature in which these properties are proved in connection with autoregressive models. In the proofs of our results, we employ a variational principle from stochastic optimization theory which was proved by Shapiro [31]. We claim that our approach may be applied to other problems of estimation theory.

The classical theory of least squares estimators and maximum-likelihood estimators of nonlinear smooth autoregressive models is presented in [10,15]. In the situation of linear autoregressive models, Koul and Zhu [17] obtained results which are similar to ours. These authors also established Bahadur–Kiefer representation for M-estimators. Asymptotic normality of M-estimators in autoregressive models was studied by Bustos [3] in the linear case, and by Koul [16] in the nonlinear case. Lai [18] derived a central limit theorem for least squares estimators in regression models (including autoregressive models) where the residuals form a martingale difference. Asymptotic normality and consistency of least squares estimators in ARMA-models was proved in [11], see also [8]. Tjøstheim [33] considered a general type estimator (including the M-estimator) in nonlinear time series models and showed asymptotic normality of it. In their monograph Pötscher and Prucha [27] proved consistency and asymptotic normality of estimators for parameters of time series models in a general framework.

Section 3 illustrates how to apply the main results of Section 2 to continuous threshold models having relevance in applications. We extend the results of Petruccelli [24] and show that least squares estimators have the strong convergence rate of the law of the iterated logarithm.

It should be mentioned that a lot of authors examined nonparametric estimators in autoregressive models, cf. [9,20,23]. Moreover, there is an extensive literature about the performance of least squares estimators in nonlinear regression models. We refer to earlier papers by Jennrich [14], Malinvaud [22] and Wu [37] as well as to the more recent paper by Richardson and Bhattacharyya [29] and the monograph by Prakasa Rao [28]. Concerning M-estimators, some references are the classical monograph by Huber [12] and the paper by Yohai and Maronna [38]. More recent accounts are due to Liese and Vajda [21], Arcones [2] and Van de Geer [36].

2. Main results

In this paper we consider the nonlinear autoregressive model. More precisely, let $\{X_t\}_{t=1,2,\dots}$ be a strictly stationary sequence of random variables fulfilling

$$X_{t+1} = g(X_t, \dots, X_{t-p+1} \mid \theta_0) + \varepsilon_{t+1} \quad (t = p, p+1, \dots), \tag{2.1}$$

where $\{\varepsilon_t\}_{t=p+1,p+2,\dots}$ is a sequence of independent random variables which are independent of $X_1, \dots, X_p$ (cf. for example [35, Chapter 3]). $\theta_0 \in \Theta \subset \mathbb{R}^q$ is the vector of the true parameters of the model. Let Θ be a bounded and closed set where θ_0 is an inner point of Θ. Assume that $g : \mathbb{R}^p \times \Theta \to \mathbb{R}$ is a measurable function such that

$\theta \mapsto g(x \mid \theta)$ is continuous for π-almost all x, where π is the distribution of $\tilde{X}_t := (X_t, \ldots, X_{t-p+1})^\top$.

In this section the asymptotic performance of M-estimators $\hat{\theta}_n$ of the parameters in model (2.1) is studied. Let us define the estimator $\hat{\theta}_n$ as a global minimizer:

$$\hat{\theta}_n \in \arg\min_{\theta \in \Theta} \sum_{t=p}^{n-1} \rho(X_{t+1} - g(\tilde{X}_t \mid \theta)). \tag{2.2}$$

The continuity assumption on g ensures the existence of $\hat{\theta}_n$ with probability 1. Least squares estimators $\check{\theta}_n$ represent a special case of $\hat{\theta}_n$:

$$\check{\theta}_n \in \arg\min_{\theta \in \Theta} \sum_{t=p}^{n-1} (X_{t+1} - g(\tilde{X}_t \mid \theta))^2. \tag{2.3}$$

We assume that ρ is a real function which satisfies a nonuniform Lipschitz-condition

$$|\rho(x) - \rho(y)| \leqslant L_\rho(|x|^\tau + |y|^\tau + 1)|x - y| \quad \forall x, y \in \mathbb{R} \tag{2.4}$$

with some $\tau \geqslant 0$. This Lipschitz condition is fulfilled in the case $\rho(x) = x^2$ $(\tau = 1)$ as well as in the case $\rho(x) = |x|^p$, $1 \leqslant p < 2$ $(\tau = p - 1)$. Let

$$\Phi(\theta) = \mathbb{E}\rho(X_{t+1} - g(\tilde{X}_t \mid \theta)).$$

For proving convergence rates, we need a theorem about consistency.

Theorem 2.1. *Suppose that* $\{X_t\}$ *is ergodic*,

$$\mathbb{E} \sup_{\theta \in \Theta} |g(\tilde{X}_t \mid \theta)|^{\tau+1} < +\infty, \quad \mathbb{E}|\varepsilon_t|^{\tau+1} < +\infty \tag{2.5}$$

and

$$\Phi(\theta) > \Phi(\theta_0) \quad \text{for all } \theta \in \Theta, \theta \neq \theta_0. \tag{2.6}$$

Then $\lim_{n \to \infty} \hat{\theta}_n = \theta_0$ *a.s.*

This theorem follows immediately from the uniform strong law of large numbers given in Lemma A2(b) of Pötscher and Prucha [26], and from Lemma 3.1 (including remarks on p. 16) of the monograph by Pötscher and Prucha [27]. Obviously,

$$\mathbb{E}(\rho(\varepsilon_t + a)) > \mathbb{E}(\rho(\varepsilon_{t+1})) \quad \text{a.s. for all } a \neq 0 \tag{2.7}$$

and

$$\mathbb{P}\{g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \theta) \neq 0\} > 0 \quad \text{for all } \theta \in \Theta, \theta \neq \theta_0 \tag{2.8}$$

imply condition (2.6).

Note that $\{\tilde{X}_t\}$ is a homogeneous Markov chain, the so-called associated Markov chain. We suppose that the density of $\tilde{X}_t$ exists and use the following assumptions:

Condition $\mathscr{L}$. Inequality (2.4) and

$$|g(x\,|\,\theta_1)-g(x\,|\,\theta_2)|\leqslant L_g(||x||^{\zeta}+1)||\theta_1-\theta_2|| \quad \forall x\in\mathbb{R}^p,\ \ \theta_1,\theta_2\in\Theta,$$

$\mathbb{E}|\varepsilon_t|^{\gamma}<+\infty$, $\mathbb{E}||\tilde{X}_t||^{\gamma}<+\infty$ with $\zeta\in[0,1]$, $\gamma>2(\tau+1)$ and $L_g>0$.

Condition $\mathscr{L}'$. (i) Condition $\mathscr{L}$ is satisfied and
(ii) ρ' exists on $\mathbb{R}$,

$$|\rho'(x_1)-\rho'(x_2)|\leqslant L_{\rho'}(|x_1|^{\xi}+|x_2|^{\xi}+1)|x_1-x_2| \quad \forall x_1,x_2\in\mathbb{R}.$$

(iii) $g(x\,|\,\theta)$ is differentiable w.r.t. θ for all $\theta\in\Theta$ and all $x\in\mathbb{R}^p\backslash A$, $A:=\{x\in\mathbb{R}^p\colon\ \exists j: z_j^{\top}x=h_j(\theta_0)\}$ where $z_1,\dots,z_J\in\mathbb{R}^p\backslash\{0\}$ are some vectors and $h_1,\dots,h_J:\Theta\to\mathbb{R}$ some functions such that $|h_j(\theta_1)-h_j(\theta_2)|\leqslant L_h||\theta_1-\theta_2||$ $\forall\theta_1,\theta_2\in\Theta$, $j=1,\dots,J$.
(iv) There is a real number $s\in(0,1]$ such that

$$\left|\frac{\partial}{\partial\theta_i}g(x\,|\,\theta_1)-\frac{\partial}{\partial\theta_i}g(x\,|\,\theta_2)\right|\leqslant L_{g'}(||x||^{\zeta}+1)||\theta_1-\theta_2||^{s} \tag{2.9}$$

(ζ as in condition $\mathscr{L}$) for all $x\in\mathbb{R}^p\backslash A$ and all $\theta_1,\theta_2\in\Theta$ with $z_j^{\top}x\neq h_j(\theta_1+\psi(\theta_2-\theta_1))$ $\forall j,\psi\in[0,1]$.
(v) Moreover, $\mathbb{E}|\varepsilon_t|^{\gamma}<+\infty$, $\mathbb{E}||\tilde{X}_t||^{\gamma}<+\infty$ with $\gamma>2(\xi+2)$, $\xi\in[0,1]$. $L_h,L_{g'},L_{\rho'}>0$ are constants.

Condition $\mathscr{A}$. There is a neighbourhood V of θ_0 such that

$$\Phi(\theta)\geqslant\Phi(\theta_0)+\alpha||\theta-\theta_0||^2 \quad \forall\theta\in V, \tag{2.10}$$

where $\alpha>0$ is a constant.

Condition (2.9) is fulfilled if for each $x\in\mathbb{R}^p$, there exist second-order partial derivatives of $g(x\,|\,.)$ on the set $\{\theta\colon\ z_j^{\top}x\neq h_j(\theta)\ \forall j\}$ and the absolute values of these derivatives are bounded by $\mathrm{const}\cdot(||x||^{\zeta}+1)$. Condition $\mathscr{A}$ is very similar to the second-order growth condition (Assumption A) of Shapiro [31]. If Φ is differentiable and strongly convex in a convex neighbourhood of θ_0, and if Θ is convex, then (2.10) is satisfied [32, p. 102]. Furthermore, we have the following lemma which contains sufficient conditions for $\mathscr{A}$.

Lemma 2.1. *For almost all $x\in\mathbb{R}^p$ (w.r.t. Lebesgue-measure), let the gradient vector $\nabla g(x\,|\,\theta)$ of g w.r.t. θ exist for θ in a neighbourhood of θ_0 and be continuous at θ_0. Suppose that the second-order partial derivatives of Φ exist in a neighbourhood of θ_0 and are continuous at θ_0. Assume that* (2.8), *condition $\mathscr{L}$ and*

$$\mathbb{P}\{v^{\top}\nabla g(\tilde{X}_t\,|\,\theta_0)\neq 0\}>0 \quad \textit{for any } v\in\mathbb{R}^q, v\neq 0 \tag{2.11}$$

are fulfilled, and one of the following assumptions is satisfied:
(i) *either* $\mathbb{E}\varepsilon_t=0$ *and* $\rho(x)=x^2$ *or*

(ii) med $\varepsilon_t = 0$, $\rho(x) = |x|$ *and* ε_t *has a bounded density* h *on* $\mathbb{R}$ *with* $h(0) > 0$ *or*
(iii) (2.7) *is fulfilled,* ρ *has a second derivative on* $\mathbb{R}$ *and* $\mathbb{E}\rho''(\varepsilon_t) > 0$.
Then condition $\mathscr{A}$ *and* (2.6) *hold true.*

Next the main results are provided. For the definition of geometric ergodicity, we refer to Tjøstheim [34].

Theorem 2.2. *Assume that conditions* $\mathscr{A}$, $\mathscr{L}$ *and* (2.6) *are satisfied. Suppose that* $\{\tilde{X}_t\}$ *is geometrically ergodic. Then*

$$||\hat{\theta}_n - \theta_0|| = O\left(\sqrt{\frac{\ln n}{n}}\right) \quad a.s.$$

Theorem 2.3. *Suppose that the assumptions of Theorem* 2.2 *and condition* $\mathscr{L}'$ *are satisfied. Let the density of* $\tilde{X}_t$ *be bounded. Then*

$$||\hat{\theta}_n - \theta_0|| \leqslant C_0 \sqrt{\frac{\ln \ln(n)}{n}} \quad a.s. \text{ for } n \geqslant n_0(\omega)$$

with a constant $C_0 > 0$.

In these two theorems (2.6) and condition $\mathscr{A}$ represent the global and the local assumptions about minimizing properties of Φ, respectively. Theorem 2.3 states that the estimator $\hat{\theta}_n$ of (2.2) tends to the true parameter vector θ_0 at the rate corresponding to the law of the iterated logarithm. Here we do not assume the existence of second-order derivatives of g.

Sufficient conditions for geometric ergodicity of $\{\tilde{X}_t\}$ can be found in papers by Tjøstheim [34], by Ango Nze [1] and by Masry and Tjøstheim [23]. Remember that stationary geometrically ergodic Markov chains are absolutely regular with β-mixing coefficients which decay to zero exponentially fast. This fact is utilized in the proofs in order to obtain an inequality of Bernstein type.

3. Threshold models

Let $\{\varepsilon_k\}$ be a sequence of i.i.d. random variables with $\mathbb{E}\varepsilon_k = 0$. In this section the continuous SETAR(p, l, d)-model

$$X_t = \begin{cases} a_0 + \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t & \text{if } X_{t-d} \in R_1, \\ a_0 + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{k=2}^{j} b_k (X_{t-d} - r_{k-1}) + \varepsilon_t & \text{if } X_{t-d} \in R_j,\ j = 2, \ldots, l \end{cases} \tag{3.1}$$

$(t = p, p+1, \ldots)$ is considered where $r_1 < r_2 < \cdots < r_{l-1}$ are the thresholds, and $R_1, \ldots, R_l$ are the regions of the different process regimes. These regions are defined by $r_0 = -\infty, R_i = (r_{i-1}, r_i]$ for $i < l$, $R_l = (r_{l-1}, \infty)$. The parameter vector of the model is given by $\theta_0 = (a_0, \ldots, a_p, b_2, \ldots, b_l, r_1, \ldots, r_{l-1})^\top \in \Theta \subset \mathbb{R}^q$,

$\theta = (\bar{a}_0, \dots, \bar{r}_{l-1})^\top \in \Theta,\ q = p + 2l - 1$. Therefore

$$g(y \mid \theta) = \bar{a}_0 + \sum_{i=1}^{p} \bar{a}_i y_i + \sum_{k=2}^{l} \bar{b}_k (y_d - \bar{r}_{k-1}) \mathbf{1}(y_d > \bar{r}_{k-1}) \quad (\theta \in \Theta),$$

where $\mathbf{1}(y > a) = 1$ if $y > a$, $\mathbf{1}(y > a) = 0$ otherwise. Model (3.1) is a special case of SETAR-models described in [35]. In contrast to Chan [4], the delay parameter d is fixed and not a component of θ.

The aim of this section is to give the convergence rate of least squares estimators $\check{\theta}_n$ defined in (2.3). Asymptotic normality of maximum likelihood estimators in a special threshold model is shown in [13]. Chan [4] proved that the least squares estimator for the threshold in a discontinuous threshold model has a faster convergence rate than the usual one. The paper by Pham et al. [25] deals with strong convergence of least squares estimators in threshold models including the case of a nonergodic time series. A further approach to overcome the difficulties arising from nondifferentiability of g at some points in threshold models is described in the paper by Chan and Tong [6] where smooth threshold models (STAR-models) are considered.

We assume that $\{X_t\}$ is stationary and geometrically ergodic. In the case $p = 1$, the paper by Chan et al. [5] contains sufficient conditions for geometric ergodicity. The following condition is used in the result of this section.

Condition $\mathscr{T}$. Suppose that ε_t has the density h and the density f of X_t is continuous and has a support including the interval $[r_{\min} - \eta, r_{\max} + \eta]$, $\eta > 0$ where $r_{\min} = \min\{\bar{r}_1\colon\ \theta \in \Theta\}$, $r_{\max} = \max\{\bar{r}_{l-1}\colon\ \theta \in \Theta\}$. There is some $\varepsilon > 0$ such that $\bar{r}_{k-1} \leqslant \bar{r}_k - \varepsilon$ for all $\theta \in \Theta$ and $k = 2, \dots, l$.

Note that under Condition $\mathscr{T}$, Condition $\mathscr{L}'$ is satisfied for least squares estimators where

$$h_j(\theta) = \theta_{p+l+j}, \quad z_j = (0, \dots, 1_d, 0, \dots, 0)^\top \in \mathbb{R}^p \quad (j = 1, \dots, l-1),$$

$$A = \{y \in \mathbb{R}^p\colon\ \exists j\colon\ y_d = r_j\},$$

$$\begin{aligned}\nabla g(y \mid \theta) = (&1, y_1, \dots, y_p, (y_d - \bar{r}_1)\mathbf{1}(y_d > \bar{r}_1), \dots, (y_d - \bar{r}_{l-1})\mathbf{1}(y_d > \bar{r}_{l-1}),\\ &- \bar{b}_2 \mathbf{1}(y_d > \bar{r}_1), \dots, -\bar{b}_l \mathbf{1}(y_d > \bar{r}_{l-1}))^\top \quad \text{for } y \notin A.\end{aligned}$$

Hence (2.8) and (2.11) are fulfilled. Now we are in a position to formulate the result of this section. The following statement is a direct consequence of Lemma 2.1 and Theorem 2.3.

Corollary 3.1. *Assume that Condition $\mathscr{T}$ is satisfied and $\mathbb{E}|\varepsilon_t|^\gamma < \infty$, $\mathbb{E}||\tilde{X}_t||^\gamma < \infty$ with $\gamma > 4$. Then*

$$||\check{\theta}_n - \theta_0|| \leqslant C_1 \sqrt{\frac{\ln \ln(n)}{n}} \quad \textit{a.s. for } n \geqslant \bar{n}_1(\omega)$$

with a constant $C_1 > 0$.

In the case of continuous threshold models, this statement extends Theorem 1 of Chan [4]. Considering the smooth version of SETAR models (STAR-models), the law of the iterated logarithm and a statement on asymptotic normality is given in [6]. The convergence rate for M-estimators can be derived in a similar way.

4. Proofs

Throughout this section we assume that the time series $\{X_t\}$ follows model (2.1), is stationary and geometrically ergodic. Suppose that the density of $\tilde{X}_t$ exists. Moreover, let $\varepsilon_{p+1}, \varepsilon_{p+2}, \ldots$ be i.i.d. random variables not depending on $\tilde{X}_p$. We denote the Hessian matrix of Φ at θ by $H(\theta)$.

Proof of Lemma 2.1. (a) *Assertion*: $H(\theta_0)$ is positive definite. Obviously, $\nabla\Phi(\theta_0) = 0$.

Case (i): Note that

$$\nabla\Phi(\theta) = -2\mathbb{E}(g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \theta))\nabla g(\tilde{X}_t \mid \theta).$$

By means of the Lipschitz-condition on g and the dominated convergence theorem, we obtain

$$\begin{aligned}\frac{\partial^2}{\partial\theta_j \partial\theta_k}\Phi(\theta_0) &= -2\lim_{\eta\to 0}\eta^{-1}\mathbb{E}(g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \tilde{\theta}_k))\frac{\partial}{\partial\theta_j}g(\tilde{X}_t \mid \tilde{\theta}_k)\\ &= 2\mathbb{E}\frac{\partial}{\partial\theta_k}g(\tilde{X}_t \mid \theta_0)\frac{\partial}{\partial\theta_j}g(\tilde{X}_t \mid \theta_0) \quad (j,k = 1,\ldots,q)\end{aligned}$$

with $\tilde{\theta}_k = (\theta_{01}, \ldots, \theta_{0k} + \eta, \ldots, \theta_{0q})^\top, \theta_0 = (\theta_{01}, \ldots, \theta_{0q})^\top$. Consequently, by (2.11),

$$H(\theta_0) = 2\mathbb{E}\nabla g(\tilde{X}_t \mid \theta_0)\nabla g(\tilde{X}_t \mid \theta_0)^\top$$

is positive definite.

Case (ii): Here we obtain

$$\begin{aligned}\nabla\Phi(\theta) &= -\mathbb{E}(\mathrm{sgn}(\varepsilon_{t+1} + g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \theta))\nabla g(\tilde{X}_t \mid \theta))\\ &= \mathbb{E}((-1 + 2F_\varepsilon(g(\tilde{X}_t \mid \theta) - g(\tilde{X}_t \mid \theta_0)))\nabla g(\tilde{X}_t \mid \theta)),\end{aligned}$$

where F_ε is the distribution function of ε_t. Since $\mathrm{med}(\varepsilon_t) = 0,\ F_\varepsilon(0) = 0.5$, we have

$$\begin{aligned}\frac{\partial^2}{\partial\theta_j \partial\theta_k}\Phi(\theta_0) &= \lim_{\eta\to 0}\eta^{-1}\mathbb{E}\left((-1 + 2F_\varepsilon(g(\tilde{X}_t \mid \tilde{\theta}_k) - g(\tilde{X}_t \mid \theta_0)))\frac{\partial}{\partial\theta_j}g(\tilde{X}_t \mid \tilde{\theta}_k)\right)\\ &= 2h(0)\mathbb{E}\frac{\partial}{\partial\theta_k}g(\tilde{X}_t \mid \theta_0)\frac{\partial}{\partial\theta_j}g(\tilde{X}_t \mid \theta_0) \quad (j,k = 1,\ldots,q)\end{aligned}$$

with $\tilde{\theta}_k$ as above. This implies assertion (a).

Case (iii): We deduce

$$\nabla\Phi(\theta) = -\mathbb{E}\rho'(\varepsilon_{t+1} + g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \theta))\nabla g(\tilde{X}_t \mid \theta).$$

Since $\mathbb{E}\rho'(\varepsilon_t) = 0$, we have

$$\begin{aligned}\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Phi(\theta_0) &= -\lim_{\eta\to 0}\eta^{-1}\mathbb{E}\left(\rho'(\varepsilon_{t+1} + g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \tilde{\theta}_k))\frac{\partial}{\partial\theta_j}g(\tilde{X}_t \mid \tilde{\theta}_k)\right)\\ &= -\lim_{\eta\to 0}\eta^{-1}\mathbb{E}(\rho'(\varepsilon_{t+1} + g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \tilde{\theta}_k)) - \rho'(\varepsilon_{t+1}))\\ &\quad\times \frac{\partial}{\partial\theta_j}g(\tilde{X}_t \mid \tilde{\theta}_k)\\ &= \mathbb{E}\rho''(\varepsilon_{t+1})\mathbb{E}\frac{\partial}{\partial\theta_k}g(\tilde{X}_t \mid \theta_0)\frac{\partial}{\partial\theta_j}g(\tilde{X}_t \mid \theta_0) \quad (j,k = 1,\dots,q)\end{aligned}$$

with $\tilde{\theta}_k$ as above. This proves assertion (a).

(b) *Proof of condition $\mathscr{A}$*: An application of Taylor's formula leads to

$$\Phi(\theta) = \Phi(\theta_0) + (\theta-\theta_0)^T H(\theta^*)(\theta-\theta_0),$$
$$\theta^* = \theta_0 + \psi(\theta-\theta_0), \quad 0<\psi<1.$$

We choose the neighborhood $V\subset\Theta$ of θ_0 such that $H(\theta)$ is positive definite for $\theta\in V$, and $\alpha>0$ is a lower bound for the smallest eigenvalues of $H(\theta)$ for $\theta\in V$. This is possible since the elements of H are continuous at θ_0. Consequently,

$$(\theta-\theta_0)^T H(\theta^*)(\theta-\theta_0) \geqslant \alpha\|\theta-\theta_0\|^2.$$

This completes the proof. □

We suppose that conditions $\mathscr{A}$, $\mathscr{L}$ and (2.6) are satisfied. Define

$$\Phi_n(\theta) \coloneqq \frac{1}{n-p}\sum_{t=p}^{n-1}\rho(X_{t+1} - g(\tilde{X}_t \mid \theta)) \quad (\theta\in\Theta)$$

such that

$$\Phi(\theta) = \mathbb{E}\Phi_n(\theta) = \mathbb{E}\rho(X_{t+1} - g(\tilde{X}_t \mid \theta)) \quad (\theta\in\Theta).$$

F_n denotes the empirical distribution function of the sample $(X_{p+1},\tilde{X}_p)$, $(X_{p+2},\tilde{X}_{p+1}),\dots,(X_n,\tilde{X}_{n-1})$. Let F be the distribution function of $(X_{i+1},\tilde{X}_i)$, and

$$\delta(\theta) = \Phi_n(\theta) - \Phi(\theta) = \int_{\mathbb{R}^{p+1}}\rho(x - g(y \mid \theta))\,\mathrm{d}(F_n(x,y) - F(x,y)).$$

Now we provide a variational principle which was proved by Shapiro.

Theorem 4.1 (Shapiro [31, Lemma 4.1]). *Assume that $\hat{\theta}_n\in V$ and condition $\mathscr{A}$ is satisfied. Then*

$$\|\hat{\theta}_n - \theta_0\| \leqslant \frac{1}{\alpha}\sup\left\{\frac{|\delta(\theta)-\delta(\theta_0)|}{\|\theta-\theta_0\|}\colon\ \theta\in\Theta\cap V, \theta\neq\theta_0\right\}.$$

This theorem is the crucial statement for the following proofs. Our next task is to prove the following lemma:

Lemma 4.1.

$$\sup_{\theta\in\tilde U}\frac{|\delta(\theta)-\delta(\theta_0)|}{||\theta-\theta_0||}=O(n^{-1/2}\sqrt{\ln n})\quad a.s.,$$

where $\tilde U := \{\theta\in\Theta\colon\ ||\theta-\theta_0||\geqslant n^{-1}\}$.

Let $\{\alpha_k\}$ be the α-mixing coefficients of the sequence $\{\check X_t, t=p+1,p+2,\dots\}$ where $\check X_t := (X_t, X_{t-1},\dots,X_{t-p})^\top$. Note that geometric ergodicity of $\{\tilde X_k\}$ implies $\alpha_k = O(\rho^k),\ \rho\in(0,1)$ (see [7, pp. 88,89]). For the proof of Lemma 4.1, we need an inequality of Bernstein-type and some further lemmas.

Proposition 4.2. *Let* $\{Z_i\}_{i=1,2,\dots}$ *be a stationary* α*-mixing sequence of real r.v. with mixing coefficients* $\{\alpha_j^Z\}$. *Assume that* $\mathbb{E}Z_1=0$, $\mathbb{E}|Z_1|^{\tilde\gamma}<+\infty$ *and* $\sum_{i=1}^\infty (\alpha_i^Z)^{1-2/\tilde\gamma}<\infty$ *for some* $\tilde\gamma>2$. *Then, for* $n,N\in\mathbb{N}, 0<N\leqslant n/2$, *for* $S,\varepsilon>0$,

$$\mathbb{P}\left\{\left|\sum_{i=1}^n Z_i\right| I\left(\max_{i=1,\dots,n}|Z_i|\leqslant S\right)>\varepsilon\right\}$$
$$\leqslant 4\exp\left\{-\frac{\varepsilon^2}{16}\left(2nd_N+\frac{1}{3}\varepsilon SN\right)^{-1}+\sum_{i=1}^n\mathbb{P}\{|Z_i|>S\}\right\}+32\frac{S}{\varepsilon}n\alpha_N^Z$$

and $d_N=(\mathbb{E}|Z_1|^{\tilde\gamma})^{2/\tilde\gamma}(1+20\sum_{i=1}^\infty(\alpha_i^Z)^{1-2/\tilde\gamma})$.

Proof. The proposition follows immediately from Proposition 5.1 of Liebscher [20] and Lemma 2.2 of Liebscher [19]. □

Lemma 4.2. *For each* n *and* $\theta\in\bar V\subset\Theta$, *let* $W_{np}(\theta), W_{n,p+1}(\theta),\dots$ *be a stationary* α*-mixing sequence of random variables with mixing coefficients which are bounded by the coefficients* $\{\alpha_k\}$ *of* $\{\check X_t\}$. *Moreover, let* $\mathbb{E}W_{nt}(\theta)=0$ *for all* $\theta\in\bar V$, $t=p,p+1,\dots$ *Assume that there is a stationary sequence of random variables* $\tilde W_p, \tilde W_{p+1},\dots$ *(not depending on* n*) with* $\mathbb{E}|\tilde W_t|^{\bar\gamma}<+\infty$ *for some* $\bar\gamma>2$,

$$\sup_{\theta\in\bar V}|W_{nt}(\theta)|\leqslant\tilde W_t \tag{4.1}$$

and

$$\sup_{\theta\in\bar V}(\mathbb{E}|W_{nt}(\theta)|^{\bar\gamma})^{1/\bar\gamma}=O(V_n). \tag{4.2}$$

Then

$$\max_{k=1,\dots,v}|\bar W_n(u_k)|=O(n^{-1/2}V_n\sqrt{\ln(n)})\ a.s.$$

with $\bar{W}_n(x) = (n-p)^{-1}\sum_{t=p}^{n-1} W_{nt}(x)$, $u_1,\dots,u_v \in \bar{V}$ *provided that* $v = v(n) \leqslant \text{const} \cdot n^{\bar{q}}$ *with some* $\bar{q} \geqslant 3$ *and* $V_n \geqslant (\ln(n))^{-1}$.

Proof. By (4.1), a standard argument leads to

$$\max_{t=p,\dots,n-1} \sup_{\theta\in\bar{V}} |W_{nt}(\theta)| \leqslant n^{1/\bar{\gamma}}(\ln n)^{1/\bar{\gamma}+\kappa} \quad \text{for all } n \geqslant n_0(\omega) \tag{4.3}$$

with some $\kappa > 0$. Let $a_n \coloneqq n^{-1/2} V_n \sqrt{\ln n}$, $A_{nk} = \{\omega\colon \max_{t=p,\dots,n-1} |W_{nt}(u_k)| \leqslant n^{1/\bar{\gamma}}(\ln n)^{1/\bar{\gamma}+\kappa}\}$ and $I_n = I(\bigcap_{k=1}^{v} A_{nk})$. Note that

$$\max_{k=1,\dots,v} \sum_{t=p}^{n-1} \mathbb{P}\{|W_{nt}(u_k)| > n^{1/\bar{\gamma}}(\ln n)^{1/\bar{\gamma}+\kappa}\} \leqslant n^{-1}(\ln n)^{-1-\bar{\gamma}\kappa} \sum_{t=p}^{n-1} \mathbb{E}|\tilde{W}_t|^{\bar{\gamma}} = o(1).$$

An application of Proposition 4.2 and (4.2) leads to

$$\begin{aligned}
&\mathbb{P}\left\{\max_{k=1,\dots,v} |\bar{W}_n(u_k)| I_n > \varepsilon a_n\right\} \\
&\quad\leqslant \sum_{k=1}^{v} \mathbb{P}\{|\bar{W}_n(u_k)| I(A_{nk}) > \varepsilon a_n\} \\
&\quad\leqslant C_2 v\left(\exp\{-C_3 \varepsilon^2 a_n^2 (n^{-1} V_n^2 + \varepsilon a_n N n^{-1+1/\bar{\gamma}}(\ln n)^{1/\bar{\gamma}+\kappa})^{-1}\} + \frac{n}{N}\alpha_N\right) \\
&\quad\leqslant C_4 (n^{\bar{q}} \exp\{-C_5 \varepsilon^2 a_n^2 (n^{-1} V_n^2 + \varepsilon a_n n^{-1+1/\bar{\gamma}}(\ln n)^{1+1/\bar{\gamma}+\kappa})^{-1}\} + n^{1-\bar{q}})
\end{aligned}$$

for any $\varepsilon > 0$ where $N \coloneqq \lceil 2\bar{q}|\ln \rho|^{-1} \ln n \rceil$. C_2 to C_5 are positive constants not depending on n or ε. Consequently, the series

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\max_{k=1,\dots,v} |\bar{W}_n(u_k)| \cdot I_n > \varepsilon a_n\right\}$$

converges for large $\varepsilon > 0$. An application of the Borel–Cantelli lemma and (4.3) leads to Lemma 4.2. □

Since $\tilde{U}$ is a compact set, then, for any n, $\tilde{U}$ can be covered with q-dimensional closed cubes $\tilde{U}_1,\dots,\tilde{U}_v$ having the properties:

$$\begin{aligned}
&\|\theta_1 - \theta_2\| \leqslant n^{-3}, \qquad \|\theta_1 - \theta_0\| \geqslant \frac{1}{2n} \quad \forall \theta_1, \theta_2 \in U_i, \ i = 1,\dots,v, \\
&\quad v \leqslant \text{const}\, n^{3q}, \quad \tilde{U}_i \cap \tilde{U} \neq \emptyset \quad (i = 1,\dots,v).
\end{aligned}$$

We denote $\tilde{U}_i \cap \tilde{U}$ by U_i $(i = 1,\dots,v)$. Let u_i be any point of U_i, $i = 1,\dots,v$. We obtain

$$\sup_{\theta\in\tilde{U}} \frac{|\delta(\theta) - \delta(\theta_0)|}{\|\theta - \theta_0\|} = \sup_{\theta\in\tilde{U}} |\bar{Z}_n(\theta)|,$$

where

$$\bar{Z}_n(\theta) = (n-p)^{-1} \sum_{t=p}^{n-1} (Z_t(\theta) - \mathbb{E}Z_t(\theta)),$$

$$Z_t(\theta) = ||\theta - \theta_0||^{-1}(\rho(X_{t+1} - g(\tilde{X}_t \mid \theta)) - \rho(X_{t+1} - g(\tilde{X}_t \mid \theta_0))).$$

Thus

$$\sup_{\theta \in U_k} |\bar{Z}_n(\theta)| \leqslant \sup_{\theta \in U_k} \left|(n-p)^{-1} \sum_{t=p}^{n-1} (Z_t(\theta) - Z_t(u_k))\right| + \sup_{\theta \in U_k} |\mathbb{E}Z_t(\theta) - \mathbb{E}Z_t(u_k)| + |\bar{Z}_n(u_k)| \quad (k = 1, \ldots, v). \tag{4.4}$$

Lemma 4.3.

$$\max_{k=1,\ldots,v} |\bar{Z}_n(u_k)| = O(n^{-1/2}\sqrt{\ln n}) \quad a.s.$$

Proof. Obviously,

$$\begin{aligned}
|Z_t(\theta)| &\leqslant L_\rho(2|\varepsilon_{t+1}|^\tau + |g(\tilde{X}_t \mid \theta_0) - g(\tilde{X}_t \mid \theta)|^\tau + 1) \\
&\quad \times ||\theta - \theta_0||^{-1}|g(\tilde{X}_t \mid \theta) - g(\tilde{X}_t \mid \theta_0)| \\
&\leqslant L_\rho(2|\varepsilon_{t+1}|^\tau + L_g^\tau(||\tilde{X}_t||^\zeta + 1)^\tau \sup_{\theta \in U} ||\theta - \theta_0||^\tau + 1) \cdot L_g(||\tilde{X}_t||^\zeta + 1) \\
&\leqslant C_6(|\varepsilon_{t+1}|^{\tau+1} + ||\tilde{X}_t||^{\zeta(\tau+1)} + 1) \quad (\theta \in \tilde{U})
\end{aligned}$$

with an appropriate constant $C_6 > 0$. Let $\mu > 2$ such that $\mu(\tau + 1) < \gamma$ and $W_{nt}(\theta) = Z_t(\theta)$. Then (4.1) and (4.2) with $\bar{\gamma} = \mu$ are satisfied. Now apply Lemma 4.2 to get Lemma 4.3. □

Lemma 4.4.

$$\max_{k=1,\ldots,v} \sup_{\theta \in U_k} \left| n^{-1} \sum_{t=p}^{n-1} (Z_t(\theta) - Z_t(u_k)) \right| = O(n^{-1}) \quad a.s.$$

and

$$\max_{k=1,\ldots,v} \sup_{\theta \in U_k} |\mathbb{E}Z_t(\theta) - \mathbb{E}Z_t(u_k)| = O(n^{-1}).$$

Proof. By the strong law of large numbers, we obtain the first part of the lemma as follows:

$$\begin{aligned}
&\max_{k=1,\ldots,v} \sup_{\theta \in U_k} \left| n^{-1} \sum_{t=p}^{n-1} (Z_t(\theta) - Z_t(u_k)) \right| \\
&\quad \leqslant n^{-1} \max_{k=1,\ldots,v} \sup_{\theta \in U_k} \sum_{t=p}^{n-1} ||\theta - \theta_0||^{-1} |\rho(X_{t+1} - g(\tilde{X}_t \mid \theta)) - \rho(X_{t+1} - g(\tilde{X}_t \mid u_k))|
\end{aligned}$$

$$
\begin{aligned}
&+n^{-1} \max_{k=1,\dots,v} \sup_{\theta\in U_k} \sum_{t=p}^{n-1} (|\rho(X_{t+1}-g(\tilde{X}_t\,|\,u_k))-\rho(X_{t+1}-g(\tilde{X}_t\,|\,\theta_0))| \\
&\times |\,||\theta-\theta_0||^{-1}-||u_k-\theta_0||^{-1}|) \\
\leqslant\, &\text{const}\times \sup_{\theta\in U_k} ||u_k-\theta|| \sum_{t=p}^{n-1} (|\varepsilon_{t+1}|^{\tau+1}+||\tilde{X}_t||^{\zeta(\tau+1)}+1)=O(n^{-1}) \quad \text{a.s.}
\end{aligned}
$$

Analogously, one proves the second part of the lemma. □

Proof of Lemma 4.1. Lemma 4.1 is a consequence of Lemmas 4.3, 4.4 and (4.4). □

Proof of Theorem 2.2. In view of Theorem 2.1, the assumptions of Theorem 2.2 imply $\hat{\theta}_n \to \theta_0$ a.s. as $n\to\infty$ such that $\hat{\theta}_n\in V$ for $n\geqslant n_1(\omega)$. Without loss of generality, let $||\hat{\theta}_n-\theta_0||\geqslant n^{-1}$. Now Theorem 2.2 is a consequence of Lemma 4.1 and Theorem 4.1. □

Now we turn to prove Theorem 2.3. We assume that, in addition, condition $\mathscr{L}'$ is satisfied.

Lemma 4.5. *We have*

$$
\sup_{\theta\in\bar{U}_n} \frac{|\delta(\theta)-\delta(\theta_0)|}{||\theta-\theta_0||} \leqslant C_7\sqrt{\frac{\ln\ln n}{n}} \quad \text{a.s. for } n\geqslant n_2(\omega)
$$

with $\bar{U}_n=\tilde{U}\cap\{\theta\colon ||\theta-\theta_0||\leqslant n^{-1/2}\ln n\}$ *and a constant* $C_7>0$.

Proof. Define $A_n=\{y\colon \min_{j=1,\dots,J}||z_j^\top y-h_j(\theta_0)||\geqslant(1+L_h)n^{-1/2}\ln(n)\}$. Let $y\in A_n$. Hence $g(y\,|\,.)$ is differentiable in $\bar{U}_n$ and

$$
\left|\frac{\partial}{\partial\theta_j}g(y\,|\,\theta)-\frac{\partial}{\partial\theta_j}g(y\,|\,\theta_0)\right| \leqslant L_{g'}(||y||^{\zeta}+1)||\theta-\theta_0||^{s}
$$

for $\theta\in\bar{U}_n$ since $||z_j^\top y-h_j(\theta)||\geqslant||z_j^\top y-h_j(\theta_0)||-L_h||\theta-\theta_0||\geqslant n^{-1/2}\ln(n)$ for $\theta\in\bar{U}_n$. Observe that

$$
\begin{aligned}
&\sup_{\theta\in\bar{U}_n} \frac{|\delta(\theta)-\delta(\theta_0)|}{||\theta-\theta_0||} \\
&\leqslant \sup_{\theta\in\bar{U}_n} \left|\left|\int_{\mathbb{R}\times A_n}\int_0^1 \rho'(x-g(y\,|\,\theta_t))\nabla g(y\,|\,\theta_t)\,\mathrm{d}t\,\mathrm{d}(F_n(x,y)-F(x,y))\right|\right| + B_n \\
&\leqslant \left|\left|\int_{\mathbb{R}^{p+1}} \rho'(x-g(y\,|\,\theta_0))\nabla g(y\,|\,\theta_0)\,\mathrm{d}(F_n(x,y)-F(x,y))\right|\right| \\
&\quad + \breve{B}_n O(n^{-s/2}\ln(n)) + B_n + \bar{B}_n \quad \text{a.s.,} \qquad (4.5)
\end{aligned}
$$

where $\theta_t = \theta_0 + t(\theta - \theta_0)$ for $t \in [0,1]$,

$$B_n := \sup_{\theta \in \bar{U}_n} \bigg| \int_{\mathbb{R}\times A_n^c} (\rho(x - g(y \mid \theta)) - \rho(x - g(y \mid \theta_0))) \times ||\theta - \theta_0||^{-1} \, \mathrm{d}(F_n(x,y) - F(x,y)) \bigg|,$$

$$\bar{B}_n := \bigg|\bigg| \int_{\mathbb{R}\times A_n^c} \rho'(x - g(y \mid \theta_0)) \nabla g(y \mid \theta_0) \, \mathrm{d}(F_n(x,y) - F(x,y)) \bigg|\bigg|,$$

$$\breve{B}_n := \sup_{\theta \in \bar{U}_n} \bigg|\bigg| \int_{\mathbb{R}\times A_n} (\rho'(x - g(y \mid \theta)) \nabla g(y \mid \theta) - \rho'(x - g(y \mid \theta_0)) \nabla g(y \mid \theta_0)) \times ||\theta - \theta_0||^{-s} \, \mathrm{d}(F_n(x,y) - F(x,y)) ||, \quad A_n^c = \mathbb{R}^p \backslash A_n.$$

For y with $z_j^\top y = h_j(\theta_0)$ for some j, we put $\nabla g(y \mid \theta_0) = 0$. Rio's [30] law of the iterated logarithm (Theorem 2 and comments at p. 1191) yields

$$\left| \int \rho'(x - g(y \mid \theta_0)) \frac{\partial}{\partial \theta_j} g(y \mid \theta_0) \, \mathrm{d}(F_n(x,y) - F(x,y)) \right| \leqslant C_8 \sqrt{\frac{\ln \ln n}{n}} \quad \text{a.s.} \quad (4.6)$$

for $j = 1, \ldots, q$, $n \geqslant n_2(\omega)$ with a constant $C_8 > 0$. Obviously, $\mathbb{P}\{\tilde{X}_t \in A_n^c\} = O(n^{-\bar{\kappa}})$ with some $\bar{\kappa} > 0$. Let γ_1, γ_2 such that $2 < \gamma_1 < \gamma_2$, $\gamma_2(\tau + 1) < \gamma$. We deduce

$$\begin{aligned}
&\sup_{\theta \in \bar{U}_n} \mathbb{E}|I(\tilde{X}_t \in A_n^c)(\rho(X_{t+1} - g(\tilde{X}_t \mid \theta)) - \rho(X_{t+1} - g(\tilde{X}_t \mid \theta_0)))||\theta - \theta_0||^{-1}|^{\gamma_1} \\
&\quad \leqslant \mathbb{E}|I(\tilde{X}_t \in A_n^c)(C_1(|\varepsilon_{t+1}|^{\tau+1} + ||\tilde{X}_t||^{\zeta(\tau+1)} + 1))|^{\gamma_1} \\
&\quad \leqslant C_1^{\gamma_1}(\mathbb{E}(|\varepsilon_{t+1}|^{\tau+1} + ||\tilde{X}_t||^{\zeta(\tau+1)} + 1)^{\gamma_2})^{\gamma_1/\gamma_2}(\mathbb{P}\{\tilde{X}_t \in A_n^c\})^{1-\gamma_1/\gamma_2} \\
&\quad = O(n^{-\bar{\kappa}(1-\gamma_1/\gamma_2)}).
\end{aligned}$$

Using Lemma 4.2, we obtain

$$B_n = o(n^{-1/2}) \quad \text{and} \quad \bar{B}_n = o(n^{-1/2}). \quad (4.7)$$

Applying Lemma 4.2, one proves that

$$\sup_{\theta \in \bar{U}_n} \left| \int_{\mathbb{R}\times A_n} \Delta_j(x,y,\theta) ||\theta - \theta_0||^{-s} \, \mathrm{d}(F_n(x,y) - F(x,y)) \right| = o(n^{-1/2}) \quad \text{a.s.} \quad (4.8)$$

for $j = 1, \ldots, q$ where

$$\Delta_j(x,y,\theta) = \rho'(x - g(y \mid \theta)) \frac{\partial}{\partial \theta_j} g(y \mid \theta) - \rho'(x - g(y \mid \theta)) \frac{\partial}{\partial \theta_j} g(y \mid \theta).$$

Eqs. (4.5)–(4.8) imply the lemma. □

Proof of Theorem 2.3. Theorem 2.2 states that $||\hat{\theta}_n - \theta_0|| = O(n^{-1/2}\sqrt{\ln(n)})$ a.s. Hence $||\hat{\theta}_n - \theta_0|| \leqslant n^{-1/2} \ln(n)$ and $\hat{\theta}_n \in V$ for $n \geqslant n_3(\omega)$. Lemma 4.5 and Theorem 4.1 imply Theorem 2.3. □

Acknowledgments

The author is grateful to the referee for valuable suggestions.

References

[1] P. Ango Nze, Criteria of ergodicity for some models with a markovian representation (in French), C. R. Acad. Sci. Paris Sér. I t. 315 (1992) 1301–1304.
[2] M.A. Arcones, Some strong limit theorems for M-estimators, Stochastic Process. Appl. 53 (1994) 241–268.
[3] O.H. Bustos, General M-estimates for contaminated *p*th-order autoregressive processes: consistency and asymptotic normality. Robustness in autoregressive processes, Z. Wahrsch. Verw. Gebiete 59 (1982) 491–504.
[4] K.S. Chan, Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, Ann. Statist. 21 (1993) 520–533.
[5] K.S. Chan, J.D. Petruccelli, H. Tong, S.W. Woolford, A multiple threshold AR(1) model, J. Appl. Probab. 22 (1985) 267–279.
[6] K.S. Chan, H. Tong, On estimating thresholds in autoregressive models, J. Time Ser. Anal. 7 (1986) 179–190.
[7] P. Doukhan, Mixing: Properties and Examples, Lecture Notes in Statistics, Vol. 85, Springer, Berlin, 1994.
[8] C. Francq, J.-M. Zakoïan, Estimating linear representations of nonlinear processes, J. Statist. Plann. Inference 68 (1998) 145–165.
[9] L. Györfi, W. Härdle, P. Sarda, P. Vieu, Non-parametric Curve Estimation from Time Series, Lecture Notes in Statistics, Vol. 60, Springer, Berlin, 1989.
[10] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press, New York, 1980.
[11] E.J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, Wiley, New York, 1988.
[12] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[13] S.Y. Hwang, I.V. Basawa, Large sample inference based on multiple observations from nonlinear autoregressive processes, Stochastic Process. Appl. 49 (1994) 127–140.
[14] R.I. Jennrich, Asymptotic properties of non-linear least squares estimators, Ann. Math. Statist. 40 (1969) 633–643.
[15] L.A. Klimko, P.L. Nelson, On conditional least-squares estimation for stochastic processes, Ann. Statist. 6 (1978) 629–642.
[16] H.L. Koul, Asymptotics of some estimators and sequential residual empiricals in non-linear time series, Ann. Statist. 24 (1996) 380–404.
[17] H.L. Koul, Z. Zhu, Bahadur–Kiefer representations for GM-estimators in autoregression models, Stochastic Process. Appl. 57 (1995) 167–189.
[18] T.L. Lai, Asymptotic properties of nonlinear least squares estimates in stochastic regression models, Ann. Statist. 22 (1994) 1917–1930.
[19] E. Liebscher, Strong convergence of sums of α-mixing random variables with applications to density estimation, Stochastic Process. Appl. 65 (1996) 69–80.
[20] E. Liebscher, Estimation of the density and the regression function under mixing conditions, Statist. Decisions 19 (2001) 9–26.
[21] F. Liese, I. Vajda, Consistency of M-estimates in general regression models, J. Multivariate Anal. 50 (1994) 93–114.
[22] E. Malinvaud, The consistency of nonlinear regression, Ann. Math. Statist. 41 (1970) 956–969.
[23] E. Masry, D. Tjøstheim, Nonparametric estimation and identification of nonlinear ARCH time series, Econometric Theory 11 (1995) 258–289.
[24] J.D. Petruccelli, On the consistency of least squares estimator of a threshold AR(1) model, J. Time Ser. Anal. 7 (1986) 269–278.

[25] D.T. Pham, K.S. Chan, H. Tong, Strong consistency of the least squares estimator for a non-ergodic threshold autoregressive model, Statistica Sinica 1 (1991) 361–369.
[26] B.M. Pötscher, I.R. Prucha, A class of partially adaptive one-step M-estimators for a nonlinear regression model with dependent observations, J. Econometrics 32 (1986) 219–225.
[27] B.M. Pötscher, I.R. Prucha, Dynamic Nonlinear Econometric Models, Asymptotic Theory, Springer, Berlin, 1997.
[28] B.L.S. Prakasa Rao, Asymptotic Theory of Statistical Inference, Wiley, New York, 1987.
[29] G.D. Richardson, B.B. Bhattacharyya, Consistent estimators in nonlinear regression for a noncompact parameter space, Ann. Statist. 14 (1986) 1591–1596.
[30] E. Rio, The functional law of the iterated logarithm for stationary strongly mixing sequences, Ann. Statist. 23 (1995) 1188–1203.
[31] A. Shapiro, Asymptotic analysis of stochastic programs, Ann. Oper. Res. 30 (1991) 169–186.
[32] A. Shapiro, Quantitative stability in stochastic programming, Math. Programming 67 (1994) 99–108.
[33] D. Tjøstheim, Estimation in nonlinear time series models, Stochastic Process. Appl. 21 (1986) 251–273.
[34] D. Tjøstheim, Non-linear time series and Markov chains, Adv. Appl. Probab. A 22 (1990) 587–611.
[35] H. Tong, Non-linear Time Series: A Dynamical Approach, Oxford University Press, Oxford, 1990.
[36] S.A. Van de Geer, Applications of Empirical Process Theory, Cambridge Series on Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2000.
[37] C.-F. Wu, Asymptotic theory of nonlinear least squares estimation, Ann. Statist. 9 (1981) 501–513.
[38] V.J. Yohai, R.A. Maronna, Asymptotic behavior of M-estimators for the linear model, Ann. Statist. 7 (1979) 258–268.