Biometrika (2003), **90**, 3, *pp.* 703–716

Printed in Great Britain

Testing and estimation of thresholds based on wavelets in heteroscedastic threshold autoregressive models

BY WAI-CHEUNG IP, HEUNG WONG

Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

mathipwc@polyu.edu.hk mathwong@polyu.edu.hk

YUAN LI

Department of Mathematics, Guangzhou University, Guangzhou 510405, China

yli57@hotmail.com

AND HONGZHI AN

Institute of Applied Mathematics, The Chinese Academy of Sciences, Beijing 100080, China

ahz@mail.amt.ac.cn

SUMMARY

We consider the testing and estimation of thresholds in heteroscedastic threshold autoregressive models with an unknown number of thresholds. A test statistic based on empirical wavelet coefficients is proposed. The asymptotic distribution of the test statistic is established and consistent estimators of the thresholds and the number of thresholds are given. A Monte Carlo study and a real example are used to assess the performance of our method.

Some key words: Heteroscedasticity; Jump point; Threshold; Time delay; Wavelet.

1. INTRODUCTION

The threshold autoregressive, TAR, model introduced by Tong (1978) is a nonlinear time series model that is able to describe sudden changes over time. The most important parameters of the TAR model are the thresholds since they determine its nonlinear structure. The thresholds divide the model into different regimes, in each of which the skeleton function is linear. Hence, if the thresholds are known, it is relatively easy to estimate other parameters such as regression coefficients. However it is not easy to estimate the thresholds because their number is usually unknown, and they are themselves changepoints of the skeleton function.

Chan & Tong (1986) gave an estimator for the threshold in the case of known time delay with one threshold. In a University of Chicago technical report, K. S. Chan used the conditional least squares method to obtain strongly consistent estimators of the thresholds and of time delay with a known number of thresholds. Tsay (1989) estimated the thresholds and time delay by arranged autoregression. Geweke & Terui (1993) and Chen & Lee (1995) used a Bayesian approach to identify the threshold and time delay.

Recently many statisticians have been interested in wavelets, for example to detect aberrants such as jumps or cusps in regression functions; see for example Donoho et al. (1995), Wang (1995) and an Australian National University technical report by M. H. Neumann. Noting that the thresholds are nothing but the jumps or cusps of the skeleton function in a TAR model, Li & Xie (1999) successfully applied the wavelet method to identify the thresholds and time delay. However, their method has a drawback in practice because distributional results for the test statistics were not obtained. The critical values were constructed by repeated simulations and intuition. In this paper, we first use the kernel method to estimate the wavelet coefficients and then construct the test statistics. Asymptotic distributions of the statistics are derived, and critical values are obtained. Lastly, by Li & Xie's (1999) ρ-division of a set, consistent estimators of the thresholds and the number of thresholds are given.

Section 2 presents a test procedure for thresholds based on wavelets and proposes wavelet estimators for the thresholds and the number of thresholds. Monte Carlo simulations for our test statistic and an empirical example are reported in § 3 and proofs are collected in the Appendix.

2. MAIN RESULTS

2·1. *The* SETAR *model*

The TAR model we address in this paper is the self-exciting threshold autoregressive, SETAR, model with heteroscedasticity:

$$x_t = \sum_{l=1}^{r+1} \left(b_0^{(l)} + \sum_{m=1}^{p_l} b_m^{(l)} x_{t-m} \right) I_{(\lambda_{l-1}, \lambda_l]}(x_{t-d}) + \sigma(x_{t-1}, \ldots, x_{t-p})\varepsilon_t, \tag{2·1}$$

where $\{\varepsilon_t\}$, for $t = 1, 2, \ldots$, are independently identically distributed random variables with mean zero and variance 1, $\lambda_0 = -\infty$ and $\lambda_{r+1} = \infty$. It is assumed that $p_l \leqslant p$, for $l = 1, 2, \ldots, r+1$, where p is a known integer and $d \leqslant p$, and $a < \lambda_1 < \lambda_2 < \ldots < \lambda_r < b$, where a and b are two known constants. We make the following assumptions on $\{x_t\}$ and the noise $\{\varepsilon_t\}$.

Assumption 1. The $\{x_t\}$ are geometrically ergodic.

Assumption 2. The probability density function f of $(x_1, x_2, \ldots, x_p)^{\mathrm{T}}$ is bounded away from zero and infinity on $[a, b]^p$:

$$M^{-1} \leqslant f(x_1, x_2, \ldots, x_p) \leqslant M, \quad (x_1, x_2, \ldots, x_p)^{\mathrm{T}} \in [a, b]^p,$$

where $M > 0$ is a constant. Furthermore, $f(x_1, x_2, \ldots, x_p)$ is continuously differentiable.

Assumption 3. The $\{\varepsilon_t\}$ are independently identically distributed with $E(\varepsilon_t) = 0$ and $\operatorname{var}(\varepsilon_t) = 1$.

Let $b_s^{(l)} = 0$ when $s > p_l$ $(l = 1, 2, \ldots, r+1)$. Then model (2·1) can be written as

$$x_t = \sum_{l=1}^{r+1} \left(b_0^{(l)} + \sum_{m=1}^{p} b_m^{(l)} x_{t-m} \right) I_{(\lambda_{l-1}, \lambda_l]}(x_{t-d}) + \sigma(x_{t-1}, \ldots, x_{t-p})\varepsilon_t \tag{2·2}$$

and is called a self-exciting heteroscedastic threshold autoregressive model with order p, denoted by SETAR$(d, r; p, \ldots, p)$, where $\{\lambda_l\}$ are called thresholds and d is the time delay. If σ is constant, then model (2·2) is the common SETAR model. Here we study the more general SETAR model with heteroscedasticity. Li & Li (1996) studied the TAR model incorporating heteroscedasticity. There have been many discussions of the estimation of the time delay d. Hence, we assume that d is known in the following discussion.

Let

$$T(x_1, x_2, \ldots, x_p) = \sum_{l=1}^{r+1} \left(b_0^{(l)} + \sum_{m=1}^{p} b_m^{(l)} x_m \right) I_{(\lambda_{l-1}, \lambda_l]}(x_{t-d}).$$

Then model (2·2) is written as

$$x_t = T(x_{t-1}, \ldots, x_{t-p}) + \sigma(x_{t-1}, \ldots, x_{t-p})\varepsilon_t.$$

It follows from Li & Xie (1999) that λ_l is a threshold of SETAR$(d, r; p, \ldots, p)$, if and only if either

(i) $x^{(l)} = (x_1, \ldots, x_{d-1}, \lambda_l, x_{d+1}, \ldots, x_p)^{\mathrm{T}}$ is a jump point of T, that is there exists a point $x_0^{(l)} = (t_1, \ldots, t_{d-1}, \lambda_l, t_d, \ldots, t_{p-1})^{\mathrm{T}}$ in R^p such that

$$T(x_0^{(l)} - 0) \neq T(x_0^{(l)} + 0),$$

or

(ii) $x^{(l)}$ is a cusp point of T.

Properties (i) and (ii) imply that there exists at least one s $(1 \leqslant s \leqslant p)$ such that $b_s^{(l)} - b_s^{(l+1)} \neq 0$. Then, for almost all $\tilde{t} \in [a, b]^{p-1}$, it holds that

$$\begin{aligned} &b_0^{(l)} + b_1^{(l)} t_1 + \ldots + b_{d-1}^{(l)} t_{d-1} + b_d^{(l)} \lambda_l + b_{d+1}^{(l)} t_d + \ldots + b_p^{(l)} t_{p-1} \\ &\qquad \neq b_0^{(l+1)} + b_1^{(l+1)} t_1 + \ldots + b_{d-1}^{(l+1)} t_{d-1} + b_d^{(l+1)} \lambda_l + b_{d+1}^{(l+1)} t_d + \ldots + b_p^{(l+1)} t_{p-1}. \end{aligned}$$

We take such fixed $\tilde{t} \in [a, b]^{p-1}$ in the following discussion. Then

$$T(t_1, \ldots, t_{d-1}, x, t_d, \ldots, t_{p-1})$$

is a one-variable function of the variable x. It is easily deduced from (i) and (ii) that, if λ_l is a threshold of SETAR$(d, r; p, \ldots, p)$, it is a jump or cusp point of $T(t_1, \ldots, t_{d-1}, x, t_d, \ldots, t_{p-1})$. Hence, we can use methods developed for the detection of jumps and cusps to identify the thresholds.

2·2. *Empirical wavelet coefficients*

For $T(t_1, \ldots, t_{d-1}, x, t_d, \ldots, t_{p-1})$, with $x \in [a, b]$, its wavelet coefficient is

$$\beta_{j,k} = \int_a^b T(t_1, \ldots, t_{d-1}, x, t_d, \ldots, t_{p-1}) \psi_{j,k}^{\mathrm{per}}(x)\, dx, \tag{2·3}$$

where

$$\psi_{j,k}^{\mathrm{per}}(x) = \sum_n (b-a)^{-\frac{1}{2}} \psi_{j,k}\left(\frac{x-a}{b-a} + n \right)$$

(Li & Xie, 1999), in which $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ and $\psi(x)$, compactly supported on $[-A, A]$ with $A > 1$, is of bounded variation on $[-A, A]$, and $\psi(x) = 0$ for $x \in [-1, 1]$. Furthermore,

$$\int_{-A}^{A} \psi(x)\, dx = 0, \quad \int_{-A}^{A} x\psi(x)\, dx = 0, \quad \int_{-A}^{A} \psi^2(x)\, dx < \infty, \tag{2·4}$$

and $\int_1^A \psi(x)\, dx \neq 0$, $\int_1^A x\psi(x)\, dx \neq 0$.

Simple computation shows that, when $a + k2^{-j}(b-a)$ is near the threshold λ_l, $\beta_{j,k}$ has relatively large absolute values, while it decays fast to zero as soon as $a + k2^{-j}(b-a)$ shifts away from λ_l. Therefore $\beta_{j,k}$ exhibits high peaks near λ_l, based on which we can construct a statistic for testing and estimating the thresholds.

Suppose that x_t $(1 \leqslant t \leqslant n)$ are sampled from model (2·2). Let

$$I(s, \delta) = \left\{k : \left| a + \frac{k}{2^j}(b-a) - s \right| \leqslant \delta \right\}, \quad I_j = \{0, 1, 2, \ldots, 2^j - 1\},$$

$$\tilde{x}_{l-1} = (x_{l-1}, x_{l-2}, \ldots, x_{l-p})^{\mathrm{T}}, \quad \tilde{t}(x) = (t_1, \ldots, t_{d-1}, x, t_d, \ldots, t_{p-1})^{\mathrm{T}}$$

for fixed j. Then the estimator of $\beta_{j,k}$, called the empirical wavelet coefficient, is defined by

$$W_{j,k} = \int_a^b \psi^{\mathrm{per}}_{j,k}(x) \frac{\sum_{l=p+1}^n K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\} x_l}{\sum_{l=p+1}^n K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}} \, dx, \tag{2·5}$$

where $K_h(\tilde{y}) = K(y_1/h)K(y_2/h)\ldots K(y_p/h)$ with $K(.)$ being a kernel function and $\tilde{y} = (y_1, y_2, \ldots, y_p)$. In practice, one often assumes that $K(.)$ is a symmetric probability density with finite support $[-c, c]$, and $h = h_n$ is a sequence of bandwidths with $h \to 0$ and $nh \to \infty$.

The estimator $W_{j,k}$ of $\beta_{j,k}$ is intuitively very appealing. Note that

$$T\{\tilde{t}(x)\} = E\{x_t \mid \tilde{x}_{t-1} = \tilde{t}(x)\},$$

so that $T(.)$ is the conditional mean of x_t, given $\tilde{x}_{t-1}$. Hence, we can use the kernel method to estimate $T\{\tilde{t}(x)\}$ by

$$\hat{T}\{\tilde{t}(x)\} = \frac{\sum_{l=p-1}^n K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\} x_l}{\sum_{l=p+1}^n K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}}.$$

The empirical wavelet coefficient $W_{j,k}$ is obtained if we replace $T\{\tilde{t}(x)\}$ in (2·3) with $\hat{T}\{\tilde{t}(x)\}$.

2·3. *Asymptotic distribution*

THEOREM 1. *Suppose that Assumptions 1–3 hold. If j satisfies*

$$\lim_{n \to \infty} \frac{nh^{p-1}}{2^{3j}} = \infty,$$

then, as n tends to infinity,

(i) *for $k \in I\{\lambda_l, 2^{-j}(b-a)\}$,*

$$\surd(nh^{p-1})(W_{j,k} - \gamma_{j,k}) \to N\{0, \tau^2(\lambda_l)\},$$

in distribution, for $l = 1, 2, \ldots, p$, where

$$\begin{aligned}\gamma_{j,k} = {} & 2^{-j/2}(b-1)^{1/2} \int_1^A \psi(x)\, dx \\ & \times \left[b_0^{(l+1)} - b_0^{(l)} + (b_1^{(l+1)} - b_1^{(l)})t_1 + \ldots + \ldots + (b_{d-1}^{(l-1)} - b_{d-1}^{(l)})t_{(d-1)} \right. \\ & \left. + (b_d^{(l+1)} - b_d^{(l)}) \left\{ a + \frac{k}{2^j}(b-a) \right\} + (b_{d+1}^{(l+1)} - b_{d+1}^{(l)})t_d + \ldots + (b_p^{(l+1)} - b_p^{(l)})t_{p-1} \right] \\ & + 2^{-3j/2}(b-a)^{3/2}(b_d^{(l+1)} - b_d^{(l)}) \int_1^A x\psi(x)\, dx,\end{aligned} \tag{2·6}$$

$$\tau^2(x) = \left\{ \int_{-c}^{c} K^2(y)\, dy \right\}^{p-1} \int_{-A}^{A} \psi^2(x)\, dx \, \frac{\sigma^2\{\tilde{t}(x)\}}{f\{\tilde{t}(x)\}},$$

where $f(.)$ is the joint probability density function of $(x_{t-1}, x_{t-2}, \ldots, x_{t-p})^{\mathrm{T}}$;
(ii) *for $k \notin \cup_{l=1}^{r} I\{\lambda_l, 2^{-j/2}(b-a)\}$,*

$$\surd(nh^{p-1})W_{j,k} \to N\{0, \tau^2(t_z)\},$$

in distribution, where $t_z = a + z(b-a)$, $k/2^j \to z$ for $0 < z < 1$.

We consider testing if λ_l is a threshold; that is we test

$$H_0 : b_i^{(l)} = b_i^{(l-1)} \; (i = 1, 2, \ldots, r+1)$$

versus

$$H_1 : \text{at least one } b_i^{(l)} \neq b_i^{(l+1)}.$$

We know from (2·6) that, under H_0, $\gamma_{j,k} = 0$ for all $k \in I_j$, and, under H_1, $|\gamma_{j,k}| \geqslant c2^{-3j/2}$, where c is a positive constant. Then we construct the statistic

$$T_H^N = \frac{\surd(nh^{p-1})W_{j,k}}{\tau(\lambda_l)}. \tag{2·7}$$

COROLLARY 1. *Assume that the conditions of Theorem 1 hold. Then, under H_0,*

$$T_H^N \to N(0, 1),$$

in distribution, for all $k \in I_j$. Under H_1,

$$T_H^N \to \infty,$$

in probability, for $k \in I\{t_l, 2^{-j}(b-a)\}$.

It follows from Corollary 1 that, if α denotes the size of the test, we reject H_0 if $|T_H^N| > \Phi^{-1}(1-\alpha/2)$, where Φ is the standard normal distribution, and that this is a consistent test.

2·4. *Data-driven test*

The test statistic T_H^N contains the probability density function $f(y_1, y_2, \ldots, y_p)$ of $(x_{t-1}, x_{t-2}, \ldots, x_{t-p})^{\mathrm{T}}$ and the conditional variance $\sigma^2(y_1, y_2, \ldots, y_p)$. In practice, these are usually unknown and we need to estimate them in order to apply the above testing procedure.

The probability density function $f(y_1, y_2, \ldots, y_p)$ can be estimated easily by the kernel method:

$$\hat{f}(y_1, y_2, \ldots, y_p) = \frac{1}{nh^p} \sum_{i=p+1}^{n} K_h(x_{i-1} - y_1, \ldots, x_{i-p} - y_p),$$

which is a consistent estimator of f under mild conditions. Although it is not easy to estimate the conditional variance $\sigma^2(y_1, y_2, \ldots, y_p)$ there has been extensive related discussions, especially for $p = 1$. Here we suggest an estimator based on Truong & Stone (1992, 1994).

Note that

$$\sigma^2(y_1, y_2, \ldots, y_p) = E[\{x_t - T(x_{t-1}, x_{t-2}, \ldots, x_{t-p})\}^2 | x_{t-1} = y_1, \ldots, x_{t-p} = y_p].$$

Then the estimator of σ^2 is given by

$$\hat{\sigma}^2(y_1, y_2, \ldots, y_p) = \frac{\sum_{l=p+1}^{n} K_h(\tilde{x}_{l-1} - \tilde{y})\{\tilde{x}_l - T(\tilde{x}_{l-1})\}^2}{\sum_{l=p+1}^{n} K_h(\tilde{x}_{l-1} - \tilde{y})}, \qquad (2\cdot8)$$

where $K_h(y_1, y_2, \ldots, y_p) = K(y_1/h)K(y_2/h)\ldots K(y_p/h)$, with $K(y) = \frac{1}{2}I_{[-1,1]}(x)$. In (2·8), $T(.)$ is still unknown. We can estimate it by replacing the parameters with their consistent estimators. Under mild conditions, $\hat{\sigma}^2$ is a consistent estimator of σ^2. Hence, we can obtain consistent estimators of both $f(y_1, y_2, \ldots, y_p)$ and $\sigma^2(y_1, y_2, \ldots, y_p)$. Then a consistent estimator of $\tau^2(x)$ is given by

$$\hat{\tau}^2(x) = \left\{\int_{-c}^{c} K^2(y)\,dy\right\}^{p-1} \int_{-A}^{A} \psi^2(x)\,dx\, \frac{\hat{\sigma}^2\{\tilde{t}(x)\}}{\hat{f}\{\tilde{t}(x)\}}.$$

Replacing $\tau^2(x)$ with $\hat{\tau}^2(x)$ in (2·7), we obtain the data-driven test statistic

$$\hat{T}_H^N = \frac{\surd(nh^{p-1})W_{j,k}}{\hat{\tau}(\lambda_l)}.$$

Theorem 2. *Assume that the conditions of Theorem* 1 *hold. Then, under* H_0,

$$\hat{T}_H^N \to N(0, 1),$$

in distribution, for all $k \in I_j$. *Under* H_1,

$$\hat{T}_H^N \to \infty,$$

in probability, for $k \in I\{t_l, 2^{-j}(b-a)\}$.

2·5. *Estimation of thresholds*

Li & Xie (1999) proposed a wavelet method for estimating the thresholds but they did not obtain the asymptotic distributions of the test statistics. Here we derive the asymptotic distribution of the empirical wavelet coefficient, which leads to consistent estimators of the thresholds.

Let $\lambda_k^* = a + k2^{-j}(b-a)$. It follows from Corollary 1 and the proof of Theorem 1 that, for all λ_k^* such that $k \in \bigcup_{l=1}^{r} I\{\lambda_l, 2^{-j}(b-a)\}$, $\surd(nh^{p-1})W_{j,k}/\hat{\tau}(\lambda_k^*)$ is distributed as $N(0, 1)$ asymptotically. Let $c_0 = \Phi^{-1}(1-\alpha/2)$. Under the confidence level $1-\alpha$, we consider the set

$$E(j) = \{k : |W_{j,k}| \geqslant c_0\tau(\lambda_k^*)(nh^{p-1})^{-1/2}, k \in I_j\}.$$

This $E(j)$ can be divided into q parts by ρ-division (Li & Xie, 1999) with $\rho = 2^{j/2+1}(b-a)$:

$$E(j) = \bigcup_{l=1}^{q} E_l(j).$$

Let

$$\hat{r}_n = \begin{cases} q, & \text{if } E(j) \text{ is not empty}, \\ 0, & \text{if } E(j) \text{ is empty}, \end{cases} \qquad \hat{\lambda}_l = a + \frac{k_l}{2^j}(b-a),$$

where k_l satisfies

$$|W_{j,k_l}| = \max_{k \in E_l(j)} |W_{j,k}| \quad (l = 1, 2, \ldots, \hat{r}).$$

The following theorem is similar to Theorem 3.3 of Li & Xie (1999).

THEOREM 3. *Under the conditions of Theorem* 1, *when* $n \to \infty$, *we have the following*:
(i) $\lim_{n \to \infty} \mathrm{pr}(\hat{r}_n = r) = 1$;
(ii) *for* $r > 0$, $\hat{\lambda}_l \to \lambda_l$, *in probability, for* $l = 1, 2, \ldots, r$.

3. NUMERICAL RESULTS

3·1. *Monte Carlo results*

Throughout this section, all Monte Carlo models are simulated 500 times and the wavelet $\psi(x)$ takes the following form:

$$\psi(x) = \begin{cases} 5(x-1)^4, & \text{if } 1 \leqslant x \leqslant 2, \\ \frac{20}{3}(x+1)^3 + 2(x+1)^2, & \text{if } -2 \leqslant x \leqslant -1, \\ 0, & \text{otherwise}, \end{cases} \tag{3·1}$$

as in Li & Xie (1999). The kernel function $K(x)$ used is the Gaussian kernel:

$$K(x) = \frac{1}{\sqrt{(2\pi)}} \exp(-x^2/2). \tag{3·2}$$

The resolution level is $j = 6$. Five hundred pairs of samples of sizes $n = 50$, 100 and 200 are generated respectively in the following examples.

Table 1. *Empirical size* (%) *of the test for model* (3·3), *based on* 500 *simulations*

		$\lambda = 0·35$		$\lambda = 0·5$	
σ	n	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$
0·2	50	6·6	15·4	6·8	9·4
	100	6·0	15·2	5·6	9·8
	200	5·4	12·0	5·4	10·4
0·4	50	5·4	12·4	4·0	10·6
	100	5·0	13·8	3·2	9·4
	200	5·0	9·4	5·0	9·6
0·6	50	5·6	9·8	6·4	9·6
	100	5·0	12·0	6·0	11·2
	200	4·8	8·2	5·2	10·6

Table 1 shows the empirical sizes under the null hypothesis H_0 that

$$x_t = 0·6x_{t-1} + \sigma\varepsilon_t, \tag{3·3}$$

with σ taking three different values 0·2, 0·4 and 0·6, where the ε_t are independently identically distributed as $N(0, 1)$. We see from Table 1 that, with increasing sample size n, the size of the test gets closer to the nominal value. Tables 2–4 present the powers of the test under the alternative hypothesis respectively for the following SETAR(1, 2; 1, 1, 1), SETAR(2, 2; 2, 2, 2) and SETAR(8, 2; 12, 12, 12) models:

$$x_t = \mu_1(x_{t-1}) + \sigma\varepsilon_t, \tag{3·4}$$

where

$$\mu_1(x_{t-1}) = \begin{cases} 0\cdot6x_{t-1}, & \text{if } x_{t-1} \leqslant 0\cdot35, \\ -4\cdot2x_{t-1}, & \text{if } 0\cdot35 < x_{t-1} \leqslant 0\cdot5, \\ 0\cdot8x_{t-1}, & \text{if } 0\cdot5 < x_{t-1}; \end{cases}$$

$$x_t = \mu_2(x_{t-1}, x_{t-2}) + \sigma\varepsilon_t, \tag{3·5}$$

where

$$\mu_2(x_{t-1}, x_{t-2}) = \begin{cases} 0\cdot4x_{t-1} + 0\cdot26x_{t-2}, & \text{if } x_{t-2} \leqslant 0\cdot35, \\ 0\cdot2x_{t-1} - 4\cdot2x_{t-2}, & \text{if } 0\cdot35 < x_{t-2} \leqslant 0\cdot5, \\ 0\cdot3x_{t-1} + 0\cdot6x_{t-2}, & \text{if } 0\cdot5 < x_{t-2}; \end{cases}$$

$$x_t = \mu_3(x_{t-1}, x_{t-8}, x_{t-12}) + \sigma\varepsilon_t, \tag{3·6}$$

where

$$\mu_3(x_{t-1}, x_{t-8}, x_{t-12}) = \begin{cases} 0\cdot4x_{t-1} + 0\cdot26x_{t-8} - 0\cdot15x_{t-12}, & \text{if } x_{t-8} \leqslant 0\cdot35, \\ 0\cdot2x_{t-1} - 4\cdot2x_{t-8} + 0\cdot2x_{t-12}, & \text{if } 0\cdot35 < x_{t-8} \leqslant 0\cdot5, \\ 0\cdot3x_{t-1} + 0\cdot2x_{t-8} + 0\cdot25x_{t-12}, & \text{if } 0\cdot5 < x_{t-8}. \end{cases}$$

Model (3·5) is used in Li & Xie (1999) and (3·6) is a higher-order model. It can be seen from Tables 2 and 3 that the test becomes more powerful as the sample size increases. On the other hand, as σ increases the threshold estimate becomes less satisfactory. This may be attributed to the contamination of the true signals by the noise as σ increases. Similar results are also found in Li & Xie (1999). Furthermore, by comparing Table 3 with Table 2, we see that the powers of the test for model (3·5) appear less satisfactory than those for model (3·4). This is because model (3·5) is a second-order TAR model and the dimension affects the empirical wavelet coefficients (2·5). This is the so-called curse of dimensionality. To assess the influence of dimensionality on our method, a SETAR model (3·6) of order 12 is included in our simulation. The results in Table 4 show some trace of the curse of dimensionality, but the problem does not appear to be serious. Nonetheless, the method is seen to provide satisfactory threshold estimates even for a model of an order as high as $p = 12$.

Table 2. *Empirical power (%) of the test for model* (3·4), *based on* 500 *simulations*

		$\alpha = 5\%$				$\alpha = 10\%$			
		$\lambda_1 = 0\cdot35$		$\lambda_2 = 0\cdot5$		$\lambda_1 = 0\cdot35$		$\lambda_2 = 0\cdot5$	
σ	n	Power	$\hat{\lambda}_1$	Power	$\hat{\lambda}_2$	Power	$\hat{\lambda}_1$	Power	$\hat{\lambda}_2$
0·2	50	81·0	0·34 (0·06)	40·8	0·53 (0·08)	87·2	0·33 (0·06)	46·0	0·53 (0·08)
	100	94·6	0·35 (0·03)	68·6	0·51 (0·06)	96·2	0·35 (0·04)	75·0	0·52 (0·06)
	200	99·2	0·36 (0·01)	93·2	0·51 (0·06)	99·8	0·36 (0·01)	96·6	0·51 (0·06)
0·4	50	82·2	0·34 (0·05)	84·6	0·58 (0·14)	87·2	0·33 (0·06)	87·2	0·58 (0·14)
	100	92·6	0·35 (0·04)	98·0	0·57 (0·14)	95·6	0·36 (0·04)	99·4	0·57 (0·15)
	200	99·8	0·36 (0·02)	99·6	0·57 (0·15)	99·6	0·36 (0·02)	100·0	0·57 (0·15)
0·6	50	73·6	0·32 (0·08)	85·6	0·61 (0·15)	81·0	0·31 (0·08)	89·4	0·61 (0·15)
	100	92·4	0·34 (0·06)	98·0	0·61 (0·16)	94·8	0·34 (0·06)	98·6	0·61 (0·16)
	200	97·6	0·36 (0·02)	99·8	0·61 (0·16)	99·4	0·35 (0·03)	100·0	0·61 (0·16)

$\hat{\lambda}_i$ is the mean estimated threshold in 500 simulations; numbers in parentheses are standard errors.

Table 3. *Empirical power (%) of the test for model (3·5), based on 500 simulations*

		$\alpha = 5\%$				$\alpha = 10\%$			
		$\lambda_1 = 0{\cdot}35$		$\lambda_2 = 0{\cdot}5$		$\lambda_1 = 0{\cdot}35$		$\lambda_2 = 0{\cdot}5$	
σ	n	Power	$\hat{\lambda}_1$	Power	$\hat{\lambda}_2$	Power	$\hat{\lambda}_1$	Power	$\hat{\lambda}_2$
0·2	50	85·2	0·34 (0·09)	39·6	0·51 (0·05)	87·8	0·34 (0·09)	46·4	0·51 (0·05)
	100	96·2	0·37 (0·06)	76·0	0·51 (0·05)	96·8	0·37 (0·06)	81·0	0·51 (0·05)
	200	99·4	0·39 (0·03)	94·4	0·50 (0·04)	99·4	0·39 (0·03)	95·4	0·50 (0·04)
0·4	50	67·4	0·33 (0·10)	56·0	0·58 (0·12)	71·6	0·33 (0·10)	61·2	0·58 (0·12)
	100	88·4	0·36 (0·07)	81·8	0·58 (0·13)	91·6	0·36 (0·08)	86·6	0·58 (0·13)
	200	98·2	0·39 (0·04)	98·2	0·56 (0·13)	99·2	0·39 (0·04)	99·0	0·56 (0·13)
0·6	50	46·0	0·32 (0·10)	47·2	0·62 (0·15)	51·6	0·32 (0·10)	51·0	0·62 (0·15)
	100	73·0	0·34 (0·08)	73·2	0·62 (0·15)	79·2	0·34 (0·09)	79·8	0·63 (0·15)
	200	93·4	0·35 (0·08)	92·6	0·59 (0·14)	95·0	0·35 (0·08)	94·8	0·59 (0·14)

$\hat{\lambda}_i$ is the mean estimated threshold in 500 simulations; numbers in parentheses are standard errors.

Table 4. *Empirical power (%) of the test for model (3·6), based on 500 simulations*

		$\alpha = 5\%$				$\alpha = 10\%$			
		$\lambda_1 = 0{\cdot}35$		$\lambda_2 = 0{\cdot}5$		$\lambda_1 = 0{\cdot}35$		$\lambda_2 = 0{\cdot}5$	
σ	n	Power	$\hat{\lambda}_1$	Power	$\hat{\lambda}_2$	Power	$\hat{\lambda}_1$	Power	$\hat{\lambda}_2$
0·2	50	92·4	0·33 (0·08)	36·2	0·52 (0·03)	94·6	0·33 (0·08)	41·6	0·52 (0·03)
	100	98·8	0·36 (0·03)	67·0	0·52 (0·02)	99·0	0·36 (0·03)	74·4	0·52 (0·02)
	200	99·4	0·36 (0·01)	95·4	0·52 (0·01)	99·4	0·36 (0·01)	97·4	0·52 (0·01)
0·4	50	60·0	0·35 (0·05)	30·6	0·55 (0·08)	67·2	0·35 (0·06)	36·2	0·55 (0·08)
	100	85·8	0·36 (0·02)	61·2	0·53 (0·03)	89·2	0·36 (0·02)	69·0	0·53 (0·04)
	200	99·2	0·36 (0·01)	92·4	0·52 (0·02)	99·8	0·36 (0·02)	95·4	0·53 (0·02)
0·6	50	29·2	0·33 (0·07)	18·2	0·58 (0·11)	37·0	0·34 (0·07)	25·0	0·58 (0·11)
	100	51·0	0·36 (0·04)	33·6	0·56 (0·10)	61·0	0·36 (0·04)	42·0	0·56 (0·10)
	200	80·0	0·36 (0·02)	64·2	0·54 (0·07)	86·2	0·36 (0·02)	74·8	0·54 (0·07)

$\hat{\lambda}_i$ is the mean estimated threshold in 500 simulations; numbers in parentheses are standard errors.

We have compared our method with the arranged autoregression of Tsay (1989). The model used in Tsay (1989) is given by

$$y_t = \phi_0^{(j)} + \sum_{i=1}^{p} \phi_i^{(j)} y_{t-i} + a_t^{(j)}, \quad r_{j-1} \leqslant y_{t-d} < r_j,$$

where $j = 1, 2, \ldots, k$, d is a positive integer and $\{a_t^{(j)}\}$ is a sequence of martingale differences. The thresholds are

$$-\infty = r_0 < r_1 < \ldots < r_k = \infty.$$

The F- and Portmanteau tests were used in Tsay (1989). In our simulation $p = d = 1$, $\phi_1^{(2)} = 0{\cdot}5$ and two sets of the other ϕ values are used. The simulated results in Table 5 show that our method is generally more capable of capturing the correct size than is Tsay's arranged autoregression. In terms of the power, our method and that of Tsay's are comparable.

3·2. *A real example*

As an empirical illustration, we apply the wavelet-based test to the Sunspots data. Using their wavelet approach, Li & Xie (1999) came to the same conclusion as Tong (1983,

Table 5. *Comparison of empirical powers (%) for wavelet versus arranged autoregression models, based on 500 simulations*

	$n=100$						$n=50$					
	$\alpha=1\%$			$\alpha=5\%$			$\alpha=1\%$			$\alpha=5\%$		
$\phi_1^{(2)}$	W	F	P	W	F	P	W	F	P	W	F	P
				$\phi_0^{(1)}=1,\ \phi_1^{(1)}=0{\cdot}5,\ \phi_0^{(2)}=1,\ r_1=1,\ \sigma_1^2=1,\ \sigma_2^2=1$								
−2·0	100·0	100·0	100·0	100·0	100·0	100·0	100·0	100·0	90·7	100·0	100·0	97·7
−1·0	96·2	100·0	92·8	99·1	100·0	97·8	88·7	91·7	58·4	95·9	98·8	78·4
−0·5	86·6	83·6	58·9	95·0	94·3	79·0	43·0	41·6	15·4	70·6	67·2	47·8
0	10·7	8·7	12·1	35·0	23·7	25·0	7·0	3·5	7·4	29·8	12·0	17·0
0·5	1·0	0·6	1·0	5·7	2·6	3·9	1·4	0·9	2·0	6·9	3·8	5·7
				$\phi_0^{(1)}=0,\ \phi_1^{(1)}=0{\cdot}5,\ \phi_0^{(2)}=0,\ r_1=0,\ \sigma_1^2=1,\ \sigma_2^2=1$								
−2·0	79·1	74·0	94·5	82·2	85·6	99·5	75·4	41·8	45·6	77·3	61·2	76·0
−1·0	66·6	67·2	53·9	78·5	80·4	82·0	49·1	35·7	16·7	66·6	53·9	41·4
−0·5	46·9	53·5	24·6	69·6	70·1	51·9	31·0	25·1	7·1	57·5	42·6	23·2
0	22·9	22·9	5·6	51·9	37·4	19·5	7·1	9·3	2·7	26·8	20·7	9·5
0·5	1·0	1·9	2·3	5·9	5·9	7·2	0·6	1·8	1·6	5·2	5·9	7·8

W, wavelet; F, F-statistic; P, Portmanteau test.

p. 320) that there was only one threshold between 1749 and 1924. They reached their conclusion by observing high peaks on the wavelet graph, which is to a certain extent subjective. Here we apply our wavelet testing method. We make the transformation $x_t = (w_t^{1/2} - 1)/10$, as did Li & Xie (1999). From Tong (1983) and Li & Xie (1999) we know that the time delay can be taken as $d = 8$, and $n = 176$. With $a = 0{\cdot}25$, $b = 1{\cdot}2$, $p = 11$, $\tilde{z} = (0{\cdot}5, 0{\cdot}5, 0{\cdot}6, 0{\cdot}7, 0{\cdot}55, 0{\cdot}4, 0{\cdot}45, 0{\cdot}65, 0{\cdot}75, 0{\cdot}5)^{\mathrm{T}}$ (Li & Xie, 1999), and the wavelet and kernel functions respectively given in (3·1) and (3·2), we calculate $W_{j,k}$, for $k \in I_j$ at $j = 7$, and find that only at $k = 63$, where $\surd(nh^{p-1})|W_{j,k}|/\tau(\lambda_k^*) = 2{\cdot}00$, do we obtain a result that is significant at the 5% level. Therefore the threshold is estimated to be 14·4, which is close to that of Tong (1983, p. 230) and Li & Xie (1999).

Acknowledgement

This research was supported by a research grant from The Hong Kong Polytechnic University. Yuan Li's research was also supported by The National Natural Scientific Foundation of China. We thank the editor and two referees whose comments have led to valuable improvements in the manuscript.

Appendix

Technical details

LEMMA A1. *Suppose that Assumption 1 is satisfied. Then, for sufficiently large B, as $n \to \infty$,*

$$\mathrm{pr}\left\{ \sup_{a \leqslant \tilde{x} \leqslant b} \left| \frac{1}{nh^p} \sum_{t=p+1}^{n} K_h(\tilde{x}_{t-1} - \tilde{x}) - f(\tilde{x}) \right| > B(nh_n)^{-p/2} \log^{\frac{1}{2}} n \right\} \to 0.$$

See Masry (1996) for the proof.

LEMMA A2 (*Volkonskii & Pozanov*, 1959). *Let $V_1, \ldots, V_m$ be random variables measurable with respect to the σ-algebras $\mathcal{F}_{i_1}^{j_1}, \ldots, \mathcal{F}_{i_m}^{j_m}$ respectively with $1 \leqslant i_1 < j_1 < \ldots < j_m \leqslant n$, $i_{l+1} - j_l \geqslant w \geqslant 1$*

and $|V_j| \leqslant 1$, *for* $j = 1, 2, \ldots, m$. *Then*

$$\left| E\left(\prod_{j=1}^{m} V_j\right) - \sum_{j=1}^{m} EV_j \right| \leqslant 16(m-1)\alpha(w),$$

where $\mathscr{F}_a^b = \mathscr{F}\{V_i, a < i \leqslant b\}$ *denotes the* σ*-field generated by* $V_{a+1}, V_{a+2}, \ldots, V_b$.

Proof of Theorem 1. Now $W_{j,k}$ can be written as

$$W_{j,k} = \int_a^b \psi_{j,k}^{\text{per}}(x) T\{\tilde{t}(x)\}\, dx + \int_a^b \psi_{j,k}^{\text{per}}(x) \frac{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}[T(\tilde{x}_{l-1}) - T\{\tilde{t}(x)\}]}{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}}\, dx$$

$$+ \int_a^b \psi_{j,k}^{\text{per}}(x) \frac{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}\sigma(\tilde{x}_{l-1})\varepsilon_l}{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}}\, dx$$

$$= \beta_{j,k} + W_{j,k}^{(R)} + W_{j,k}^{(\varepsilon)}, \tag{A·1}$$

where

$$\beta_{j,k} = \int_a^b \psi_{j,k}^{\text{per}}(x) T\{\tilde{t}(x)\}\, dx,$$

$$W_{j,k}^{(R)} = \int_a^b \psi_{j,k}^{\text{per}}(x) \frac{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}[T(\tilde{x}_{l-1}) - T\{\tilde{t}(x)\}]}{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}}\, dx,$$

$$W_{j,k}^{(\varepsilon)} = \int_a^b \psi_{j,k}^{\text{per}}(x) \frac{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}\sigma(\tilde{x}_{l-1})\varepsilon_l}{\sum_{l=p+1}^{n} K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}}\, dx.$$

Note that

$$\int_{-A}^{A} \psi(x)\, dx = 0, \quad \int_{-A}^{A} x\psi(x)\, dx = 0.$$

Then simple computations show that, for all $k \in I\{\lambda_l, 2^{-j}(b-a)\}$,

$$\beta_{j,k} = 2^{-j/2}\sqrt{(b-a)} \int_1^A \psi(x)\, dx \Big[b_0^{(l+1)} - b_0^{(l)} + (b_1^{(l+1)} - b_1^{(l)})t_1 + \ldots + (b_{d-1}^{(l+1)} - b_{d-1}^{(l)})t_{d-1}$$

$$+ (b_d^{(l+1)} - b_d^{(l)})\left\{a + \frac{k}{2^j}(b-a)\right\} + (b_{d+1}^{(l+1)} - b_{d+1}^{(l)})t_d$$

$$+ \ldots + (b_p^{(l+1)} - b_p^{(l)})t_{p-1}\Big]$$

$$+ 2^{-3j/2}(b-a)^{3/2}(b_d^{(l+1)} - b_d^{(l)}) \int_1^A x\psi(x)\, dx; \tag{A·2}$$

for $k \notin \cup_{l=1}^{r} I\{\lambda_l, 2^{-j/2}(b-a)\}$,

$$\beta_{j,k} = 0. \tag{A·3}$$

As in Lemma A.2 of a Hong Kong Polytechnic University technical report by G. M. Chen and Y. Zhou, it can be shown that

$$W_{j,k}^{(R)} = O_p\left(h 2^{-j/2} \frac{\log n}{(nh)^{p/2}}\right). \tag{A·4}$$

It follows from Lemma A1 that

$$W_{j,k}^{(\varepsilon)} = W_{j,k}^{(\varepsilon)}(M) + O_p\left(h 2^{-j/2} \frac{\log n}{(nh)^{p/2}}\right), \tag{A·5}$$

where

$$W^{(\varepsilon)}_{j,k}(M) = \frac{1}{nh^p}\sum_{l=p+1}^{n} Z_{n,l}, \quad Z_{n,l} = \Gamma_{n,l}\sigma(\tilde{x}_{l-1})\varepsilon_l, \quad \Gamma_{n,l} = \int_a^b \psi^{\text{per}}_{j,k}(x)\,\frac{K_h\{\tilde{x}_{l-1} - \tilde{t}(x)\}}{f\{\tilde{t}(x)\}}\,dx.$$

Obviously $E(Z_{n,l}) = 0$ and $\operatorname{var}(Z_{n,l}) = E(Z^2_{n,l}) = E\{\Gamma^2_{n,l}\sigma^2(\tilde{x}_{l-1})\}$. Trivial computations yield the variance of $Z_{n,l}$ as

$$\operatorname{var}(Z_{n,l}) = h^{p+1}\tau^2(\lambda_l) + o(h^{p-1}),$$

where

$$\tau^2(\lambda_l) = \left\{\int_{-c}^{c} K^2(y)\,dy\right\}^{p-1}\int_{-A}^{A}\psi^2(x)\,dx\,\frac{\sigma^2\{\tilde{t}(\lambda_l)\}}{f\{\tilde{t}(\lambda_l)\}}.$$

Hence

$$\frac{\sqrt{(nh^{p-1})}W^{(\varepsilon)}_{j,k}(M)}{\tau(\lambda_l)} = \frac{1}{\sqrt{n}}\sum_{l=p+1}^{n} V_{n,l}, \tag{A·6}$$

where $V_{n,l} = Z_{n,l}/h^{(p+1)/2}\tau(\lambda_l)$. Obviously, $E(V_{n,l}) = 0$ and $\operatorname{var}(V_{n,l}) = 1 + o(1)$.

In what follows, we shall show that

$$\frac{1}{\sqrt{n}}\sum_{l=p+1}^{n} V_{n,l} \to N(0, 1), \tag{A·7}$$

in distribution. Let $S_n = \sum^{n}_{l=p+1} V_{n,l}$. For convenience, let the subscript l of $V_{n,l}$ start from 1. Partition the set $\{1, 2, \ldots, n\}$ into $2q_n + 1$ subsets with large block size $u = u_n$ and small block size $v = v_n$. Let

$$q = q_n = \left[\frac{n}{u_n + v_n}\right].$$

We define the random variables

$$\zeta_l = \sum_{i=l(u+v)+1}^{l(u+v)+u} V_{n,i}, \quad \xi_l = \sum_{i=l(u+v)+u+1}^{(l+1)(u+v)} V_{n,i} \quad (0 \leqslant l \leqslant q-1), \quad \zeta_q = \sum_{i=q(u+v)+1}^{n} V_{n,i}.$$

Then

$$S_n = \sum_{l=0}^{q-1}\zeta_l + \sum_{l=0}^{q-1}\xi_l + \zeta_q = S_{n_1} + S_{n_2} + S_{n_3}.$$

According to Theorem 18.4.1 of Ibragimov & Linnik (1971), for (A·7) to hold, it is sufficient to prove that

$$\frac{1}{n}E(S^2_{n_2}) \to 0, \quad \frac{1}{n}E(S^2_{n_3}) \to 0, \tag{A·8}$$

$$\left|E\{\exp(itS_{n_1})\} - \prod_{l=0}^{q-1}E\{\exp(it\zeta_l)\}\right| \to 0, \tag{A·9}$$

$$\frac{1}{n}\sum_{l=0}^{q-1}E(\zeta_l^2) \to 1, \tag{A·10}$$

$$\frac{1}{n}\sum_{l=0}^{q-1}E\{\zeta_l^2 I(|\zeta_l| > \varepsilon\sqrt{n}\} \to 0. \tag{A·11}$$

Now we shall show that (A·8)–(A·11) hold. We first choose the block size. Take the large block

size u_n and the

where $\alpha(.)$ is th
Note that $E(V_n$

Thus (A·8) hol

which tends to
It is straight

for $v_n/u_n \to 0$, le
Finally, it re

for some positi

as $n \to \infty$ and
(1971), and (2·
Corollary 1
consistent estin
is similar to th

Chan, K. S. &
7, 179–90.
Chen, C. W. S. &
16, 483–92.
Donoho, D. L.,
(with Discussi
Geweke, J. & T
Ser. Anal. 14,
Ibragimov, I. A
Groningen: W
Li, W. C. & Li, V
Economet. 11,
Li, Y. & Xie, Z
models. Statis
Masry, E. (1996
and rates. J. T

size u_n and the small block size v_n to satisfy

$$\frac{v_n}{u_n} \to 0, \quad \frac{u_n}{n} \to 0, \quad \frac{u_n\sqrt{\log n}}{(n2^{-j})^{\frac{1}{2}}} \to 0,$$

$$\frac{n}{u_n}\,\alpha(v_n) \to 0, \tag{A·12}$$

where $\alpha(.)$ is the α-mixing coefficient of $\{x_t\}$, which is ensured by the geometric ergodicity of $\{x_t\}$. Note that $E(V_{n,i}V_{n,l}) = 0$ for $i \neq l$. Then straightforward computations show that

$$\frac{1}{n}E(S_{n_2}^2) = \frac{qv}{n} \to 0, \quad \frac{1}{n}E(S_{n_3}^2) = \frac{1}{n}\{n - q(u+v)\} \to 0.$$

Thus (A·8) holds. By Lemma A2 with $V_l = \exp(it\zeta_l)$ and $i_{l+1} - j_l = v_n + 1$, we have that

$$\left| E\left[\prod_{i=0}^{q-1}\exp(it\zeta_i) - \prod_{i=0}^{q-1}E\{\exp(it\zeta_i)\}\right]\right| \leqslant 16q_n\alpha(v_n+1),$$

which tends to zero by (A·12). Hence (A·9) holds.

It is straightforward to show that $\operatorname{var}(\zeta_l) = \operatorname{var}(\zeta_0) = u_n\{1 + o(1)\}$, which implies that

$$\frac{1}{n}\sum_{i=0}^{q-1}E(\zeta_i^2) = \frac{q_n u_n}{n}\{1 + o(1)\} \to 1$$

for $v_n/u_n \to 0$, leading to (A·10).

Finally, it remains to establish (A·11). Since

$$\max_{0\leqslant i\leqslant q-1}|\zeta_i| \leqslant C2^{-j/2}u_n\sqrt{\log n}$$

for some positive constant C, which implies that the set $\{|\zeta_i| \geqslant \varepsilon\sqrt{n}\}$ is empty, therefore

$$\max_{1\leqslant i\leqslant q-1} E\{\zeta_i^2 I(|\zeta_i| > \varepsilon\sqrt{n})\} \to 0$$

as $n \to \infty$ and (A·11) follows. Expression (A·7) holds by Theorem 18.4.1 of Ibragimov & Linnik (1971), and (2·4) and (2·6) follow from (A·1), (A·2)–(A·5) and (A·6).

Corollary 1 follows directly from Theorem 1, and thus we omit its proof. Note that $\hat{\tau}(\lambda_l)$ is a consistent estimator of $\tau(\lambda_l)$. Theorem 2 is the direct result of Theorem 1. The proof of Theorem 3 is similar to that of Theorem 3.3 of Li & Xie (1999).

References

Chan, K. S. & Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time Ser. Anal.* **7**, 179–90.

Chen, C. W. S. & Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *J. Time Ser. Anal.* **16**, 483–92.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: asymptopia? (with Discussion). *J. R. Statist. Soc.* B **57**, 301–69.

Geweke, J. & Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *J. Time Ser. Anal.* **14**, 441–54.

Ibragimov, I. A. & Linnik, Y. A. (1971). *Independent and Stationary Sequences of Random Variables.* Groningen: Wolters-Noordohoff.

Li, W. C. & Li, W. K. (1996). On a double threshold autoregressive heteroscedastic time series model. *J. Appl. Economet.* **11**, 253–74.

Li, Y. & Xie, Z. (1999). The wavelet identification of thresholds and time delay of threshold autoregressive models. *Statist. Sinica* **9**, 153–66.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform and strong consistency and rates. *J. Time Ser. Anal.* **17**, 571–99.

TONG, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, NATO ASI Series E: Applied SC. No. 29, Ed. C. H. Chen, pp. 575–86. Groningen: Sijthoff & Noordhoff.

TONG, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*, Springer Lecture Notes in Statistics, 21. New York: Springer.

TRUONG, Y. K. & STONE, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20**, 77–97.

TRUONG, Y. K. & STONE, C. J. (1994). Semiparametric time series regression. *J. Time Ser. Anal.* **15**, 405–28.

TSAY, R. S. (1989). Testing and modelling threshold autoregressive processes. *J. Am. Statist. Assoc.* **84**, 231–40.

VOLKONSKII, V. A. & ROZANOV, Y. A. (1959). Some limit theorems for random functions. *Theory Prob. Applic.* **4**, 178–97.

WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385–97.

[*Received September* 2001. *Revised October* 2002]

Biometrika (2003),
© 2003 Biometrika
Printed in Great Br

Estin

Department

We use the s
its related quar
distribution for
resampling tec
the correspond
The new propo

Some key words
hazards model.

Relatively lit
portional haza
survival functi
Breslow cumul
estimates for t
Andersen, 1986
dent increment
prediction pro
Lin et al. (199
bands for such
The accelera
covariates, has
prediction pro
used a so-called
survival proba
on all the unk
It is difficult
Moreover, the
specific surviv
Recently, J
coefficients of
variance estim
procedure wit