The Annals of Statistics
1990, Vol. 18, No. 4, 1886–1894

TESTING FOR THRESHOLD AUTOREGRESSION[1]

By K. S. Chan

The University of Chicago

We consider the problem of determining whether a threshold autoregressive model fits a stationary time series significantly better than an autoregressive model does. A test statistic λ which is equivalent to the (conditional) likelihood ratio test statistic when the noise is normally distributed is proposed. Essentially, λ is the normalized reduction in sum of squares due to the piecewise linearity of the autoregressive function. It is shown that, under certain regularity conditions, the asymptotic null distribution of λ is given by a functional of a central Gaussian process, i.e., with zero mean function. Contiguous alternative hypotheses are then considered. The asymptotic distribution of λ under the contiguous alternative is shown to be given by the same functional of a noncentral Gaussian process. These results are then illustrated with a special case of the test, in which case the asymptotic distribution of λ is related to a Brownian bridge.

1. Introduction. A useful class of nonlinear time series model is the threshold autoregressive (TAR) model. For an introduction to the TAR model, see Tong (1983). Tong (1987) surveyed some recent developments of TAR and other nonlinear time series models. An interesting problem is to test if a TAR model provides a significantly better fit to the data than an autoregressive (AR) model does. Petruccelli and Davis (1986) proposed a portmanteau test for threshold autoregressive-type nonlinearity. Their test is a CUSUM-type test based on the predictive residuals of ordered autoregressions. A variant of the test is considered in Tsay (1989). Luukkonen, Saikkonen and Teräsvirta (1988) considered Lagrange multiplier tests for nonliner smooth transition autoregressive models which exclude the TAR model as its autoregressive function is, in general, discontinuous.

In this paper, we consider a test statistic λ which is equivalent to the (conditional) likelihood ratio test statistic when the noise is normally distributed. Essentially, λ is the normalized reduction in sum of squares due to the piecewise linearity of the autoregressive function. The performance of the present test and the portmanteau test of Petruccelli and Davis was studied in Moeanaddin and Tong (1988). Their simulation results seem to suggest that, in general, the present test is more powerful.

The testing problem here is nonstandard since one of the parameters is absent under the null hypothesis. Thus, classical theory of asymptotics is not

Received August 1988; revised October 1989.

[1]Manuscript prepared using computer facilities supported in part by NSF Grants DMS-86-01732 and DMS-87-03942 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

AMS 1980 *subject classifications*. Primary 62M10; secondary 62J05.

Key words and phrases. Asymptotics, autoregressive model, Brownian bridge, contiguity, ergodicity, Gaussian process, least squares, nuisance parameter present only under the alternative hypothesis, ρ-mixing, stationarity, threshold autoregressive model.

applicable. See, for example, Davies (1977, 1987). It is shown in Section 2 that, under certain regularity conditions, the asymptotic null distribution of λ is given by the distribution of a functional of a central Gaussian process, i.e., with zero mean function. Then, in Section 3, we discuss the contiguous alternative hypothesis, under which the asymptotic distribution of λ is shown to be given by the same functional of a noncentral Gaussian process. These results are then illustrated with a special case of the test, in which case the asymptotic distribution of λ is related to a Brownian bridge.

2. The asymptotic null distribution. Let $(X_t)_{t\in Z}$ be a discrete parameter stationary time series. It is assumed that X_t satisfies the difference equation,

$$\begin{aligned} H\colon X_t - \theta_0 - \theta_1 X_{t-1} - \cdots - \theta_p X_{t-p} \\ - I(X_{t-d} \le r)(\phi_0 + \phi_1 X_{t-1} + \cdots + \phi_q X_{t-q}) = \varepsilon_t \end{aligned} \tag{2.1}$$

where (ε_t) is iid with zero mean and finite nonzero variance σ^2 and ε_t independent of the past $X_{t-1}, X_{t-2}, \dots$; I is the indicator function; θ_i's and ϕ_i's are scalars; p, d, q are known nonnegative integers and $p \ge q$; $r \in \tilde{R} \subseteq \mathbb{R}$ and $\tilde{R}$ is a known bounded subset. For simplicity, we first assume that ε_t is Gaussian. In a remark near the end of this section, it is shown that the Gaussian assumption can be relaxed a bit. We also assume the following condition holds:

(C1) All the roots of the characteristic equation, $z^p - \theta_1 z^{p-1} - \cdots - \theta_p = 0$, lie strictly inside the unit circle.

Suppose we observe $(X_0, X_1, \dots, X_N)$. The null hypothesis we want to test is

$$H_0\colon \phi_i = 0, \qquad 0 \le i \le q. \tag{2.2}$$

Under H, X_t is said to follow a threshold autoregressive model. The parameter r is called the threshold parameter and d the delay parameter. For a discussion on the ergodicity and stationarity of TAR models, see, for example, Chan and Tong (1985).

To begin with, suppose the true r is known. Then H is just a linear model. Define

$$\begin{aligned} \hat{\sigma}^2(r) = \min \sum \Bigg[X_t - \theta_0 - \sum_{j=1}^{p} \theta_j X_{t-j} \\ - I(X_{t-d} \le r)\Bigg(\phi_0 + \sum_{j=1}^{q} \phi_j X_{t-j}\Bigg)\Bigg]^2 \Bigg/ n, \end{aligned} \tag{2.3}$$

$$\hat{\sigma}^2 = \min \sum (X_t - \theta_0 - \theta_1 X_{t-1} - \cdots - \theta_p X_{t-p})^2 / n, \tag{2.4}$$

all summations are from l to N unless stated otherwise, $l = \max(p, d)$, $n = N - l + 1$ and the minimum being over all θ's and ϕ's. Then a strictly increasing function of the (conditional) likelihood ratio test statistic for H_0 against $H_1 = H \setminus H_0$ is $\lambda(r) = (RSS_{\text{AR}} - RSS_{\text{TAR}(r)})/\hat{\sigma}^2(r)$, where $RSS_{\text{AR}} = n\hat{\sigma}_0^2$ and $RSS_{\text{TAR}(r)} = \text{in}\hat{\sigma}^2(r)$.

Some conventions and notation, to be adopted throughout, follow. Unless stated otherwise, all expectations are taken under the true probability distribution for which H_0 holds. Let $\mathbf{X}' = (X_l, X_{l+1}, \dots, X_N)$, $\varepsilon' = (\varepsilon_l, \varepsilon_{l+1}, \dots, \varepsilon_N)$ and $\boldsymbol{\theta}' = (\theta_0, \theta_1, \dots, \theta_p)$. Denote by X the $n \times (p+1)$ matrix with its first column consisting of 1's and the $(i, j+1)$th entry equal to $X_{l+i-1-j}$, $j \ge 1$. Abbreviate $I(X_t \le r)$ by $I_r(X_t)$. Y_r stands for the $n \times (q+1)$ matrix whose $(i, 1)$th entry is $I_r(X_{l+i-1-d})$ and whose $(i, j+1)$th entry is $X_{l+i-1-j} I_r(X_{l+i-1-d})$, $j \ge 1$. $\Sigma_{r,k}$ is a $(k+1) \times (k+1)$ symmetric matrix whose $(i+1, j+1)$th entry is

$$(2.5)\quad \Sigma_{r,k}(i+1, j+1) = \begin{cases} E(I_r(X_{t-d})), & j = 0, i = 0, \\ E(X_{t-i} I_r(X_{t-d})), & j = 0, i \ne 0, \\ E(X_{t-i} X_{t-j} I_r(X_{t-d})), & ij \ne 0. \end{cases}$$

Define $\Sigma_r = \Sigma_{r,q}$, Λ_r the upper $(q+1) \times (p+1)$ submatrix of $\Sigma_{r,p}$ and $\Sigma = \Sigma_{\infty,p}$.

Let $Q(r) = RSS_{\text{AR}} - RSS_{\text{TAR}(r)}$. It is just the reduction in sum of squares due to "adding" the variables $I(X_{t-d} \le r)$, $X_{t-1} I(X_{t-d} \le r), \dots,$ $X_{t-q} I(X_{t-d} \le r)$ to the "original" variables $1, X_{t-1}, \dots, X_{t-p}$ in the autoregression. It can be verified that

$$(2.6)\qquad Q(r) = T_r' \left(Y_r'Y_r/n - Y_r'X/n (X'X/n)^{-1} X'Y_r/n \right)^{-1} T_r,$$

where

$$(2.7)\qquad T_r = n^{-1/2} \left(Y_r' - Y_r'X/n (X'X/n)^{-1} X' \right) \mathbf{X}.$$

Note that $T_r = n^{-1/2}(Y_r' - Y_r'X/n(X'X/n)^{-1}X')\varepsilon$ if H_0 holds.

LEMMA 2.1. *Suppose that* (C1) *holds. Let $b > 0$. Then, under H_0, we have the following*:

(i) *Σ is invertible; Σ_r and Λ_r are continuous in r;*

(ii) *$X'X/n \to \Sigma$ a.s.;*

(iii) *$\sup_{-b \le r \le b} |Y_r'X/n - \Lambda_r| = o_p(1)$, where $|\cdot|$ denotes the norm = square root of sum of squares of all entries of the matrix;*

(iv) *for every $r \in R$, $\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda_r'$ is positive definite;*

$$\text{(v)}\qquad \sup_{-b \le r \le b} \left| \left(\frac{Y_r'Y_r}{n} - \frac{Y_r'X}{n} \left(\frac{X'X}{n} \right)^{-1} \frac{X'Y_r}{n} \right)^{-1} - \left(\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda_r' \right)^{-1} \right| = o_p(1);$$

$$\text{(vi)}\qquad \sup_{-b \le r \le b} |T_r - n^{-1/2} Y_r' \varepsilon - n^{-1/2} \Lambda_r \Sigma^{-1} X' \varepsilon| = o_p(1).$$

The proof of the lemma is given in the appendix. In the rest of this section, assume H_0 is true. Let $C = \Lambda_r \Sigma^{-1}$ and write $C = (c_{ij})$. In view of (vi) in

Lemma 2.1, asymptotically,

$$(2.8)\quad T_r = \begin{pmatrix} n^{-1/2}\Sigma\varepsilon_t[I_r(X_{t-d}) - c_{11} - c_{12}X_{t-1} - \cdots - c_{1(p+1)}X_{t-p}] \\ n^{-1/2}\Sigma\varepsilon_t[X_{t-1}I_r(X_{t-d}) - c_{21} - c_{22}X_{t-1} - \cdots - c_{2(p+1)}X_{t-p}] \\ n^{-1/2}\Sigma\varepsilon_t[X_{t-q}I_r(X_{t-d}) - c_{(q+1)1} - \cdots - c_{(q+1)(p+1)}X_{t-p}] \end{pmatrix}.$$

Each component of T_r is a normalized sum of a martingale difference sequence. Employing the Cramer–Wold device and a martingale central limit theorem [see, e.g., Theorem 23.1 in Billingsley (1968)], it is readily seen that T_r converges to $N(0, \sigma^2(\Sigma_r - \Lambda_r\Sigma^{-1}\Lambda'_r))$ weakly. As $\hat{\sigma}^2(r) \to \sigma^2$ in probability, $\lambda(r)$ is asymptotically χ^2 with d.f. $= q + 1$.

In the general case, r is only known to lie inside $\tilde{R}$. Let

$$(2.9)\qquad \hat{\sigma}^2 = \min_{r\in\tilde{R}} \hat{\sigma}^2(r)$$

and

$$(2.10)\qquad \lambda = \sup_{r\in\tilde{R}} Q(r)/\hat{\sigma}^2,$$

which is a monotone increasing function of the LRT statistic. Let $\{\xi_r, -\infty < r < \infty\}$ be a Gaussian process with zero mean function and covariance kernel $K(r,s) = \Sigma_{\min(r,s)} - \Lambda_r\Sigma^{-1}\Lambda'_s$.

Let $b > 0$. Let $D_k(-\infty,\infty)$ ($D_k[-b,b]$) denote the function spaces of all functions, mapping $\mathbb{R}$ ($[-b,b]$) into $\mathbb{R}^k$, that are right continuous and have left-hand limits. Equip $D_k(-\infty,\infty)$ ($D_k[-b,b]$) with the topology of uniform convergence over compact sets. Let $C_k(-\infty,\infty)$ be the subspace of $D_k(-\infty,\infty)$ consisting of functions continuous everywhere. See, for example, Pollard (1984) for more details on these spaces. Now, $\{T_r, -\infty < r < \infty\}$ lives on $D_{q+1}(-\infty,\infty)$.

The argument showing that T_r converges weakly to $\sigma\xi_r$ can be adapted to show that $\{T_r\}$ converges to $\{\sigma\xi_r\}$ in terms of finite dimensional distributions. To derive the asymptotic null distribution of λ, we need the following theorem.

THEOREM 2.2. *Suppose H_0 and condition* (C1) *hold. Also, assume that ε_t is Gaussian. Then $\{T_r\}$ converges weakly to $\{\sigma\xi_r\}$ in $D_{q+1}(-\infty,\infty)$. Furthermore, each realization of $\{\xi_r\}$ belongs to $C_{q+1}(-\infty,\infty)$ a.s.*

SKETCH OF PROOF. Let $b > 0$. It suffices to verify the tightness of $\{T_r, -b \le r \le b\}$ componentwise. Without loss of generality, consider the last component of $\{T_r, -b \le r \le b\}$. It is tight iff $g_n(r) = n^{-1/2}\Sigma\varepsilon_t X_{t-q} I_r(X_{t-d})$ is tight. Under (C1), the Gaussian process (X_t) is ρ-mixing with an exponential decreasing rate, i.e.,

(C2) $\exists\, 0 < \tau < 1$ such that $\rho(m) = O(\tau^m)$, $m \in \mathbb{N}$, where $\rho(m) = \sup|\mathrm{corr}(f,g)|$, the supremum being over all square integrable f and g which are measurable w.r.t. $\{X_t, t \le 0\}$ and $\{X_t, t \ge m\}$, respectively.

For a proof of this result, see Kolmogorov and Rozanov (1960). Let $-b \le s \le r \le b$ be two arbitrary numbers. Let M_0, M_1, M_2, K_1 and K_2 be constants independent of n. Then

$$g_n(r) - g_n(s) = n^{-1/2}\Sigma\varepsilon_t X_{t-q} I(s < X_{t-d} \le r). \tag{2.11}$$

For $i = 1, 2, \dots, q$, $\delta = 1, 2, 3, 4$, we have

$$\begin{aligned} &E\left(|\varepsilon_t X_{t-i} I(s < X_{t-d} \le r)|^{\delta}\right) \\ &\quad \le E\left(1 + |\varepsilon_t|^4\right) E\left(I(s < X_{t-d} \le r) E\left(1 + X_{t-i}^4 | X_{t-d}\right)\right) \end{aligned} \tag{2.12}$$

Now, the conditional distribution of X_{t-i} given X_{t-d} is $N(\mu + \gamma(X_{t-d} - \mu), \sigma_x^2(1 - \gamma^2))$, where $\mu = E(X_t)$, σ_x^2 = variance of X_t and γ is the correlation between X_{t-i} and X_{t-d}. Therefore, $\exists\ M_0$, independent of b, such that $E(1 + X_{t-i}^4 | X_{t-d}) \le M_0(1 + X_{t-d}^4)$. Thence,

$$E\left(|\varepsilon_t X_{t-i} I(s < X_{t-d} \le r)|^{\delta}\right) \le M_1 E\left[\left(1 + |X_1|^4\right) I(s < X_1 \le r)\right], \tag{2.13}$$

and hence

$$E\left(|\varepsilon_t X_{t-i} I(s < X_{t-d} \le r)|^{\delta}\right) \le M_2(r - s), \qquad i = 1, 2, \dots, q,\ \delta = 1, 2, 3, 4. \tag{C3}$$

Let $\zeta_t = n^{-1/2}\varepsilon_t X_{t-q} I(s < X_{t-d} \le r)$. Now by an inequality, for a ρ-mixing process, in Peligrad [(1982), Lemma 3.6], we have

$$\begin{aligned} E\left(|g_n(r) - g_n(s)|^4\right) &\le K_1\left(n^{1/4}\|\zeta_t\|_4 + n^{1/2}\|\zeta_t\|_2\right)^4 \\ &\le K_2\left((r - s)/n + (r - s)^2\right), \end{aligned} \tag{2.14}$$

where $\|\cdot\|_\delta$ denotes the usual L^δ norm. The second line in the preceding inequalities follows from (C3). Let $u > 0$ and $\{-b = r_0 < r_1 < \cdots < r_L = b\}$ be a partition of $[-b, b]$ with $r_j = r_{j-1} + u$, $0 \le j \le L - 1$ and $r_L - r_{L-1} \le u$. Let $\psi_{t,i} = n^{-1/2}|\varepsilon_t X_{t-q}| I(r_{i-1} < x_{t-d} \le r_i)$. Then, $\forall\ i$, for $r_{i-1} \le r \le r_i$,

$$|g_n(r) - g_n(r_i)| \le \Sigma\psi_{t,i}. \tag{2.15}$$

Using Peligrad's inequality and (C3), it can be verified that

$$\sup_i \Sigma\psi_{t,i} = u O_p(n^{1/2}). \tag{2.16}$$

Now, we can adapt the proof of Theorem 22.1 in Billingsley (1968) to show the tightness of $\{g_n(r),\ -b \le r \le b\}$.

THEOREM 2.3. *Suppose that condition* (C1) *holds and* ε_t *is Gaussian. Then*

(i) *under* H_0, $\hat{\sigma}^2 \to \sigma^2$ *in probability*;

(ii) *the asymptotic null distribution of* λ *is given by the distribution of*

$$\sup_{r \in \tilde{R}} \xi_r' \left(\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda_r' \right)^{-1} \xi_r.$$

PROOF. Without loss of generality, assume $\tilde{R} \subseteq [-b, b]$. Define the functional

$$L\colon x(\cdot) \in D_{q+1}[-b, b] \to \sup_{\tilde{R}} x(r)' \left(\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda_r' \right)^{-1} x(r).$$

Now $(\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda_r')^{-1}$ is a continuous matrix function over $[-b, b]$ and $\sup_{\tilde{R}} |x(r)| < \infty$ for each $x(\cdot) \in D[-b, b]$. Thus, L is a continuous functional. Similarly, the functional $\sup_{\tilde{R}} |x(r)|$ is also continuous. It follows from Theorem 2.2 that $\sup_{\tilde{R}} |T_r| = O_p(1)$. Using results in Lemma 2.1, it is easy to see that $\sup_{\tilde{R}} Q(r) = \sup_{\tilde{R}} T_r'(\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda_r')^{-1} T_r + O_p(1)$. Now, $\hat{\sigma}^2 = \hat{\sigma}_0^2 - \sup_{\tilde{R}} Q(r)/n$. Hence, $\hat{\sigma}^2 \to \sigma^2$ in probability and λ converges weakly to $\sup_{\tilde{R}} \xi_r'(\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda_r')^{-1} \xi_r$. □

REMARK. The Gaussian assumption of ε_t in the previous results can be replaced by (C2), (C3) and

(C4) ε_t has zero mean, finite fourth moment and is absolutely continuous with a continuous pdf, say g, which is positive everywhere.

Also, under (C4), it follows from (2.12) that (C3) is equivalent to $\forall\ b > 0\ \exists\ M$ such that $\forall\ -b \le s < r \le b$,

$$E\left(X_{t-i}^4 I(s < X_{t-d} \le r) \right) \le M(r - s), \qquad i = 1, 2, \dots q.$$

As can be readily seen from the proofs, the conclusions of Lemma 2.1, Theorems 2.2 and 2.3, continue to hold if (C1)–(C4) are satisfied.

3. Contiguous alternative hypothesis. As a first step to understanding the asymptotic power of the test proposed in the previous section, we consider local alternatives. For each N, the null hypothesis is

$H_{0N}\colon (X_0, X_1, \dots, X_N)$ follows the model H with $\phi_i = 0,\ 0 \le i \le q$,

versus the alternative hypothesis

$H_{1N}\colon (X_0, X_1, \dots, X_N)$ follows the model H with $\phi_i = \gamma_i n^{-1/2}$, $0 \le i \le q$ and $r = r_0 \in \mathbb{R}$.

Here, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)' \in \mathbb{R}^{q+1}$ is a fixed vector, and r_0 is a fixed scalar. Let P_{0N} and P_{1N} be the probability measure of $(X_0, X_1, \dots, X_N)$ under H_{0N} and H_{1N}, respectively.

THEOREM 3.1. *Suppose that* (C1) *holds and* $g(\cdot)$, *the pdf of* ε_t, *has finite Fisher information, that is,*

$$0 < I(g) = \int \left(\frac{g'(x)}{g(x)} \right)^2 g(x)\, dx < \infty. \tag{C5}$$

Then $\{P_{1N}\}$ *is contiguous to* $\{P_{0N}\}$. *Furthermore, under* H_{0N}, *the log likelihood ratio* Λ_N *is equal to*

$$n^{-1/2} \Sigma I_{r_0}(X_{t-d}) \left(\gamma_0 + \sum_{j=1}^{q} \gamma_j X_{t-j} \right) g'(\varepsilon_t)/g(\varepsilon_t)$$

$$+ I(g) \Sigma I_{r_0}(X_{t-d}) \left(\gamma_0 + \sum_{j=1}^{q} \gamma_j X_{t-j} \right)^2 /(2n) + o_p(1).$$

PROOF. This follows from Theorem 12 in Jeganathan (1988). The only nontrivial condition that needs verification is to show that, under H_{0N}, the sequence of pdf of $(X_0, \dots, X_{l-1})$ under H_{1N} tends to that under H_{0N} in probability. However, this can be done by adapting the proof of Theorem 2.2 in Chan and Tong (1986). □

THEOREM 3.2. *Suppose that* (C1)–(C5) *hold. Then*

(i) *under* H_{1N}, $\{T_r\}$ *converges weakly in* $D_{q+1}(-\infty, \infty)$ *to* $\{\sigma\xi_r + m_r\}$, *where* ξ_r *is as in Theorem* 2.2 *and* $m_r = (\Sigma_{\min(r, r_0)} - \Lambda_r \Sigma^{-1} \Lambda'_{r_0})\gamma$;

(ii) *the asymptotic distribution, under* H_{1N}, *of* $\lambda = \sup_{\tilde{R}} Q(r)/\hat{\sigma}^2$ *is given by*

$$\sup_{\tilde{R}} (\xi_r + \sigma^{-1} m_r)' (\Sigma_r - \Lambda_r \Sigma^{-1} \Lambda'_r)^{-1} (\xi_r + \sigma^{-1} m_r).$$

PROOF. First, the tightness of $\{T_r\}$ under H_{1N} follows from the tightness of $\{T_r\}$ under H_{0N} in view of contiguity. So, it suffices to show the convergence, under H_{1N}, of finite dimensional distribution of $\{T_r\}$ to those of $\{\sigma\xi_r + m_r\}$. This can be done as follows. From Theorem 3.1 and a martingale central limit theorem, under H_{0N}, $(\Sigma_{i=1}^{k} c_i T_{r_i}, \Lambda_n)$ converges weakly to a multivariate normal distribution with $\text{cov}(\Sigma c_i T_{r_i}, \Lambda_n) = \Sigma c_i^2 m_{r_i}$. Then, the required convergence follows by applying Le Cam's third lemma [page 208 in Hájek and Šidák (1967)]. This proves (i). As (ii) follows readily from (i), its proof is omitted. □

It is instructive to examine the results in a special case: $p = q = 0$. Then the covariance kernel $K(r, r')$ of ξ_r is equal to $s(\min(r, r')) - s(r)s(r'))$, where $s(r) = E(I(X_t \le r))$. Let $r(s)$ denote the inverse function of $s(r)$. Hence, $\xi_{r(s)}$ is just a Brownian bridge. Thus, the asymptotic null distribution of λ is given by the distribution of $\sup_{\tilde{S}} B_s^2/(s - s^2)$, where B_s stands for a Brownian bridge and $\tilde{S} \subseteq [0,1]$ is the image of $\tilde{R}$ under the map s. Tabulation for λ in this special case is then possible. For details, see Chan and Tong (1991).

As $m_r = \gamma(s(\min(r, r_0)) - s(r)s(r_0))$, the asymptotic distribution of λ under the contiguous alternative,

$$H_{1N}: X_t - \theta - n^{-1/2}\gamma I(X_{t-d} \le r_0) = \varepsilon_t,$$

is given by the distribution of

$$\sup_{\tilde{S}} \left(B_s + \sigma^{-1}\gamma(\min(s, s_0) - ss_0)\right)^2 / (s - s^2),$$

where $s_0 = s(r_0)$.

APPENDIX

Proof of Lemma 2.1. Part (i) is obvious and part (ii) follows from the ergodicity of (X_t).

To prove part (iii), it suffices to demonstrate the desired uniform convergence entrywise. Consider, the (i, j)th entry of $Y_r'X/n$, $S_n(r) \equiv \Sigma I_r(X_{t-d})X_{t-i}X_{t-j}/n$. From ergodicity, for each r, $S_n(r) \to S(r) \equiv E(I_r(X_{t-d})X_{t-i}X_{t-j})$ a.s. Let $c < d$ be two arbitrary numbers in $[-b, b]$. Condition (C3), as stated in the proof of Theorem 2.2, implies that $\max_{1 \le k \le l} E(X_{t-k}^2 I(c < X_{t-d} \le d)) \to 0$ as $d \to c$. Then, by routine analysis, we can show the uniform convergence of $S_n(r)$ to $S(r)$, over $-b \le r \le b$, in probability.

For part (iv), first suppose $q = p$. Then $\Lambda_r = \Sigma_r$. Since Σ_r and $\Sigma - \Sigma_r$ are positive definite, $\exists$ invertible Q and diagonal D such that $Q\Sigma Q' = I$ and $Q\Sigma_r Q' = D$, and with all diagonal entries of D being strictly between 1 and 0. Hence, $\Sigma_r - \Lambda_r\Sigma^{-1}\Lambda_r'$ is positive definite. The general case can be proved similarly.

Parts (v) and (vi) follows readily from the other parts of the lemma.

Acknowledgment. I am grateful to the two referees for much useful advice, leading to an improvement in the presentation, the elimination of an error in a previous draft and to a strengthened version of Theorem 3.1.

REFERENCES

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time Series Anal.* **7** 179–190.

Chan, K. S. and Tong, H. (1991). On likelihood ratio tests for threshold autoregression. *J. Roy. Statist. Soc. Ser. B* **53**.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254.

DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43.

HÁJEK, J. and ŠIDAK, Z. (1967). *Theory of Rank Tests*. Academic, New York.

JEGANATHAN, P. (1988). Some aspects of asymptotic theory with applications to time series model. Technical Report. Dept. Statist., Univ. Michigan.

KOLMOGOROV, A. N. and ROZANOV, YU. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory Probab. Appl.* **5** 204–207.

LUUKKONEN, R., SAIKKONEN, P. and TERÄSVIRTA, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* **75** 491–499.

MOEANADDIN, R. and TONG, H. (1988). A comparison of likelihood ratio test and CUSUM test for threshold autoregression. *The Statistician* **37** 213–225. (Addendum and corrigendum, **37** 473–474.)

PELIGRAD, M. (1982). Invariance principles for mixing sequences of random variables. *Ann. Probab.* **10** 968–981.

PETRUCCELI, J. D. and DAVIS, N. (1986). A portmanteau test for self exciting threshold autoregressive-type nonlinearity in time series. *Biometrika* **73** 687–694.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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 Scientific Press, The Netherlands.

TSAY, R. S. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.* **84** 231–240.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637