Statistica Sinica 1(1991), 361-369

STRONG CONSISTENCY OF THE LEAST SQUARES ESTIMATOR FOR A NON-ERGODIC THRESHOLD AUTOREGRESSIVE MODEL

Dinh Tuan Pham, K. S. Chan and Howell Tong

Laboratory of Modelling and Computation, Grenoble, France,
University of Chicago and University of Kent

Abstract: We have shown that the least squares estimator for a non-ergodic, first order, self-exciting, threshold autoregressive model is strongly consistent under quite general conditions.

Key words and phrases: Least squares estimator, martingale, nonlinear unit root, stationarity, strong consistency, threshold autoregressive model.

1. Introduction

The class of self-exciting, threshold, autoregressive models (SETAR) has proved to be quite useful in nonlinear time series modelling. This class was introduced by Tong (1978) and has been studied by various authors (see Tong (1990) for references). In particular, it has been shown (Chan (1988)) that the least squares estimators (LSE) of the parameters of the model (including the thresholds and delay parameters) are strongly consistent. This result, however, depends crucially on the fact that the model is stationary and ergodic.

In this paper, we shall relax the above stationarity and ergodicity condition in the case of a simple model. Consider the first order SETAR model with only one threshold:

$$X_t = \begin{cases} \alpha X_{t-1} + \gamma + e_t, & \text{if} \quad X_{t-1} \le r, \\ \beta X_{t-1} + \delta + e_t, & \text{if} \quad X_{t-1} > r, \end{cases} \tag{1.1}$$

where e_t is a sequence of independent identically distributed (i.i.d.) random variables with zero mean and variance σ^2. The above model is stationary and geometrically ergodic if and only if $\alpha < 1$, $\beta < 1$ and $\alpha\beta < 1$, and is transient (as a Markov chain) if (α, β) is in the exterior of this region, but the situation is rather complex when (α, β) lies on the boundary (see Chan et al. (1985) and Guo and Petruccelli (1990)). Here, we are interested in the case where ergodicity may not hold. To simplify matters we shall assume that the threshold r is known, so that by simple substraction we may take $r = 0$. However, the above 4-parameter model is still too difficult to analyze (in the non-ergodic case). Therefore we assume further that the autoregression function is continuous at 0 and its value at this point is known, i.e. δ and γ are equal and known). For this simple model we shall show that the LSE of the parameters are strongly consistent if and only if $\alpha \le 1$ (< 1 if $\gamma < 0$) and $\beta \le 1$ (< 1 if $\gamma > 0$). The boundary case of $\alpha\beta = 1$ is also considered. This may be called the nonlinear unit root problem by analogy with the linear case.

2. Strong Consistency of the LSE

The least squares estimators of the parameters α, β of the model based on a sample $X_1, \ldots, X_n$, say, are obtained by minimizing the sum of squares

$$Q_n(\alpha,\beta) = \sum_{t=2}^{n}[X_t - g(X_{t-1},\alpha,\beta)]^2 \tag{2.1}$$

where $g(x,\alpha,\beta) = [\alpha I(x \le 0) + \beta I(x > 0)]x + \gamma$ and $I(\cdot)$ denotes the set indicator function. Simple computation shows that these estimators are given by

$$\widehat{\alpha}_n = \left[\sum_{t=1}^{n-1} I(X_t \le 0)X_t(X_{t+1} - \gamma)\right] \Big/ \left[\sum_{t=1}^{n-1} I(X_t \le 0)X_t^2\right],$$

$$\widehat{\beta}_n = \left[\sum_{t=1}^{n-1} I(X_t > 0)X_t(X_{t+1} - \gamma)\right] \Big/ \left[\sum_{t=1}^{n-1} I(X_t > 0)X_t^2\right].$$

Let α_0, β_0 denote the true values of the parameters α, β. Then from (1.1) with $\alpha = \alpha_0$, $\beta = \beta_0$, $\gamma = \delta$ and $r = 0$, we get

$$\widehat{\alpha}_n - \alpha_0 = M_n^- / S_n^-, \quad \widehat{\beta}_n - \beta_0 = M_n^+ / S_n^+, \tag{2.2}$$

where

$$M_n^- = \sum_{t=1}^{n-1} I(X_t \le 0)X_t e_{t+1}, \quad S_n^- = \sum_{t=1}^{n-1} I(X_t \le 0)X_t^2,$$

$$M_n^+ = \sum_{t=1}^{n-1} I(X_t > 0)X_t e_{t+1}, \quad S_n^+ = \sum_{t=1}^{n-1} I(X_t > 0)X_t^2.$$

From (2.2), it is clear that the LSE is strongly consistent if and only if the following conditions hold:

(C0) $\lim_{n\to\infty} M_n^- / S_n^- = 0, \quad \lim_{n\to\infty} M_n^+ / S_n^+ = 0,$ almost surely.

Observe that for $n > 1$, M_n^- and M_n^+ are martingales adapted to the σ-field $\mathcal{A}_n$ generated by $X_1, \ldots; X_n$. That is, they are $\mathcal{A}_n$-measurable and $E(M_n^- \mid \mathcal{A}_{n-1}) = M_{n-1}^-$, $E(M_n^+ \mid \mathcal{A}_{n-1}) = M_{n-1}^+$. Now let $M_n, n > 1$, be any martingale adapted to some σ-field $\mathcal{A}_n$ and denote by D_n the sum of the conditional variances $\sum_{t=2}^n E[(M_t - M_{t-1})^2 \mid \mathcal{A}_{t-1}]$. Then we have the following result (see e.g. Neveu (1965), p. 150): On the set $\{\lim_{n\to\infty} D_n = \infty\}$,

$$M_n/[D_n^{1/2}(\log D_n)^{(1/2)+\varepsilon}] \to 0, \quad \text{as } n \to \infty,$$

almost surely for every $\varepsilon > 0$, while on the set $\{\lim_{n\to\infty} D_n < \infty\}$, M_n converges almost surely to a finite limit. It is easy to see that the sum of the conditional variances for the martingales M_n^- and M_n^+ are precisely $\sigma^2 S_n^-$ and $\sigma^2 S_n^+$, respectively. Therefore the conditions (C0) are equivalent to

(C1) $\lim_{n\to\infty} S_n^- = \infty, \quad \lim_{n\to\infty} S_n^+ = \infty, \quad$ almost surely.

We now proceed to study the behaviour of S_n^- and S_n^+. To this end we shall establish some results concerning the sample path behaviour of the process X_t. For ease of reading, the proofs are relegated to the Appendix. In the sequel we shall make the following assumptions:

(A1) $P(e_1 + \gamma > 0) > 0$,

(A2) $P(e_1 + \gamma < 0) > 0$.

Note that if we exclude the case $\sigma^2 = 0$, which corresponds to a deterministic process and is without interest, then from $E(e_1) = 0$, we have $P(e_1 > 0) > 0$ and $P(e_1 < 0) > 0$. Thus (A1) holds trivially when $\gamma \geq 0$ and so does (A2) when $\gamma \leq 0$. In particular, for $\gamma = 0$, both (A1) and (A2) hold. In any case, these assumptions hold if the distribution of e_1 has infinite positive and negative tails. The latter is a mild condition.

Lemma 2.1. *If $\alpha_0 < 1$ or $\alpha_0 = 1$, $\gamma \geq 0$, there exists a positive number c such that for all x,*

$$P(X_t > c \text{ for some } t > 0 \mid X_0 = x) = 1,$$

or equivalently, for any initial distribution on X_0, the Markov chain X_t enters the interval $[c, \infty)$ infinitely often almost surely.

Similarly, if $\beta_0 < 1$ or $\beta_0 = 1$, $\gamma \leq 0$, there exists a negative number d such that for all x,

$$P(X_t < d \text{ for some } t > 0 \mid X_0 = x) = 1,$$

or equivalently, for any initial distribution on X_0, X_t enters $(-\infty, d]$ infinitely often almost surely.

Note. Assumptions (A1) and (A2) cannot be relaxed, at least in the case $\alpha \geq 0$ and $\beta \geq 0$. Indeed, for $x < 0$, $P(X_t < 0 \mid X_{t-1} = x) \leq P(e_1 + \gamma \leq 0)$. Hence, if $P(e_1 + \gamma > 0) = 0$, we have $P(X_t < 0 \mid X_{t-1} = x) = 1$, which implies that starting at a point $x < 0$, the process remains indefinitely negative. Thus when $\alpha \geq 0$ and (A1) fails, the first conclusion of the Lemma would not hold. Similarly, $P(e_1 + \gamma < 0) = 0$ implies that starting at a point $x > 0$, the process remains indefinitely positive. Therefore when $\beta \geq 0$ and (A2) fails, the second conclusion of the Lemma would not hold.

Lemma 2.2. *If $\alpha_0 > 1$ or $\alpha_0 = 1$, $\gamma < 0$, then for any $x < 0$, $P(X_t \leq x$ for all $t > 0 \mid X_0 = x) > 0$. Similarly, if $\beta_0 > 1$ or $\beta_0 = 1$, $\gamma > 0$, then for any $x > 0$, $P(X_t \geq x$ for all $t > 0 \mid X_0 = x) > 0$.*

Lemma 2.1 shows that if α_0, β_0 and γ satisfy the conditions of this Lemma then (C1) holds since $S_n^+ \geq c^2 \sum_{t=2}^n I(X_{t-1} \geq c)$ and $S_n^- \geq d^2 \sum_{t=2}^n I(X_{t-1} \leq d)$. Lemmas 2.1 and 2.2 show that if these conditions are not satisfied then (C1) will not hold. Indeed, when α_0, γ satisfy the first condition of Lemma 2.2 and β_0, γ satisfy the second condition of Lemma 2.1, then by Lemma 2.1, X_t will almost surely eventually become negative and by Lemma 2.2, there is a positive probability that it remains negative indefinitely. Thus with positive probability, X_t enters the positive real line finitely often. The same is true when the positive real line is replaced by the negative one, when β_0, γ satisfy the second condition of Lemma 2.2 and when α_0, γ satisfy the first condition of Lemma 2.1. If α_0, β_0 and γ satisfy the conditions of Lemma 2.2, then depending on the starting value, there is a positive probability that X_t will be always positive or always negative. Therefore in all cases, at least one of the sequences of random variables S_n^- and S_n^+ is bounded with positive probability. This yields the following Proposition.

Proposition 1. *The* LSE *estimator $(\widehat{\alpha}, \widehat{\beta})$ is strongly consistent if and only if one of the following sets of conditions holds:* (i) $\alpha_0 \leq 1$, $\beta_0 \leq 1$ *and* $\gamma = 0$, (ii) $\alpha_0 < 1$, $\beta_0 \leq 1$ *and* $\gamma < 0$ *and* (iii) $\alpha_0 \leq 1$, $\beta_0 < 1$ *and* $\gamma > 0$.

We now consider the boundary case, that is when it is *known* that the point (α, β) lies on the boundary of the stationary region. This boundary is formed by three curves: $\{\alpha \leq 1,\ \beta = 1\}$, $\{\alpha = 1,\ \beta \leq 1\}$ and $\{\alpha = 1/\beta < 0\}$. For the case $\alpha \leq 1$, $\beta = 1$, the LSE of α is the same as in the general case. A similar remark holds for the case $\alpha = 1$, $\beta \leq 1$. However, in case $\alpha = 1/\beta < 0$, the least squares method would estimate α by minimizing $Q_n(\alpha, 1/\alpha)$, where Q_n is as in (2.1), leading to a different estimator. The following Proposition shows that the resulting estimator is also strongly consistent.

Proposition 2. *Suppose that $\alpha_0 = 1/\beta_0 < 0$; then the estimator of α obtained* *by minimizing $Q_n(\alpha, 1/\alpha)$ is strongly consistent.*

Proof. We have

$$\begin{aligned} &Q_n(\alpha,1/\alpha) - Q_n(\alpha_0, 1/\alpha_0) \\ &= \sum_{t=2}^{n}\{[e_t + g(X_{t-1},\alpha_0,1/\alpha_0) - g(X_{t-1},\alpha,1/\alpha)]^2 - e_t^2\} \\ &= (\alpha-\alpha_0)^2 S_n^- - 2(\alpha-\alpha_0)M_n^- + (1/\alpha - 1/\alpha_0)S_n^+ - 2(1/\alpha - 1/\alpha_0)M_n^+. \end{aligned}$$

Let δ be an arbitrary positive number. Observe that $x^2 - 2bx > \delta^2 - 2|b|\delta$ for $|x| \geq \delta \geq |b|$ and $|\alpha - \alpha_0| > \delta$ implies $|1/\alpha - 1/\alpha_0| > \delta/[|\alpha_0|(|\alpha_0| + \delta)]$. Therefore when $|M_n^-|/S_n^- \leq \delta$ and $|M_n^+|/S_n^+ \leq \delta/[|\alpha_0|(|\alpha_0| + \delta)]$, we have

$$\begin{aligned} &\inf_{\alpha:|\alpha-\alpha_0|>\delta}[Q_n(\alpha,1/\alpha) - Q_n(\alpha_0,1/\alpha_0)] \\ &\geq S_n^+\delta^2 - 2|M_n^-|\delta + S_n^+\delta^2/[|\alpha_0|(|\alpha_0|+\delta)]^2 - 2|M_n^+|\delta/[|\alpha_0|(|\alpha_0|+\delta)]. \quad (2.3) \end{aligned}$$

But we have already shown that (C1) holds and hence (C0) also holds, which implies that almost surely for n large enough, $|M_n^-|/S_n^- \leq \delta$, $|M_n^+|/S_n^+ \leq \delta/[|\alpha_0|(|\alpha_0| + \delta)]$ and hence (2.3) is satisfied. Further, from (C0) and (C1), the right hand side of (2.3) tends to infinity almost surely as n goes to infinity. Therefore

$$\lim_{n\to\infty} \inf_{\alpha:|\alpha-\alpha_0|>\delta}[Q_n(\alpha,1/\alpha) - Q_n(\alpha_0,1/\alpha_0)] = \infty, \quad \text{almost surely.}$$

Thus there exists, for almost all sample paths of the X_t process, an integer N such that for all $n > N$ and for all α satisfying $|\alpha - \alpha_0| > \delta$, we have $Q_n(\alpha,1/\alpha) - Q_n(\alpha_0,1/\alpha_0) > 0$. This implies that for all $n > N$, the function $Q_n(\alpha,1/\alpha)$ cannot attain its minimum outside the closed interval $[\alpha_0 - \delta, \alpha_0 + \delta]$; and since this function is continuous, its restriction on the compact set $[\alpha_0 - \delta, \alpha_0 + \delta]$ always admits a minimum. Thus for all $\delta > 0$, there exists, for almost all sample paths, an integer N such that for all $n > N$, the function $Q_n(\alpha,1/\alpha)$ admits a minimum which is in the interval $[\alpha_0 - \delta, \alpha_0 + \delta]$. This completes the proof of our Proposition.

Remark. A close look at the above proof reveals that the LSE of α, β under the constraints $\alpha\beta = c$, $\alpha < 0$, $\beta < 0$, where c is a given positive number, are also strongly consistent provided that the true values satisfy the same constraints. However, there is no reason to consider the constraint $\alpha\beta = c$ except for $c = 1$.

3. Some Open Problems

Our work is still incomplete in that some important questions have not been addressed. The first one concerns the more general model in which the threshold is unknown and/or the autoregression function takes unknown value at this threshold, with or without discontinuity there. Again, the least squares method may be used to estimate the model parameters. Strong consistency of the estimators has been established for the ergodic case provided that the invariant probability measure of the Markov chain X_t admits a strictly positive density (see Chan (1988)). In the general, not necessarily ergodic case, we conjecture that the last condition needs to be replaced by a recurrence property of any open interval for the Markov chain X_t. The second open problem concerns the convergence rate of the estimators. This rate may be faster than the usual rate $n^{-1/2}$ since, intuitively, the process would explode or drift to infinity with alternating signs, according to $\alpha_0\beta_0 > 1$ or $\alpha_0\beta_0 = 1$ ($\alpha_0 < 0$, $\beta_0 < 0$). The third open problem is the limiting distribution of the estimators.

Appendix : Proof of Lemmas 2.1 and 2.2

Proof of Lemma 2.1. We shall prove only the first part of the Lemma since the proof for the second part is similar.

We begin by establishing the equivalence between the two statements of the Lemma. Suppose that X_t enters $[c, \infty)$ infinitely often, regardless of the initial distribution of X_0. Then, taking this distribution to be the Dirac distribution with mass at x, we get

$$P(X_t > c \;\text{ for some }\; t > 0 \,|\, X_0 = x) = 1.$$

Conversely, if the above equality holds for all x, then from the homogeneity of the Markov chain X_t, we also have for all $k > 0$ and all x,

$$P(X_{k+t} > c \;\text{ for some }\; t > 0 \,|\, X_k = x) = 1.$$

By integration with respect to the distribution of X_k, we get $P(X_t > c$ for some $t > k) = 1$, and since this is true for all k, this means that the process enters $[c, \infty)$ infinitely often almost surely.

We now show the validity of the Lemma. Let c be a positive number. We have for all $x \in (0, c)$

$$\begin{aligned} P(X_1 \geq c \,|\, X_0 = x) &= P(\beta_0 x + e_1 + \gamma \geq c) = P(e_1 + \gamma \geq c - \beta_0 x) \\ &\geq P\{e_1 + \gamma \geq c(1 + |\beta_0|)\}. \end{aligned}$$

Choose c small enough such that $P\{e_1 + \gamma \geq (1 + |\beta_0|)c\} > 0$, which is possible because of (A1). Then

$$\inf_{x \in [0,c)} P(X_t \geq c \text{ for some } t > 0 \,|\, X_0 = x) > 0.$$

Hence, by Proposition 5.1 in Orey (1971), for any initial distribution on X_0,

$$\{X_t \in [0,c) \text{ infinitely often}\} \subseteq \{X_t \geq c \text{ infinitely often}\}$$

almost surely. But clearly,

$$\begin{aligned}&\{X_t \geq 0 \text{ infinitely often}\}\\ =&\{X_t \in [0,c) \text{ infinitely often}\} \cup \{X_t \geq c \text{ infinitely often}\}.\end{aligned}$$

Therefore we get the result of the Lemma: $P\{X_t \geq c \text{ infinitely often}\} = 1$, if we can prove that the event on the above left hand side has probability one, or equivalently (by the same argument as the beginning of this proof)

$$P(X_t < 0 \text{ for all } t > 0 \,|\, X_0 = x) = 0,$$

for all x. For this purpose, it is enough to show that for all $y < 0$,

$$P(X_t < 0 \text{ for all } t > 1 \,|\, X_1 = y) = 0,$$

which would yield the desired result by integration with respect to the conditional distribution of X_1 given $X_0 = x$.

We now prove the last equality. Note that the left hand side of this equality is the same as the probability that a first order autoregressive process with parameter α_0 and constant term γ, starting at $y < 0$, remains indefinitely negative. For $\alpha_0 = 1$, this process reduces to a random walk with increments having mean γ and hence the corresponding probability, for $\gamma \geq 0$, is zero (see Feller (1966), pp. 395, 396). The same is true for $0 < \alpha_0 < 1$, since then the above process is a stationary Markov chain with invariant probability measure having support not contained in $(-\infty, 0]$. To see this, note that this measure is the probability distribution of $\sum_{k=0}^{\infty} \alpha_0^k(e_{-k} + \gamma)$. By (A1) there exists $\epsilon > 0$ such that $P(e_{-k} + \gamma > \epsilon) > 0$. Thus $\sum_{k=0}^{n-1} \alpha_0^k(e_{-k} + \gamma)$ is greater than $\epsilon(1 - \alpha_0^n)/(1 - \alpha_0)$ with positive probability. Now, choose n large enough such that $(1 - \alpha_0^n)\epsilon + \alpha_0^n\gamma$ is positive (if $\gamma \geq 0$, any value of n would do). Then the random variable

$$\sum_{k=n}^{\infty} \alpha_0^k(e_{-k} + \gamma) + \epsilon(1 - \alpha_0^n)/(1 - \alpha_0)$$

is positive with positive probability since it has positive expectation. Thus $\sum_{k=0}^{\infty} \alpha_0^k(e_{-k} + \gamma)$ can be expressed as the sum of two random variables greater than $\epsilon(1 - \alpha_0^n)/(1 - \alpha_0)$ and $-\epsilon(1 - \alpha_0^n)/(1 - \alpha_0)$ with positive probability, respectively, yielding the desired result.

Finally, for $\alpha_0 \leq 0$, one obtains the conclusion of the Lemma by taking the limit, as k goes to infinity, of the extreme sides of

$$P(X_t < 0, t = 2, \ldots, k \mid X_1 = y) \leq P(e_t + \gamma < 0, t = 2, \ldots, k) \\ = [P(e_1 + \gamma < 0)]^{k-1},$$

and noting that $P(e_1 + \gamma < 0) < 1$ by (A1).

Proof of Lemma 2.2. Again, we prove only the first part of the Lemma since the proof for the second part is similar.

Suppose that $\alpha_0 \geq 1$. Then $X_{t-1} \leq x < 0$ implies

$$X_t - x = \alpha_0(X_{t-1} - x) + e_t + \gamma + (\alpha_0 - 1)x \\ \leq X_{t-1} - x + e_t + \gamma + (\alpha_0 - 1)x.$$

Define the random walk Y_t by $Y_0 = X_0 - x$, $Y_t = Y_{t-1} + e_t + \gamma + (\alpha_0 - 1)x$. Then by induction, it is easily seen that $Y_t \leq 0$ for all $t \geq 0$ implies $X_t - x \leq Y_t$. For $\alpha_0 = 1$ and $\gamma < 0$, or $\alpha_0 > 1$ and $x < -\gamma/(\alpha_0 - 1)$, the random variable $(\alpha_0 - 1)x + e_t + \gamma$ has negative expectation, which implies that the random walk Y_t drifts to minus infinity (Feller (1966), pp. 395, 396) or equivalently

$$P(Y_t \leq 0 \text{ for all } t > 0 \mid Y_0 = 0) > 0,$$

giving

$$P(X_t \leq x \text{ for all } t > 0 \mid X_0 = x) > 0.$$

To complete the proof of the Lemma, we need only show that for $\alpha_0 > 1$, there is a positive probability that starting at $X_0 = x < 0$, the X_t process becomes less than $-\gamma/(\alpha_0 - 1)$ after a finite number of steps. (Here $\gamma > 0$, otherwise there is nothing to prove.) That this is true follows from the fact that for all $y < 0$,

$$P(X_t < y \mid X_{t-1} = y) = P\{(\alpha_0 - 1)y + e_t + \gamma < 0\} \\ \geq P(e_1 + \gamma < 0) > 0.$$

Acknowledgement

This paper was presented at the Edinburgh International Workshop on Nonlinear Time Series under the auspices of the Science and Engineering Research Council (UK) in July 1989. The research of Chan was partially supported by National Science Foundation Grant DMS-9006.

References

Chan, K. S. (1988). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Technical Report 245, University of Chicago.

Chan, K. S., Petruccelli, J. D., Tong, H. and Woolford, S. W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22**, 267-279.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications II*, 2nd edition. John Wiley, New York.

Guo, M. and Petruccelli, J. D. (1990). On the null recurrence and transience of a first order SETAR model. Unpublished manuscript.

Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco.

Orey, S. (1971). *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand and Reinhold Company, London.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing* (Edited by C. H. Chen). Sijthoff and Noordhoff, Holland.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

Laboratory of Modelling and Computation, CNRS, Grenoble, B. P. 53x, 38041 Grenoble Cedex, France.

Department of Statistics, University of Chicago, Illinois 60637, U.S.A.

Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury CT2 7NF, U.K.

(Received November 1989; accepted January 1991)