Asymptotic Behavior of Least Squares Estimator for Nonlinear Autoregressive Models

Zhaobo Liu[1] and Chanying Li[2]

Articles you may be interested in

The rate of convergence for the least squares estimator in nonlinear regression model with dependent errors
Science in China Series A-Mathematics **45**, 137 (2002);

Prediction of earth rotation parameters based on improved weighted least squares and autoregressive model
Geodesy and Geodynamics **3**, 57 (2012);

Asymptotic properties of autoregressive regime-switching models
ESAIM: Probability and Statistics **16**, 25 (2012);

CONSISTENCY OF LEAST SQUARES ESTIMATES IN LINEAR MODELS
Scientia Sinica **22**, 162 (1979);

Abnormal behavior of the least squares estimate of multiple regression
Science in China Series A-Mathematics **40**, 234 (1997);

Asymptotic Behavior of Least Squares Estimator for Nonlinear Autoregressive Models

Zhaobo Liu[1] & Chanying Li[2,3*]

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

Dear editor,

Estimations of nonlinear autoregressive (AR) models in the literature typically involve ergodic series. Based on this assumption, a series of asymptotic theory has been established accordingly (see [1], [5], [8]). However, this good property is not always true. For example, we consider

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

where θ is the $m \times 1$ unknown parameter vector, y_t, 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

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

where $\phi^{(i)} = (f_{i1}, \ldots, f_{im_i})^\tau : \mathbb{R} \to \mathbb{R}^{m_i}, i = 1, \ldots, n$ 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) estimator is known as the most efficient algorithm 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 $\phi_t = \phi(y_t, \ldots, y_{t-n+1})$. 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, [7] showed

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

whereas [2, Theorem 1] found that when $\{w_t\}$ is an appropriate martingale difference sequence,

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

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

Moreover, [2] 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, [3] successfully verified $\liminf_{t \to +\infty} t^{-1}\lambda_{\min}(t+1) > 0$, a.s. for the linear AR model

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

which is a special case of (1). Then, [3] completely solved the strong consistency of the LS estimator for this basic situation. The proof in [3] 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 $(y_t, \ldots, y_{t-n+1})^\tau$ 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. Assume

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
(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 $x \in \mathbb{R}^m$, $\ell\left(\{y \in \prod_{i=1}^n E_i : |\phi^\tau(y)x| > 0\}\right) > 0$, where ℓ denotes the Lebesgue measure.

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

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

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 $Y_t \triangleq (y_{t+n-1}, \ldots, y_t)^\tau$ and $M > 0$ is a constant. Then, in terms of $N_t(M)$, our estimate of $\lambda_{\min}(t+1)$ is readily available by

Theorem 1.   Under Assumptions A1–A2, for any constant $M > 0$,

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

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

Corollary 1.   Let Assumptions A1–A2 hold. Then,

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

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 $\phi_t \triangleq I_{\{\|Y_{t-n+1}\| \leqslant C_\phi\}}\phi(y_t, \ldots, y_{t-n+1})$. 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

$$P\left\{\liminf_{t \to +\infty} \|Y_t\| < +\infty\right\} = 1 \qquad (9)$$

and the strong consistency of the LS estimates is thus guaranteed. Observe that Assumption A1' and (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. [6]). However, judging these integrability assumptions for a nonlinear AR model is in general nontrivial.

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

A2' $f_{1i} \in C^{m_1}(E_1), i = 1, \ldots, m_1$ 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

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

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 $\theta_1, \ldots, \theta_m \in \mathbb{R}$. As can be seen, $\{y_t\}_{t \geqslant 1}$ must fall into $\bigcup_{j=1}^m D_i$ infinitely many times and (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 paper 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.

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   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.

3   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.

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   Li, D., Ling, S. On the least squares estimation of multiple-regime threshold autoregressive models. J. Econometrics, 2012, 167, 240–253.

6   Li, D., Tjøstheim, D., Gao, J.. Estimation in nonlinear regression with Harris recurrent Markov chains. Ann. Statist., 2016, 44, 1957–1987.

7   Sternby, J.. On consistency for the method of least squares using martingale theory. IEEE Trans. Autom. Control, 1977, 22, 346–352.

8   Zhao, W. X., Chen, H. F., Zheng, W. X.. Recursive identification for nonlinear ARX systems based on stochastic approximation algorithm. IEEE Trans. Autom. Control, 2010, 55, 1287–1299.