The Annals of Statistics
1993, Vol. 21, No. 1, 520–533

CONSISTENCY AND LIMITING DISTRIBUTION OF THE LEAST SQUARES ESTIMATOR OF A THRESHOLD AUTOREGRESSIVE MODEL[1]

BY K. S. CHAN

University of Iowa

It is shown that, under some regularity conditions, the least squares estimator of a stationary ergodic threshold autoregressive model is strongly consistent. The limiting distribution of the least squares estimator is derived. It is shown that the estimator of the threshold parameter is N consistent and its limiting distribution is related to a compound Poisson process.

1. Introduction. Recently, there is much interest in nonlinear time series analysis. See, for example, Tong (1987) for a review on some recent work on nonlinear time series analysis. One of the most interesting nonlinear time series models is the self exciting threshold autoregressive model (SETAR) or sometimes just called the threshold autoregressive model (TAR). The SETAR model can exhibit many nonlinear phenomena such as limit cycles, jump resonance, harmonic distortion, modulation effects, chaos and so on. Specifically, the SETAR model is defined as below:

$$x_n = \sum_{1 \le i \le m+1} (a_{i0} + a_{i1}x_{n-1} + \cdots + a_{ip}x_{n-p} + c_i e_n) I(r_{i-1} < x_{n-d} \le r_i), \tag{1.1}$$

where $-\infty = r_0 < r_1 < \cdots < r_{m+1} = \infty$; a_{ij}'s are scalars; c_i's are positive numbers; m and p are nonnegative integers and d is a positive integer; e_n is i.i.d., zero mean, of unit variance and independent of the past $x_{n-1}, x_{n-2}, \dots$. The parameters r_i's and d are called the thresholds and the delay, respectively. Heuristically speaking, x_n is generated by one of the $m + 1$ "linear" mechanisms according to the level of x_{n-d}. See Tong (1983) for an introduction of the SETAR model.

Here, we only consider the case $m = 1$. Then, (1.1) becomes

$$x_n = \begin{cases} a_{10} + a_{11}x_{n-1} + \cdots + a_{1p}x_{n-p} + c_1 e_n, & \text{if } x_{n-d} \le r, \\ a_{20} + a_{21}x_{n-1} + \cdots + a_{2p}x_{n-p} + c_2 e_n, & \text{if } x_{n-d} > r, \end{cases} \tag{1.2}$$

Received October 1987; revised March 1992.
[1]This paper was presented at the Edinburgh International Workshop on Nonlinear Time Series under the auspices for the Sciences and Engineering Research Council (UK) in July 1989. Partially supported by NSF Grants DMS-90-06464 and DMS-91-18626.
AMS 1991 *subject classifications*. Primary 62M10; secondary 62J05.
Key words and phrases. Compound Poisson process, consistency, ergodicity, least squares estimation, limiting distribution, threshold autoregressive models.

where $r = r_1$. Let $A_i = (a_{i0}, a_{i1}, \ldots, a_{ip})'$, $i = 1, 2$. Throughout, it is assumed that $A_1 \neq A_2$, $r \in \mathscr{R}$, $d \leq p$ and p is known.

Given the data $\{x_0, x_1, x_2, \ldots, x_N\}$ generated from (1.2), the parameter $\theta_0 = (A_1', A_2', r, d)'$ can be estimated by the method of conditional least squares (CLS). In this paper, we consider the properties of the CLS estimator. Suppose (x_n) satisfying (1.2) is ergodic. In the case when r and d are known, it is not difficult to show that the estimators of a's are consistent and asymptotically normal. In practice, r is unknown and needs to be estimated. For the case $p = d = 1$, Petruccelli (1986) proved that $\hat{r}_N$, the CLSE of r, is strongly consistent. In Theorem 1, we show that, for the case of arbitrary p, the CLSE $\hat{\theta}_N$ of θ_0 is strongly consistent. Next, we consider the limiting distribution of $\hat{\theta}_N$ for the case when the autoregressive function is discontinuous. Then r is the location of the discontinuity of the autoregressive function. In Theorem 2, it is shown that $\hat{r}_N$ is N consistent and $N(\hat{r}_N - r)$ converges weakly to a random variable M_-, where $[M_-, M_+)$ is the unique random interval over which a compound Poisson process attains its global minimum. (For details, see Section 2.) Furthermore, $\hat{r}_N$ is asymptotically independent of $(\hat{A}_{1N}, \hat{A}_{2N})$ and the asymptotic distribution of the latter is equal to that in the case when r is known.

The organization of the paper is as follows. The main results are stated in Section 2. The proofs of Theorems 1 and 2 are given in Sections 3 and 4, respectively.

2. Main results. Let $\bar{\mathscr{R}} = \mathscr{R} \cup \{-\infty, \infty\}$ be equipped with the metric $\tilde{\delta}(x, y) = |\arctan(x) - \arctan(y)|$. Thus $\bar{\mathscr{R}}$ is compact. The parameter space Ω is $\mathscr{R}^{2p+2} \times \bar{\mathscr{R}} \times \{1, 2, \ldots, p\}$ equipped with the product metric. A general parameter in Ω is always denoted by $\theta = (B_1', B_2', z, q)'$ and the true parameter $\theta_0 = (A_1', A_2', r, d)'$.

Some further notation to be adopted throughout: $E_\theta(\cdot \mid \cdot)$ denotes the conditional expectation assuming θ to be the true parameter; all summations are, unless stated otherwise, from $n = p$ to N; statements involving random variables are meant to hold a.s. The CLSE $\hat{\theta}_N = (\hat{A}_{1N}', \hat{A}_{2N}', \hat{r}_N, \hat{d}_N)'$ is any measurable choice of $\theta \in \Omega$ which globally minimizes the (conditional) sum of square errors function

$$L_N(\theta) = \sum \left(x_n - E_\theta(x_n | F_{n-1})\right)^2, \tag{2.1}$$

where F_n is the σ algebra generated by $\{x_0, x_1, \ldots, x_n\}$. The minimization can be done in two steps. First, for fixed z and q, $L_N(\theta) = L_N(B_1, B_2, z, q) = L_{1N}(B_1, z, q) + L_{2N}(B_2, z, q)$, where $L_{1N}(B_1, z, q) = \Sigma'(x_n - B_1 \cdot Z_n)^2$ and $L_{2N}(B_2, z, q) = \Sigma''(x_n - B_2 \cdot Z_n)^2$, $Z_n = (1, x_{n-1}, x_{n-2}, \ldots, x_{n-p})'$; $\cdot$ stands for inner product; Σ' denotes summation over all n such that $x_{n-q} \leq z$ and Σ'' summation over the remaining n's. Let $L_{iN}(\cdot, z, q)$ be globally minimized at $B_{iN}(z, q)$, $i = 1, 2$. However, $L_N(B_{1N}(z, q), B_{2N}(z, q), z, q)$ has only a finite number of possible values. In general, there are infinitely many θ at which $L_N(\cdot)$ attains its global minimum, the one with the smallest z and q can be chosen as $\hat{\theta}_N$. In Theorem 2, $\hat{\theta}_N$ is assumed to be defined in this manner. We can estimate c_i^2 by $\hat{c}_i^2 = L_{iN}(\hat{A}_{iN}, \hat{r}_N, \hat{d}_N)/N_i$, $i = 1, 2$, where $N_1 = \Sigma I(x_{n-\hat{d}_N} \le \hat{r}_N)$ and $N_2 = N - N_1 - p + 1$.

The first main result obtained in this paper is the following theorem.

THEOREM 1. *Suppose that* (x_n) *satisfying* (1.2) *is stationary ergodic, having finite second moments and that the stationary distribution of* $(x_1, x_2, \dots, x_p)'$ *admits a density positive everywhere. Then,* $\hat{\theta}_N$ *is strongly consistent, that is,* $\hat{\theta}_N \to \theta_0$ *a.s. and so are* $\hat{c}_1^2$ *and* $\hat{c}_2^2$.

REMARK A.

(i) If $\max_i \Sigma_j |a_{ij}| < 1$ and e_n is absolutely continuous with a pdf positive everywhere, then the conditions in Theorem 1 hold. See Chan and Tong (1985) for some general sufficient conditions for (x_n) to be stationary ergodic. In general, the problem of determining the region of stationarity, say Ω^S, of a TAR model is still open. For $p = d = 1$, the problem is however completely solved. Then the conditions on the a's can be weakened to $a_{11} < 1$, $a_{21} < 1$ and $a_{11}a_{21} < 1$. For details, see Chan, Petruccelli, Tong and Woolford (1985). Note that Ω^S is a proper subset of the parameter space Ω.

(ii) Even if $\theta_0 \in \Omega^S$, $\hat{\theta}_N$ need not be in Ω^S. However, if θ_0 is an interior point of Ω^S, Theorem 1 implies that $P(\hat{\theta}_N \in \Omega^S) \to 1$ as $N \to \infty$.

(iii) Without the assumption of x_n being stationary ergodic, Theorem 1 may not be true as shown by the following.

EXAMPLE 1. Let $p = d = 1$ and consider

$$x_n = \begin{cases} a_{11}x_{n-1} + e_n, & \text{if } x_{n-1} \le 0, \\ a_{21}x_{n-1} + e_n, & \text{if } x_{n-1} > 0, \end{cases} \tag{2.2}$$

where e_n is i.i.d., zero mean and $E(|e|) < +\infty$. If $a_{21} > 1$, then (x_n) is transient. See, for example, Petruccelli and Woolford (1984). Specifically, if $x_0 > 0$, then there is positive probability that $x_n > 0$ for all n. Therefore, strong consistency of $\hat{\theta}_N$ is impossible.

Next, we study the limiting distribution of $\hat{\theta}_N$. Let

$$\mathbf{x}_n = (x_n, x_{n-1}, \dots, x_{n-p+1})'.$$

Then $(\mathbf{x}_n)$ is a Markov chain. Denote its l-step transition probability by $P^l(\mathbf{x}, A)$ where $\mathbf{x} \in \mathscr{R}^P$ and A is a Borel set. The following set of regularity conditions will be required later.

CONDITION 1. $(\mathbf{x}_n)$ admits a unique invariant measure $\pi(\cdot)$ such that $\exists\, K$, $\rho < 1$, $\forall\, \mathbf{x} \in \mathscr{R}^P$, $\forall\, n \in \mathscr{N}$, $\|P^n(\mathbf{x}, \cdot) - \pi(\cdot)\| \le K(1 + |\mathbf{x}|)\rho^n$, where $\|\cdot\|$ and $|\cdot|$ denote the total variation norm and the Euclidean norm, respectively.

CONDITION 2. e_n is absolutely continuous with a uniformly continuous and positive pdf. Furthermore, $E(e_n^4) < \infty$.

CONDITION 3. (x_n) is stationary with its marginal pdf denoted by $\pi(\cdot)$. Also, $E(x_n^4) < \infty$.

CONDITION 4. The autoregressive function is discontinuous, that is, $\exists$ $Z^* = (1, z_{p-1}, z_{p-2}, \ldots, z_0)'$ such that $(A_1 - A_2) \cdot Z^* \neq 0$ and $z_{p-d} = r$.

Some remarks on the above conditions will be given after the statement of Theorem 2. Suppose Conditions 1–4 hold, then it follows from Theorem 1 that $\hat{d}_N = d$ eventually. Hence, with no loss of generality, assume d is known. To describe the limiting distribution of $\hat{r}_N$, consider two independent compound Poisson processes $\{\tilde{L}^{(1)}(z),\ z \geq 0\}$ and $\{\tilde{L}^{(2)}(z),\ z \geq 0\}$, both with rate $\pi(r)$, $\tilde{L}^{(1)}(0) = \tilde{L}^{(2)}(0) = 0$ a.s. and the distributions of jump being given by the conditional distribution of $\zeta_1 \doteq (A_1 - A_2) \cdot Z_p(2c_1 e_p + (A_1 - A_2) \cdot Z_p)$ given $x_{p-d} = r_-$ and the conditional distribution of $\zeta_2 \doteq (A_2 - A_1) \cdot Z_p(2c_2 e_p + (A_2 - A_1) \cdot Z_p)$ given $x_{p-d} = r_+$, respectively. The former conditional distribution is the limiting conditional distribution of ζ_1 given $r - \delta < x_{p-d} \leq r$ as $\delta \downarrow 0$ and the latter that of ζ_2 given $r < x_{p-d} \leq r + \delta$ as $\delta \downarrow 0$. The existence of these limits as well as some of their properties stated below follow from Condition 2 and a result in Neveu [(1965), page 124]. It is worthwhile to note that although the stationary univariate marginal distribution of x_p has a continuous pdf, the pdf of a finite dimensional distribution of (x_n) is possibly discontinuous over a number of hyperplanes. These facts can easily be deduced from the invariant equation which the stationary density has to obey. Note that both distributions of jump have positive means and are absolutely continuous. So the two random walks associated with the compound Poisson processes tend to $+\infty$ a.s. Hence $\exists$ a random interval $[M_-, M_+)$ on which the process $\{\tilde{L}^{(1)}(-z)I(z \leq 0) + \tilde{L}^{(2)}(z)I(z > 0),\ z \in \mathscr{R}\}$ attains its global minimum and nowhere else. [We work with the left continuous version for $\tilde{L}^{(1)}(\cdot)$ and the right continuous version for $\tilde{L}^{(2)}(\cdot)$.]

THEOREM 2. *Suppose Conditions* 1–4 *hold. Then* $N(\hat{r}_N - r)$ *converges in distribution to* M_-. *Furthermore,* $N(\hat{r}_N - r)$ *is asymptotically independent of* $\sqrt{N}(\hat{A}_{1N} - A_1, \hat{A}_{2N} - A_2)$ *and the latter is asymptotically normal with a distribution same as that for the case when r is known.*

REMARK B.

(i) Geometric ergodicity of $(\mathbf{x}_n)$ only requires that $\|P^n(\mathbf{x}, \cdot) - \pi(\cdot)\| = O(\rho^n)$ for some $\rho < 1$. [For a discussion of geometric ergodicity, see, e.g., Nummelin (1984)]. Hence, Condition 1 is stronger than geometric ergodicity. However, it is shown in Chan (1989) that if $(\mathbf{x}_n)$ satisfies an appropriate drift condition, then Condition 1 holds. Indeed, if Condition 2 holds and $\max_{i=1,2} \sum_{j=1}^{p} |a_{ij}| < 1$, then Condition 1 obtains and $E(x_n^4) < \infty$. This follows from the discussions in Chan and Tong (1985) and Chan (1989). For $p = 1$, the above condition on the a's can be weakened to $a_{11} < 1$, $a_{21} < 1$ and $a_{11}a_{21} < 1$.

(ii) Condition 2 entails that $\pi(\cdot)$ is absolutely continuous with its pdf bounded away from 0 and ∞ over each bounded set. This follows from the invariant equation for the stationary distribution of $(\mathbf{x}_n)$ given in, for example, Chan and Tong (1985). It is clear that $\pi(\cdot)$ is the marginal distribution of the first coordinate of $(\mathbf{x}_n)$. Hence $\pi(\cdot)$ is positive everywhere.

(iii) Under Condition 4, r can be considered as the "location" parameter of the discontinuity of the autoregressive function. For i.i.d. observations, statistical inference of the location of the discontinuity in the density function is considered in, for example, Chernoff and Rubin (1956). See also Chapter V in Ibragimov and Has'minskii (1981).

We conclude Section 2 by mentioning two areas of future research. It would be interesting to have efficient methods to compute the percentage points for the limiting distribution of $\hat{r}_N$. Another problem is to work out the limiting distribution of the CLSE when the autoregressive function is continuous.

3. Strong consistency of the CLSE.

Proof of Theorem 1. We consider the simple case that $c_1 = c_2 = \sigma_e$, the proof for the general case being similar and hence omitted. Then (1.2) becomes

$$(3.1)\qquad x_n = \begin{cases} a_{10} + a_{11}x_{n-1} + \cdots + a_{1p}x_{n-p} + e_n, & \text{if } x_{n-d} \le r, \\ a_{20} + a_{21}x_{n-1} + \cdots + a_{2p}x_{n-p} + e_n, & \text{if } x_{n-d} > r, \end{cases}$$

where e_n is i.i.d., zero mean and of finite nonzero variance $= \sigma_e^2$. Consider the following decomposition of a summand of $L_N(\theta)$:

$$(3.2a)\qquad \left(x_n - E_\theta(x_n | F_{n-1})\right)^2 = R_{1n}(\theta) + R_{2n}(\theta) + R_{3n}(\theta) + R_{4n}(\theta),$$

where

$$(3.2b)\qquad R_{1n}(\theta) = (x_n - B_1 \cdot Z_n)^2 I(x_{n-q} \le z, x_{n-d} \le r),$$

$$(3.2c)\qquad R_{2n}(\theta) = (x_n - B_1 \cdot Z_n)^2 I(x_{n-q} \le z, x_{n-d} > r),$$

$$(3.2d)\qquad R_{3n}(\theta) = (x_n - B_2 \cdot Z_n)^2 I(x_{n-q} > z, x_{n-d} \le r),$$

$$(3.2e)\qquad R_{4n}(\theta) = (x_n - B_2 \cdot Z_n)^2 I(x_{n-q} > z, x_{n-d} > r).$$

Substituting (3.1) into (3.2b), we get

$$\begin{aligned} R_{1n}(\theta) &= e_n^2 I(x_{n-q} \le z, x_{n-d} \le r) + 2e_n(A_1 - B_1) \cdot Z_n I(x_{n-q} \le z, x_{n-d} \le r) \\ (3.3)\qquad &\quad + ((A_1 - B_1) \cdot Z_n)^2 I(x_{n-q} \le z, x_{n-d} \le r) \\ &= \phi_{1n}^{(1)}(\theta) + \phi_{2n}^{(1)}(\theta) + \phi_{3n}^{(1)}(\theta), \end{aligned}$$

where, for example, $\phi_{1n}^{(1)}(\theta) = e_n^2 I(x_{n-q} \le z,\ x_{n-d} \le r)$. $\phi_{jn}^{(k)}(\theta)$ could be defined analogously so that $R_{kn}(\theta) = \sum_{j=1}^{3} \phi_{jn}^{(k)}(\theta)$.

LEMMA 1. *Let U denote an open neighborhood of θ. Suppose that the conditions in Theorem 1 are satisfied. Then for each $\theta \in \Omega$, $\forall\ j = 1,2,3$, $\forall\ k = 1,2,3,4$,*

$$(3.4)\qquad E\Big(\sup_{\theta^* \in U} \big|\phi_{jn}^{(k)}(\theta^*) - \phi_{jn}^{(k)}(\theta)\big|\Big) \to 0 \qquad \text{as } U \text{ shrinks towards } \theta.$$

We postpone the proof of this lemma and first see how it could be used in showing the strong consistency of $\hat{\theta}_N$. This is done in three steps.

STEP 1.

CLAIM 1. $\exists\ \tau > 0$ such that, for N sufficiently large, $\hat{\theta}_N$ lies in $\Omega_1 = \{\theta \in \Omega: |z - r| \le \tau\}$ a.s.

We now verify Claim 1. Let $V = \{\theta \in \Omega: |A_i - B_i| = 1,\ i = 1,2\}$ and $\varepsilon > 0$ to be determined later. From the compactness of V and Lemma 1, $\exists$ $\{\theta_1, \theta_2, \dots, \theta_m\} \subseteq V$ and a finite partition $\{Q_1, Q_2, \dots, Q_m\}$ of V such that, $\forall\ i$ and $\forall\ 1 \le k \le 4$, $\forall\ 1 \le j \le 3$,

$$(3.5a)\qquad E\Big(\sup_{\theta \in Q_i} \big|\phi_{jn}^{(k)}(\theta) - \phi_{jn}^{(k)}(\theta_i)\big|\Big) \le \varepsilon$$

and hence,

$$(3.5b)\qquad E\Big(\inf_{\theta \in Q_i} \phi_{jn}^{(k)}(\theta)\Big) \ge E\big(\phi_{jn}^{(k)}(\theta_i)\big) - \varepsilon.$$

Note that $\phi_{1N}^{(1)}(\theta) = \phi_{1N}^{(1)}(\theta')$, $\phi_{2N}^{(1)}(\theta) = |A_1 - B_1|\phi_{2N}^{(1)}(\theta'')$ and $\phi_{3N}^{(1)}(\theta) = |A_1 - B_1|^2\phi_{3N}^{(1)}(\theta''')$ for some θ', θ'' and $\theta''' \in V$. Moreover, these can be chosen such that $\forall\ \theta$, θ', θ'' and θ''' have the same z-coordinate that θ has. Similar properties hold for the other ϕ's. Let $z_1 = r + \tau$ and $\Omega(z_1) = \{\theta \in \Omega: z \ge z_1\}$. By increasing ε in (3.5) to 2ε if necessary, it can be assumed that if $Q_i \cap \Omega(z_1) \ne \phi$, then $\theta_i \in \Omega(z_1)$. In view of (3.5), (3.3) and the preceding discussion, it is not difficult to see that $\exists$ a partition $\{P_1, P_2, \dots, P_M\}$ of $\Omega(z_1)$ such that, $\forall\ i$,

$$\begin{aligned} E\Big(\inf_{\theta \in P_i} \big(x_p - E_\theta(x_p|F_{p-1})\big)^2\Big) &\ge E\Big(\inf_{P_i}\big(R_{1p}(\theta) + R_{2p}(\theta)\big)\Big) \\ &\ge \sigma_e^2\big(I_1(z_1) + I_2(z_1)\big) - 2\varepsilon \\ &\quad - \varepsilon\big(|A_1 - B_1| + |A_2 - B_1|\big) \\ &\quad + (2\beta - \varepsilon)\big(|A_1 - B_1|^2 + |A_2 - B_1|^2\big), \end{aligned} \qquad (3.6a)$$

where

$$I_1(z_1) = \min_{1 \le k \le p} E\big(I(x_{p-k} \le z_1, x_{p-d} \le r)\big), \tag{3.6b}$$

$$I_2(z_1) = \min_{1 \le k \le p} E\big(I(x_{p-k} \le z_1, x_{p-d} > r)\big), \tag{3.6c}$$

$$\beta = \inf_{\theta \in V \cap \Omega(z_1)} \min\Big(E\big(\phi_{3p}^{(1)}(\theta)\big), E\big(\phi_{3p}^{(2)}(\theta)\big)\Big). \tag{3.6d}$$

Note that $\beta > 0$ for sufficiently large z_1 and $I_1(z_1) + I_2(z_1) \to 1$ as $z_1 \to \infty$ (equivalently, $\tau \to \infty$).

Suppose ε is chosen so that $\varepsilon < \beta$. Let $c = |A_1 - A_2|/2 > 0$. Then $\max(|A_1 - B_1|, |A_2 - B_1|) \ge c$. By routine arguments, it can be verified that, for sufficiently small $0 < \varepsilon < 1$, the RHS of (3.6a) is

$$\ge \sigma_e^2(I_1(z_1) + I_2(z_1)) - 2\varepsilon - \varepsilon^2/(4\beta) + \beta c^2 - \varepsilon c. \tag{3.7}$$

For sufficiently small ε and large enough τ, the RHS of (3.7) is greater than or equal to $\sigma_e^2 + \varepsilon$.

However, $E((x_p - E_{\theta_0}(x_p|F_{p-1}))^2) = \sigma_e^2$. By the law of large numbers for (x_n) and arguing as in Huber (1967), for N sufficiently large, $\hat{\theta}_N \notin P_i$. Since i is arbitrary, we conclude that $\hat{\theta}_N \notin \{\theta \in \Omega: z \ge z_1\}$ for sufficiently large N. By considering $R_{3p}(\theta)$ and $R_{4p}(\theta)$ instead and arguing similarly, it is easily seen that Claim 1 holds.

Step 2.

Claim 2. $\exists\ w > 0$ such that, for N sufficiently large, $\hat{\theta}_N$ lies in $\Omega_2 = \{\theta \in \Omega_1: |A_1 - B_1| < w, |A_2 - B_2| < w\}$.

This can be verified by noting that $(x_p - E_\theta(x_p|F_{p-1}))^2 \ge R_{1p}(\theta) + R_{4p}(\theta)$ and arguing as in the previous claim.

Step 3. Let $L(\theta) = E(x_{p+1} - E_\theta(x_{p+1})|\mathscr{F}_p))^2$. Then it can be verified that $L(\theta_0) = \sigma_e^2$ and $L(\theta) > L(\theta_0)$, $\forall\ \theta \ne \theta_0$. From the previous two steps, without loss of generality, the domain of θ can be restricted to Ω_2. Also, it follows from Lemma 1 that, $\forall\ \theta^* \in \Omega$,

$$E\left(\inf_{\theta \in U}\left(x_{p+1} - E_\theta\big(x_{p+1}|\mathscr{F}_p\big)\right)^2\right) \to L(\theta^*)$$

as V, an open ball of θ^*, shrinks towards θ^*. Thus, proceeding as in the proof of Theorem 1 in Huber (1967), we have $\hat{\theta}_N \to \theta_0$ and $L_N(\hat{\theta}_N)/N \to \sigma_e^2$ a.s.

This completes the proof of Theorem 1. □

Proof of Lemma 1. Let $\theta = (B_1', B_2', z, q)' \in \Omega$ and $\eta > 0$. Define

$$U(\eta) = \big\{\theta^* = (C_1', C_2', u, q)' \in \Omega: |C_i - B_i| < \eta, i = 1, 2; \tilde{\delta}(u, z) < \eta\big\}.$$

Then one can verify that

$$\lim_{\eta \to 0} E\left(\sup_{\theta^* \in U(\eta)} \left| \phi_{1p}^{(3)}(\theta^*) - \phi_{1p}^{(3)}(\theta) \right| \right) = 0. \tag{3.8}$$

This is because if $z \in \mathscr{R}$ and $\theta^* \in U(\eta)$, then

$$\begin{aligned} \left| \phi_{1p}^{(3)}(\theta^*) - \phi_{1p}^{(3)}(\theta) \right| &= \Big| \big((A_1 - C_1) \cdot Z_p \big)^2 I(x_{p-q} \le z, x_{p-d} \le r) \\ &\qquad - \big((A_1 - B_1) \cdot Z_p \big)^2 I(x_{p-q} \le u, x_{p-d} \le r) \Big| \\ &\le 2(|A_1 - B_1| + \eta)\eta |Z_p|^2 + \big((A_1 - B_1) \cdot Z_P \big)^2 \\ &\qquad \times \big| I(x_{p-q} \le u, x_{p-d} \le r) - I(x_{p-q} \le z, x_{p-d} \le r) \big| \\ &\le 2(|A_1 - B_1| + \eta)\eta |Z_p|^2 + |A_1 - B_1|^2 |Z_p|^2 \\ &\qquad \times I\big(|x_{p-q} - z| \le |u - z| \big). \end{aligned} \tag{3.9}$$

The case when $z = \pm\infty$ is similar and hence omitted. The rest of the lemma can be proved similarly. □

4. Limiting distribution of the CLSE.

As discussed in Section 2, with no loss of generality, d is assumed known. So, the parameter becomes $\theta = (B_1', B_2', z)'$ and Ω is modified accordingly. We first establish some auxiliary results.

Proposition 1. *Suppose Condition 1 to Condition 4 hold. Then* $\hat{r}_N = r + O_p(1/N)$.

Proof. Since $\hat{\theta}_N$ is strongly consistent, with no loss of generality, the parameter space can be restricted to a neighborhood of θ_0, say,

$$\omega(\Delta) = \{\theta \in \Omega : |B_i - A_i| < \Delta, i = 1, 2; |z - r| < \Delta\}, \tag{4.1}$$

for some $1 > \Delta > 0$ to be determined later. First, assume $p = d = 1$. For simplicity of notation, assume $r = 0$. Then, it suffices to verify the following claim.

Claim 1. $\forall\ \varepsilon > 0$, $\exists\ K$ such that with probability greater than $1 - \varepsilon$, $\theta \in \omega(\Delta)$, $|z| > K/N \Rightarrow L_N(B_1, B_2, z) - L_N(B_1, B_2, 0) > 0$.

First, consider the case that $z > 0$. Then

$$\begin{aligned} &L_N(B_1, B_2, z) - L_N(B_1, B_2, 0) \\ &\quad = \sum \left\{ (x_n - B_1 \cdot Z_n)^2 - (x_n - B_2 \cdot Z_n)^2 \right\} I(0 < x_{n-1} \le z) \\ &\quad \ge 2c_2(B_2 - B_1) \cdot \sum Z_n e_n I(0 < x_{n-1} \le z) \\ &\qquad + \sum \big((2A_2 - B_1 - B_2) \cdot Z_n \big)(B_2 - B_1) \cdot Z_n I(0 < x_{n-1} \le z). \end{aligned} \tag{4.2}$$

If Δ is sufficiently small, then it follows from Condition 4 that the second term is greater than or equal to $\delta^2 \sum I(0 < x_{n-1} \le z)$ for some $\delta > 0$. The first term on the RHS of (4.2) is bounded in absolute value by $\nu(|\sum e_n I(0 < x_{n-1} \le z)| + |\sum x_{n-1} e_n I(0 < x_{n-1} \le z)|)$ for some constant ν independent of N. Define

$$Q(z) = E(I(0 < x \le z)), \qquad 0 < z \le \Delta. \tag{4.3}$$

CLAIM 2. $\forall\, \varepsilon > 0$, $\forall\, \eta > 0$, $\exists\, K > 0$ such that, $\forall\, N$,

$$P\Big(\sup_{\Delta \ge z > K/N} \Big|\sum I(0 < x_{n-1} \le z)/(NQ(z)) - 1\Big| < \eta\Big) > 1 - \varepsilon, \tag{4.4a}$$

$$P\Big(\sup_{\Delta \ge z > K/N} \Big|\sum e_n I(0 < x_{n-1} \le z)/(NQ(z))\Big| < \eta\Big) > 1 - \varepsilon, \tag{4.4b}$$

$$P\Big(\sup_{\Delta \ge z > K/N} \Big|\sum x_{n-1} e_n I(0 < x_{n-1} \le z)/(NQ(z))\Big| < \eta\Big) > 1 - \varepsilon. \tag{4.4c}$$

Suppose the above claim is valid for the present moment. Let $\varepsilon > 0$ be given and $\eta > 0$ be chosen so that $-2\nu\eta + \delta^2(1 - \eta) > 0$. It follows from the preceding claim that $\exists\, K(\varepsilon, \eta) > 0$ such that with probability greater than $1 - 3\varepsilon$, $\Delta \ge z > K/N$ implies that

$$\begin{aligned} &\{L_N(B_1, B_2, z) - L_N(B_1, B_2, 0)\}/(NQ(z)) \\ &\quad \ge -2\nu\eta + \delta^2(1 - \eta) > 0 \end{aligned} \tag{4.5}$$

and hence the validity of Claim 1 under the further condition that $z > 0$.

We now verify Claim 2. Define

$$Q_N(z) = \sum I(0 < x_{n-1} \le z)/N, \tag{4.6a}$$

$$R_N(z) = \sum x_{n-1} e_n I(0 < x_{n-1} \le z)/N, \tag{4.6b}$$

$$\tilde{R}_N(z_1, z_2) = \sum |x_{n-1} e_n| I(z_1 < x_{n-1} \le z_2)/N, \tag{4.6c}$$

$$\tilde{R}(z_1, z_2) = E(\tilde{R}_N(z_1, z_2)). \tag{4.6d}$$

By choosing Δ sufficiently small, it follows from (ii) in Remark B that $\exists\, 0 < m < M < \infty$ and H, all independent of N, such that, $\forall\, z, z_1, z_2$ in $[0, \Delta)$,

$$mz \le Q(z) \le Mz, \tag{4.7a}$$

$$\operatorname{var}(I(0 < x_{n-1} \le z)) \le HQ(z), \tag{4.7b}$$

$$E(|x_{n-1} e_n| I(z_1 < x_{n-1} \le z_2)) \le H(Q(z_2) - Q(z_1)), \tag{4.7c}$$

$$\operatorname{var}(|x_{n-1} e_n| I(z_1 < x_{n-1} \le z_2)) \le H(Q(z_2) - Q(z_1)), \tag{4.7d}$$

$$\tilde{R}(z_1, z_2) \le H(Q(z_2) - Q(z_1)). \tag{4.7e}$$

Using Condition 1, it can be verified that $\forall\ b > 0$, $\exists\ H$ such that $\forall\ z, z_1, z_2 \in [-b, b]$, $\forall\ N$

$$\text{var}(NQ_N(z)) \le NHQ(z), \tag{4.8a}$$

$$\text{var}(N\tilde{R}_N(z_1, z_2)) \le NH(Q(z_2) - Q(z_1)), \tag{4.8b}$$

$$\text{var}(NR_N(z)) \le NHQ(z). \tag{4.8c}$$

Indeed, for example, (4.8b) follows readily from the fact that, by direct calculation, uniformly for z_1, $z_2 \in [-b, b]$, $\text{cov}(|e_1 I_0|, |e_k I_{k-1}|) = O(\rho^k(Q(z_2) - Q(z_1)))$ where ρ is as defined in Condition 1 and $I_k = x_k I(z_1 < x_k \le z_2)$, $\forall\ k \in \mathscr{N}$. In view of (4.7) without loss of generality, let $Q(z) = z$. The following facts are pertinent in establishing Claim 2. They follow from (4.8) and Markov's inequality. All the following suprema and summations are taken over all $i \in \mathscr{N}$ and such that all quantities involved are well defined.

Let $b > 1$, $K > 0$ and $\eta > 0$

$$P\big(\sup|Q_N(b^iK/N)/(b^iK/N) - 1| > \eta\big) \le H/(\eta^2 K(1 - b^{-1})). \tag{4.9a}$$

For $0 < x \le y \le bx \le \Delta$ with $|Q_N(x)/x - 1| < \eta$ and $|Q_N(bx)/(bx) - 1| < \eta$, we have

$$\begin{aligned}(1 - \eta)/b - 1 &\le Q_N(x)/(bx) - 1 \le Q_N(y)/y - 1\\ &\le bQ_N(by)/(by) - 1 \le b(1 + \eta) - 1.\end{aligned} \tag{4.9b}$$

$$P\big(\sup|R_N(b^iK/N)/(b^iK/N)| > \eta\big) \le H/(\eta^2 K(1 - b^{-1})). \tag{4.9c}$$

$$\begin{aligned}P\Big(\sup\big|\tilde{R}_N(b^iK/N, b^{i+1}K/N) - \tilde{R}(b^iK/N, b^{i+1}K/N)\big|\Big/\\ (b^iK/N) > \eta\Big) \le Hb/(\eta^2 K).\end{aligned} \tag{4.9d}$$

$$\sup \tilde{R}(b^iK/N, b^{i+1}K/N)/(b^iK/N) \le H(b - 1). \tag{4.9e}$$

By first choosing $\eta > 0$ and $b > 1$ sufficiently small and then K sufficiently large, (4.9a) and (4.9b) imply the validity of (4.4a) and (4.9c–e) imply that (4.4c) hold. Equation (4.4b) can be similarly proved.

The case of $z < 0$ is similar and hence the validity of Claim 1. This completes the proof for the case $p = d = 1$. For the general case, let Z^* be as in Condition 4. Then, $\exists\ \gamma > 0$ such that $(A_1 - A_2) \cdot Z$ is bounded away from 0 for all Z such that $|Z - Z^*| \le \gamma$. On the right-hand side of (4.2), replace each occurrence of $I(0 < x_{n-1} \le z)$ by $I(0 < x_{n-d} \le z;\ |Z_n - Z^*| \le \gamma)$. Then, with suitable modification, the preceding proof would go through. This completes the proof of Proposition 1. □

We now consider the limiting behavior of the normalized profile sum of squares errors function. Define, for $z \in \mathscr{R}$,

$$\tilde{L}_N(z) = L_N(B_{1N}(r + z/N), B_{2N}(r + z/N), r + z/N) - L_N(B_{1N}(r), B_{2N}(r), r). \tag{4.10}$$

Note that since d is assumed known, $B_{1N}(r + z/N, d)$ is simply written as $B_{1N}(r + z/N)$ and so on. [The B's are defined after (2.1).]

PROPOSITION 2. *Under Conditions* 1–4, $(\{\tilde{L}_N(-z),\ z \ge 0\}, \{\tilde{L}_N(+z),\ z \ge 0\})$ *converges weakly to* $(\{\tilde{L}^{(1)}(z),\ z \ge 0\}, \{\tilde{L}^{(2)}(z),\ z \ge 0\})$ *in* $D[0,\infty) \times D[0,\infty)$, *the product space being equipped with the product Skorohod metric. Here,* $\{\tilde{L}^{(1)}(z)\}$ *and* $\{\tilde{L}^{(2)}(z)\}$ *are described before the statement of Theorem* 2. [*See, e.g., Kushner* (1984, *pages* 29–33) *for a discussion of* $D[0,\infty)$ *and the Skorohod metric.*]

PROOF. For the sake of simplicity, assume $r = 0$. Let $1 \le l \le p$ and $k > 0$. Since $E(x_p^4) < \infty$, $x_p x_{p-l} I(-K/N \le x_{p-d} \le K/N) = o_p(N^{-1/2})$. Using this result and by routine but lengthy arguments, it can be verified that, $\forall\ K > 0$,

$$\sup_{|z| \le K/N} |B_{iN}(z) - B_{iN}(0)| = o_p(N^{-1/2}), \qquad i = 1, 2. \tag{4.11}$$

From this, it follows that

$$\sup_{|z| \le K} \left|\tilde{L}_N(z) - (L_N(A_1, A_2, z/N) - L_N(A_1, A_2, 0))\right| = o_p(1). \tag{4.12}$$

For details of arguments leading to (4.11) and (4.12) for the case $p = d = 1$, see Chan (1988). Owing to (4.12), we shall proceed as if $\tilde{L}_N(z) = L_N(A_1, A_2, z/N) - L_N(A_1, A_2, 0)$. For $z > 0$, $\tilde{L}_N(z) = \Sigma(A_2 - A_1) \cdot Z_n(2c_2 e_n + (A_2 - A_1) \cdot Z_n) I(0 < x_{n-d} \le z/N)$. It follows from Condition 2 that $\forall\ b > 0$, $\exists\ H > 0$ such that for any interval $I \subseteq [-b, b]$ with length $l(I)$,

$$P(x_0 \in I, x_k \in I) \le H(l(I))^2, \qquad \forall\ k \in \mathscr{N}. \tag{4.13}$$

From this follows the tightness of $(\{\tilde{L}_N(-z),\ z \ge 0\}, \{\tilde{L}_N(z),\ z \ge 0\})$. For a related argument, see, for example, the proof of Lemma 3.2 in Ibragimov and Has'minskii [(1981), page 261].

Thus, to complete the proof, it suffices to demonstrate the appropriate convergence of finite dimensional distributions. Let $z > 0$ be fixed. Let $\varepsilon = 1/N$. Consider the following process indexed by ε:

$$\begin{aligned} &x^\varepsilon(t) = X^\varepsilon_{[Nt]}, \qquad 0 \le t \le 1, \\ &X_0^\varepsilon = 0, \qquad X_{n+1}^\varepsilon = X_n^\varepsilon + J_n^\varepsilon, \qquad n \ge 1, \\ &J_n^\varepsilon = (A_2 - A_1) \cdot Z_n(2c_2 e_n + (A_2 - A_1) \cdot Z_n) I(0 < x_{n-d} \le z/N) \end{aligned} \tag{4.14}$$

Here, $[\cdot]$ denotes the integral part of the expression inside the square bracket. Define $\xi_n = (x_n, x_{n-1}, \dots, x_{n-p})'$. Note that $x^\varepsilon(1) = \tilde{L}_N(z)$ and that J_n^ε is a functional of ξ_n. Again, it follows from (4.13) that $\{x^\varepsilon(t),\ 0 \le t \le 1\}$ is tight in

$D[0,1]$. Applying the direct averaging method, it can be shown that $\{x^{\varepsilon}(t), 0 \le t \le 1\}$ converges weakly in $D[0,1]$ to $\{\wp(t), 0 \le t \le 1\}$, a compound Poisson process with rate $\pi(r)z$ and the distribution of jump same as the conditional distribution of $(A_2 - A_1) \cdot Z_p(2c_2e_p + (A_2 - A_1) \cdot Z_p)$ given $x_{p-d} = r_+$. This can be done by adapting the arguments given in Kushner [(1984), Sections 2, 3 and 7 of Chapter 5]. As a corollary of the weak convergence of $\{x^{\varepsilon}(t)\}$ to the compound Poisson process, we have the weak convergence of $\tilde{L}_N(z)$ to $\tilde{L}^{(2)}(z)$. Employing the Cramer–Wold device, similar arguments yield the convergence of finite dimensional distributions of $(\{\tilde{L}_N(-z), z \ge 0\}, \{\tilde{L}_N(z), z \ge 0\})$ to those of $(\{\tilde{L}^{(1)}(z), z \ge 0\}, \{\tilde{L}^{(2)}(z), z \ge 0\})$.

Now, we outline the adaptation of Kushner's arguments to the present case. To prove that $\{x^{\varepsilon}(t), 0 \le t \le 1\}$ converges weakly to $\{\wp(t), 0 \le t \le 1\}$, it suffices to show that the weak limit of any convergent subsequence $\{x^{\varepsilon_n}(t), 0 \le t \le 1\}$ is a solution to the following martingale problem: For any function f with compact support and continuous second derivative,

$$x(t) - \int_0^t Af(x(s))\, ds \quad \text{is a martingale,} \tag{4.15}$$

where $Af(w) = \pi(r)z\int(f(y+w) - f(w))q(dy)$ and $q(dy)$ is the probability measure induced by the conditional distribution of $(A_2 - A_1) \cdot Z_p(2c_2e_p + (A_2 - A_1) \cdot Z_p)$ given $x_{p-d} = r_+$. Then the convergence result follows from the fact that the above martingale problem admits $\{\wp(t), 0 \le t \le 1\}$ as its unique solution. In the rest of the proof, we supply further technical details in verifying (4.15).

Below, $E_m(\cdot)$ and $P_m(\cdot)$ denote, respectively, the conditional expectation and the conditional probability of the enclosed expression given $\{\xi_j, j < m\}$. It is also assumed that the probability space is sufficiently large to contain all the random variables discussed below. Let n_ε be an integer chosen so that $n_\varepsilon \to \infty$ and $\delta_\varepsilon = \varepsilon n_\varepsilon \to 0$ as $\varepsilon \to 0$. Let $k \ge d+1$ be a fixed integer. Fix $f(\cdot)$, a function with compact support and continuous second derivative, and define

$$\tilde{A}^{\varepsilon}(t) = \frac{1}{n_\varepsilon} \frac{\sum_{j=ln_\varepsilon}^{(l+1)n_\varepsilon - 1} E_{ln_\varepsilon - k}\big(f(X_{j-1}^{\varepsilon}) - f(X_j^{\varepsilon})\big)}{\varepsilon}, \tag{4.16}$$

for $t \in [l\delta_\varepsilon, (l+1)\delta_\varepsilon)$ and set $\hat{A}^{\varepsilon}(t) = \int_0^t \tilde{A}^{\varepsilon}(s)\, ds$. It is easily seen that

$$\begin{aligned} &E_{ln_\varepsilon - k}\big(f(X_{j+1}^{\varepsilon}) - f(X_j^{\varepsilon})\big) \\ &\quad = P_{ln_\varepsilon - k}(0 < x_{j-d} < z\varepsilon) E_{ln_\varepsilon - k}\big(f(X_j^{\varepsilon} + J_j^{\varepsilon}) - f(X_j^{\varepsilon}) | 0 < x_{j-d} < z\varepsilon\big). \end{aligned}$$

It follows from Condition 1 and Condition 2 that for any compact set K and any scalar X, $\exists\, M$, $\forall\, \varepsilon > 0$, $j \ge ln_\varepsilon$,

$$P_{ln_\varepsilon - k}(0 < x_{j-d} < z\varepsilon) \le M\varepsilon \tag{4.17}$$

and uniformly for $\xi_{ln_\varepsilon - k} \in K$,

$$\lim_{j \to \infty} P_{ln_\varepsilon - k}(0 < x_{j-d} < z\varepsilon)/\varepsilon = \pi(r)z + o(\varepsilon) \tag{4.18}$$

and

$$\lim_{j\to\infty} E_{ln_\varepsilon - k}\big(f(X+J_j^\varepsilon) - f(X)|0 < x_{j-d} < z\varepsilon\big) \tag{4.19}$$
$$= E\big(f(X+J_n^\varepsilon) - f(X)|0 < x_{n-d} < z\varepsilon\big).$$

It follows from (4.17) that $\{x^\varepsilon(t), \hat{A}^\varepsilon(t), 0 \le t \le 1\}$ is tight in $D^2[0,1]$. We want to characterize the weak limit of any convergent subsequence of $\{x^\varepsilon(t), \hat{A}^\varepsilon(t), 0 \le t \le 1\}$. For simplicity, assume that the latter process converges to $\{x(t), \hat{A}(t), 0 \le t \le 1\}$ with probability 1. We are going to show that

$$\hat{A}(t) = Af(x(t)) \tag{4.20}$$

and for arbitrary m, t, s and $s_1 < s_2 < \cdots < s_m < t < t+s$ and any bounded and continuous functions $h(\cdot)$,

$$E\Big\{h\big(x(s_j), j \le m\big) \times \big[f(x(t+s)) - f(x(t)) \tag{4.21}$$
$$-\big(\hat{A}(t+s) - \hat{A}(t)\big)\big]\Big\} = 0.$$

Equations (4.20) and (4.21) imply that $\{x(t), 0 \le t \le 1\}$ solves the above martingale problem.

Thus, to complete the proof, it remains to verify (4.20) and (4.21). Note that for any bounded and continuous $h(\cdot)$, we have, as $\varepsilon \to 0$,

$$E\Bigg\{h\big(x^\varepsilon(s_j), 1 \le j \le m\big) \tag{4.22}$$
$$\times\bigg[f(x^\varepsilon(t+s)) - f(x^\varepsilon(t)) - \int_t^{t+s} \tilde{A}^\varepsilon(u)\,du\bigg]\Bigg\} \to 0.$$

By using (4.17)–(4.19) and arguments similar to those employed in Sections 2, 3 and 7 of Chapter 5 of Kushner (1984), it can be verified that for any s such that $P(x(\cdot)$ is continuous at $s) = 1$,

$$\tilde{A}^\varepsilon(s) \to Af(x(s)) \tag{4.23}$$

in probability. From these results follow (4.20) and (4.21). □

PROOF OF THEOREM 2. Again, assume $r = 0$. Using Skorohod embedding, we may assume for simplicity that the convergence in Proposition 2 is almost sure convergence. Since $\hat{r}_N = r + O_p(1/N)$, it is readily seen that $N(\hat{r}_N - r)$ converges weakly to M_- where $[M_-, M_+)$ is the unique random interval of all z at which $\tilde{L}^{(1)}(z)I(z<0) + \tilde{L}^{(2)}(z)I(z \ge 0)$ attains its global minimum.

The asymptotic independence between $N(\hat{r}_N - r)$ and $N^{1/2}(\hat{A}_{1N} - A_1, \hat{A}_{2N} - A_2)$ follows readily from (4.11). Indeed, Proposition 1 and formula (4.11) imply that

$$\hat{A}_{iN} = B_{iN}(0) + o_p(N^{-1/2}), \qquad i = 1, 2, \tag{4.24}$$

but $B_{iN}(0)$ is the least squares estimator of A_i when the threshold parameter is known. From this follows the claimed limiting distributions of $\hat{A}$'s. □

Acknowledgments. I would like to thank an Associate Editor for many useful comments and pointing out an error in an earlier draft. I also thank an anonymous referee for many useful comments and suggestions.

REFERENCES

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

CHAN, K. S. (1989). A note on the geometric ergodicity of a Markov chain. *Adv. in Appl. Probab.* **21** 702–704.

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.

CHAN, K. S. and TONG, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. in Appl. Probab.* **17** 666–678.

CHERNOFF, H. and RUBIN, H. (1956). The estimation of the location of a discontinuity in density. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 19–38. Univ. California Press, Berkeley.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–234. Univ. California Press, Berkeley.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

KUSHNER, H. J. (1984). *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory.* MIT Press.

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

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-negative Operators.* Cambridge Univ. Press.

PETRUCCELLI, J. D. and WOOLFORD, S. W. (1984). A threshold *AR*(1) model. *J. Appl. Probab.* **21** 270–286.

PETRUCCELLI, J. D. (1986). On the consistency of least squares estimator of a threshold *AR*(1) model, *J. Time Ser. Anal.* **7** 269–278.

TONG, H. (1983). *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statist.* **21**. Springer, Berlin.

TONG, H. (1987). Non-linear time series models of regularly sampled data: A review. In *Proc. First World Congress of the Bernoulli Society, Tashkent, USSR* **2** 355–367. VNU Science Press, The Netherlands.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242