SCIENCE CHINA
Information Sciences

• LETTER •

June 2023, Vol. 66 169204:1–169204:2
https://doi.org/10.1007/s11432-021-3309-7

Asymptotic behavior of least squares estimators for nonlinear autoregressive models

Zhaobo LIU[1] & Chanying LI[2,3*]

[1]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China;
[2]Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;
[3]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

Received 27 January 2021/Revised 11 May 2021/Accepted 30 June 2021/Published online 17 November 2022

Citation Liu Z B, Li C Y. Asymptotic behavior of least squares estimators for nonlinear autoregressive models. Sci China Inf Sci, 2023, 66(6): 169204, https://doi.org/10.1007/s11432-021-3309-7

Dear editor,

Estimations of nonlinear autoregressive (AR) models in the literature typically involve ergodic series. Based on this assumption, the asymptotic theory has been established accordingly (see [1–3]). However, this good property is not always true [4]. For example, we consider

$$y_{t+1} = \theta^\tau \phi(y_t, \dots, y_{t-n+1}) + w_{t+1}, \quad t \geqslant 0, \tag{1}$$

where θ is the $m \times 1$ unknown parameter vector and y_t and w_t are the scalar observations and random noise signals, respectively. Moreover, $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is a known Lebesgue measurable vector function:

$$\phi(z_1, \dots, z_n) = \mathrm{col}\{\phi^{(1)}(z_1), \dots, \phi^{(n)}(z_n)\}, \tag{2}$$

where $\phi^{(i)} = (f_{i1}, \dots, f_{im_i})^\tau : \mathbb{R} \to \mathbb{R}^{m_i}, i = 1, \dots, n$ are certain known Lebesgue measurable vector functions and $m_i \geqslant 1$ are n integers satisfying $\sum_{i=1}^n m_i = m$. Without loss of generality, let $y_t = 0$ for $t < 0$. Clearly, most functions ϕ produce non-ergodic sequences $\{y_t\}$. Therefore, in this article, parameter θ in model (1), whose outputs are not necessarily ergodic, is identified.

Least squares (LS) estimators are known as the most efficient algorithms in parameter estimation. Its strong consistency for model (1) depends crucially on matrix $P_{t+1}^{-1} = I_m + \sum_{i=0}^t \phi_t \phi_t^\tau$, where $\phi_t = \phi(y_t, \dots, y_{t-n+1})$. Let $\lambda_{\min}(t+1)$ and $\lambda_{\max}(t+1)$ denote the minimal and maximal eigenvalues of P_{t+1}^{-1}, respectively. In the Gaussian case, Ref. [5] showed

$$\left\{\lim_{t\to+\infty} \lambda_{\min}(t+1) = +\infty\right\} = \left\{\lim_{t\to+\infty} \hat{\theta}_t = \theta\right\}, \tag{3}$$

whereas Ref. [6, Theorem 1] found that when $\{w_t\}$ is an appropriate martingale difference sequence,

$$\|\hat{\theta}_{t+1} - \theta\|^2 = O\left(\frac{\log(\lambda_{\max}(t+1))}{\lambda_{\min}(t+1)}\right), \quad \text{a.s.} \tag{4}$$

Moreover, Ref. [6] pointed out that $\log(\lambda_{\max}(t+1)) = o(\lambda_{\min}(t+1))$ is in some sense the weakest condition for the strong consistency of $\hat{\theta}_t$ in general.

The eigenvalues of P_{t+1}^{-1} depend on outputs $\{y_t\}$, which are produced automatically by the nonlinear random system (1). Thus, checking $\lim_{t\to+\infty} \lambda_{\min}(t+1) = +\infty$ or $\log(\lambda_{\max}(t+1)) = o(\lambda_{\min}(t+1))$ is nontrivial. However, Ref. [7] successfully verified $\liminf_{t\to+\infty} t^{-1}\lambda_{\min}(t+1) > 0$, a.s. for the linear AR model,

$$y_{t+1} = \theta_1 y_t + \theta_2 y_{t-1} + \cdots + \theta_n y_{t-n+1} + w_{t+1}, \ t \geqslant 0, \tag{5}$$

which is a special case of (1). Then, Ref. [7] completely solved the strong consistency of the LS estimator for this basic situation. The proof in [7] attributes to the linear structure of the AR model (5) to a certain extent. Regarding the nonlinear model (1), we aim to identify whether the LS estimator retains a similar asymptotic behavior.

This study establishes the asymptotic properties of the LS estimator for model (1). By assuming mild conditions on ϕ, the minimal eigenvalue of P_{t+1}^{-1} is estimated. We find that the LS estimates converge to the true parameter almost surely on the set where vectors $(y_t, \dots, y_{t-n+1})^\tau$ do not diverge to infinity. This means the LS estimator is highly likely to have a strong consistency when applied to model (1) in practice because most real systems are non-divergent. Finally, Appendix A provides the proofs of the main theorems.

Gaussian case. We assume the following conditions.

(A1) Noise $\{w_t\}$ is an i.i.d. random sequence with $w_1 \sim N(0, 1)$, and parameter $\theta \sim N(\theta_0, I_m)$ is independent of $\{w_t\}$.

(A2) Certain open sets $\{E_i\}_{i=1}^n$ belong to $\mathbb{R}$ satisfying the following:

(i) $f_{ij} \in C(\mathbb{R})$ and $f_{ij} \in C^{m_i}(E_i)$, $1 \leqslant j \leqslant m_i$, $1 \leqslant i \leqslant n$;

(ii) For every unit vector $x \in \mathbb{R}^m$, a point $y \in \prod_{i=1}^n E_i$ exists such that $|\phi^\tau(y)x| \neq 0$.

Remark 1. By assumption (A2)(ii), for every unit vector $x \in \mathbb{R}^m$, $\ell(\{y \in \prod_{i=1}^n E_i : |\phi^\tau(y)x| > 0\}) > 0$, where ℓ denotes the Lebesgue measure.

* Corresponding author (email: cyli@amss.ac.cn)

The LS estimate $\hat{\theta}_t$ for parameter θ can be recursively defined by

$$\begin{cases} \hat{\theta}_{t+1} = \hat{\theta}_t + P_{t+1}\phi_t(y_{t+1} - \phi_t^{\tau}\hat{\theta}_t), \\ P_{t+1} = P_t - (1 + \phi_t^{\tau}P_t\phi_t)^{-1}P_t\phi_t\phi_t^{\tau}P_t, \quad P_0 = I_m, \end{cases} \tag{6}$$

where $\hat{\theta}_0$ is the deterministic initial condition and ϕ_0 is the random initial vector of model (1). We provide a simple way to estimate the minimal eigenvalue of P_{t+1}^{-1}. Let $N_t(M) \triangleq \sum_{i=1}^{t} I_{\{\|Y_i\|\leqslant M\}}$, where $Y_t \triangleq (y_{t+n-1}, \dots, y_t)^{\tau}$ and $M > 0$ is a constant. Then, in terms of $N_t(M)$, our estimate of $\lambda_{\min}(t+1)$ is readily available by the following theorem.

Theorem 1. Under assumptions (A1) and (A2), for any constant $M > 0$,

$$\liminf_{t\to+\infty} \frac{\lambda_{\min}(t+1)}{N_t(M)} > 0 \quad \text{a.s. on } \Omega(M), \tag{7}$$

where $\Omega(M) \triangleq \{\lim_{t\to+\infty} N_t(M) = +\infty\}$.

Corollary 1. Let assumptions (A1) and (A2) hold. Then,

$$\lim_{t\to+\infty} \hat{\theta}_t = \theta \quad \text{a.s. on } \left\{\liminf_{t\to+\infty} \|Y_t\| < +\infty\right\}. \tag{8}$$

Remark 2. For a typical case where $\prod_{i=1}^{n} E_i = \mathbb{R}^n$, if assumption (A2)(ii) fails, then $\ell(\{y \in \mathbb{R}^n : |\phi^{\tau}(y)x| > 0\}) = 0$ for some unit vector $x \in \mathbb{R}^m$. Therefore, as $t \to +\infty$, $\lambda_{\min}(t+1) = O(1)$, a.s. In view of (3), $\hat{\theta}_t$ cannot converge to the true parameter θ. So, assumption (A2)(ii) is necessary for the strong consistency of the LS estimates $\{\hat{\theta}_t\}_{t\geqslant 0}$.

Constant parameter. Assume θ is a non-random parameter, and

(A1$'$) $\{w_t\}$ is an i.i.d random sequence with $Ew_1 = 0$ and $E|w_1|^{\beta} < +\infty$ for some $\beta > 2$. Moreover, w_1 has a density $\rho(x)$ such that for every proper interval $I \subset \mathbb{R}$, $\inf_{x\in I} \rho(x) > 0$ and $\sup_{x\in\mathbb{R}} \rho(x) < +\infty$.

In this case, the LS estimator is constructed from partial data. More specifically, for some constant $C_{\phi} > 0$, ϕ_t is modified as $\phi_t \triangleq I_{\{\|Y_{t-n+1}\|\leqslant C_{\phi}\}}\phi(y_t, \dots, y_{t-n+1})$. Then, an analogous version of Theorem 1 is deduced as follows.

Theorem 2. Under assumptions (A1$'$) and (A2), a constant $M_{\phi} > 0$ exists depending only on ϕ such that for any $C_{\phi} > M_{\phi}$ and $M > 0$,

$$\liminf_{t\to+\infty} \frac{\lambda_{\min}(t+1)}{N_t(M)} > 0 \quad \text{a.s. on } \Omega(M).$$

Furthermore, if $M \geqslant C_{\phi}$, then

$$\|\hat{\theta}_t - \theta\|^2 = O\left(\frac{\log N_t(M)}{N_t(M)}\right) \quad \text{a.s. on } \Omega(M).$$

Remark 3. Furthermore, we conclude (8) in Theorem 2. For most practical situations, the systems frequently fulfill

$$P\left\{\liminf_{t\to+\infty} \|Y_t\| < +\infty\right\} = 1, \tag{9}$$

and the strong consistency of the LS estimates is thus guaranteed. Observe that assumption (A1$'$) and Eq. (9) imply that $\{y_t\}_{t\geqslant 1}$ in model (1) is an aperiodic Harris recurrent Markov chain and hence admits an invariant measure. Certain integrability assumptions on the invariant measure may lead to the consistency of the LS estimates (e.g., [8]). However, judging these integrability assumptions for a nonlinear AR model is generally nontrivial.

When $n = 1$ in model (1), assumption (A2) in Theorems 1 and 2 can be relaxed as below.

(A2$'$) $f_{1i} \in C^{m_1}(E_1), i = 1, \dots, m_1$ are linearly independent on an open set $E_1 \in \mathbb{R}$ and ϕ is bounded in every compact set.

Example 1. Consider a parametric AR model:

$$y_{t+1} = \sum_{j=1}^{m} \theta_j g(y_t) I_{\{y_t\in D_j\}} + y_t I_{\{y_t\in D_{m+1}\}} + w_{t+1}, \tag{10}$$

where $y_0 = 0$, $g(\cdot)$ is bounded in any compact set, $\{D_j\}_{j=1}^m$ are some compact subsets of $\mathbb{R}$ with positive Lebesgue measure, and $D_{m+1} = (\bigcup_{j=1}^m D_i)^c$. Let noises $\{w_t\}_{t\geqslant 1}$ satisfy assumption (A1$'$) and unknown parameters $\theta_1, \dots, \theta_m \in \mathbb{R}$. As can be seen, $\{y_t\}_{t\geqslant 1}$ must fall into $\bigcup_{j=1}^m D_i$ infinitely many times and Eq. (9) holds. Hence Theorems 1 and 2 can be applied and the strong consistency of the LS estimates is established. If $g(x) = x$, model (10) becomes a non-ergodic threshold autoregressive (TAR) model.

Conclusion. This study discusses the LS estimator for a basic class of nonlinear AR models, whose outputs are not necessarily ergodic. Several asymptotic properties of the LS estimator have been established under mild conditions. These properties suggest the strong consistency of the LS estimates in non-divergent nonlinear AR models.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 11925109) and National Key R&D Program of China (Grant No. 2018YFA0703800).

Supporting information Appendix A. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Chan K S. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Ann Statist, 1993, 21: 520–533
2 Li D, Ling S. On the least squares estimation of multiple-regime threshold autoregressive models. J Econom, 2012, 167: 240–253
3 Zhao W-X, Chen H-F, Zheng W X. Recursive identification for nonlinear ARX systems based on stochastic approximation algorithm. IEEE Trans Automat Contr, 2010, 55: 1287–1299
4 Li C, Lam J. Stabilization of discrete-time nonlinear uncertain systems by feedback based on LS algorithm. SIAM J Control Optim, 2013, 51: 1128–1151
5 Sternby J. On consistency for the method of least squares using martingale theory. IEEE Trans Automat Contr, 1977, 22: 346–352
6 Lai T L, Wei C Z. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Ann Statist, 1982, 10: 154–166
7 Lai T L, Wei C Z. Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. J Multivariate Anal, 1983, 13: 1–23
8 Li D, Tjøstheim D, Gao J. Estimation in nonlinear regression with Harris recurrent Markov chains. Ann Statist, 2016, 44: 1957–1987