Statistica Sinica **9**(1999), 153-166

THE WAVELET IDENTIFICATION OF THRESHOLDS AND TIME DELAY OF THRESHOLD AUTOREGRESSIVE MODELS

Yuan Li and Zhongjie Xie

University of Petroleum and Peking University

Abstract: In this paper, we consider identification of the thresholds and time delay of threshold autoregressive models with $p-$ dependence and an unknown number of thresholds. By checking p different empirical wavelets of the data to see which have significantly large absolute values, the time delay is identified first. By further checking the empirical wavelets corresponding to the time delay across the fine scale levels, the thresholds and their number are identified. All estimators are shown to be strongly consistent.

Key words and phrases: Thresholds, time delay, wavelets.

1. Introduction

Recently, wavelets have interested more and more statisticians. Donoho et al. (1995) successfully applied wavelets to estimate a regression function and obtained nearly optimal estimators in a large function class, the Besov space B^{σ}_{pq}. Following Donoho's idea, Neumann (1994) used wavelets to estimate the power spectrum in time series. Wang (1995), Luan and Xie (1995) dealt with change points by wavelets. Li and Xie (1997) used wavelets in identifying the hidden periodicities in time series. In this paper, we identify the thresholds and time delay of the threshold autoregressive models by wavelets.

Threshold autoregressive models were introduced by Tong (1978) to describe complex stochastic systems. Since then, they have become some of the most widely used nonlinear time series models in the literature. One of the threshold autoregressive models, the self-exciting threshold autoregressive model, is defined as follows:

$$x_t = \sum_{l=1}^{r+1}\Big(b_0^{(l)} + \sum_{m=1}^{p_l} b_m^{(l)} x_{t-m} + \epsilon_t^{(l)}\Big) I_{(\lambda_{l-1},\lambda_l]}(x_{t-d}), \tag{1.1}$$

where for each l, $\{\epsilon_t^{(l)}, t = 1, 2, \ldots\}$ are i.i.d. random variables with mean zero and variance $\sigma_l^2, l = 1, 2, \ldots, r+1$, and $\{\epsilon_t^{(l)}, t = 1, 2, \ldots\}, l = 1, 2, \ldots, r+1$ are mutually independent, $\lambda_0 = -\infty$, and $\lambda_{r+1} = \infty$. It is assumed that $p_l \le p$ with p being a known integer and $d \le p$. Let $b_s^{(l)} = 0$ when $s > p_l, l = 1, 2, \ldots, r+1$.

Then model (1.1) can be written as

$$x_t = \sum_{l=1}^{r+1} \Big(b_0^{(l)} + \sum_{m=1}^{p} b_m^{(l)} x_{t-m} + \epsilon_t^{(l)} \Big) I_{(\lambda_{l-1}, \lambda_l]}(x_{t-d}) \tag{1.2}$$

and is called a self-exciting threshold autoregressive model with order p – $SETAR(d, r; p, \ldots, p)$, where $\{\lambda_l\}$ are called thresholds and d the time delay.

Model (1.2) is nonlinear with a piecewise linear skeleton. If the thresholds λ_l and time delay d are known, then the parameters $b_i^{(j)}$ can easily be estimated in the same way as those of $AR(p)$ models. However, it is a real challange for statisticians to estimate the thresholds λ_l and time delay d since d is an integer and the number of λ_l is unknown. There have been some results about the estimation of λ_l and d. Chan and Tong (1986) gave a method to estimate the threshold of $SETAR(d, 1; p, \ldots, p)$ models with only one threshold and known time delay. Later, Chan (1988) and Tong (1990) gave strongly consistent estimators of the time delay and thresholds by using conditional least squares under the condition that the number of thresholds is known. Furthermore, as the time delay is an integer and its estimate is strongly consistent, the authors took it as given when they discussed the asympototic properties of the estimators of thresholds and other parameters. After Chan and Tong (1986), Geweke and Terui (1993), and Chen and Lee (1995) took an alternative Bayesian approach to identify the threshold and time delay of $SETAR(d, 1; p, \ldots, p)$ with only one threshold and unknown time delay. However, there have been no results in the general case up to now.

One of the examples solved successfully by $SETAR(d, r; p, \ldots, p)$ models is the Sun spots data (Tong (1980)). The data consist of the number of Sun spots from 1749 to 1924. Through Box-Cox transformation of the original data, $x_t = 2((w_t)^{1/2} - 1)$, the transformed data x_t admits $SETAR(8, 1; 11, \ldots, 11)$:

$$x_t = \begin{cases} 1.9191 + 0.8416 x_{t-1} + 0.0728 x_{t-2} - 0.3153 x_{t-3} \\ +0.1479 x_{t-4} - 0.1985 x_{t-5} - 0.0005 x_{t-6} + 0.1875 x_{t-7} \\ -0.2701 x_{t-8} + 0.2116 x_{t-9} + 0.0091 x_{t-10} + 0.0873 x_{t-11} \\ +\epsilon_t^{(1)} \qquad \text{if } x_{t-8} \le 11.9824 \\ 4.2746 + 1.4431 x_{t-1} - 0.8408 x_{t-2} + 0.0554 x_{t-3} \\ +\epsilon_t^{(2)} \qquad \text{if } x_{t-8} > 11.9824. \end{cases} \tag{1.3}$$

There is only one threshold in this model: $\lambda = 11.9824$ and time delay $d = 8$. The threshold $\lambda = 11.9824$ determines the nonlinear structure of the model and plays a very important role in the model.

In the following, we always base our discussion on model (1.2). Let

$$T(\underline{x}) = \sum_{l=1}^{r+1} \Big(b_0^{(l)} + \sum_{m=1}^{p} b_m^{(l)} x_m \Big) I_{(\lambda_{l-1}, \lambda_l]}(x_d)$$

with $\underline{x} = (x_1, x_2, \ldots, x_p)^\tau$ and

$$\underline{x}_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})^\tau \quad , \quad \underline{\epsilon}_t = (\epsilon_t^{(1)}, \epsilon_t^{(2)}, \ldots, \epsilon_t^{(r+1)})^\tau,$$

$$\underline{\alpha}(\underline{x}_{t-1}) = (I_{(-\infty,\lambda_1]}(x_{t-d}), I_{(\lambda_1,\lambda_2]}(x_{t-d}), \ldots, I_{(\lambda_{r-1},\lambda_r]}(x_{t-d}), I_{(\lambda_r,\infty)}(x_{t-d}))^\tau.$$

Then model (1.2) has the following form:

$$x_t = T(\underline{x}_{t-1}) + \underline{\alpha}(\underline{x}_{t-1})^\tau \underline{\epsilon_t}. \tag{1.4}$$

It is easily seen from (1.2) that λ_l is a threshold of $SETAR(d, r; p, \ldots, p)$ if and only if

(1) $\underline{x}_l = (x_1, \ldots, x_{d-1}, \lambda_l, x_{d+1}, \ldots, x_p)^\tau$ is a jump point of T, i.e., there exists a point $\underline{x}_l^0 = (t_1, t_2, \ldots, t_{d-1}, \lambda_l, t_{d+1}, \ldots, t_p)^\tau$ in R^p such that

$$T(\underline{x}_l^0 - 0) \neq T(\underline{x}_l^0 + 0) \tag{1.5}$$

or

(2) $\underline{x}_l$ is a cusp point of T.

Condition (1.5) implies that there exists at least one $s \neq d$ such that $b_s^{(l)} - b_s^{(l+1)} \neq 0$, and (2) means that T does not have a derivative at $\underline{x}_l$. Then we have $b_s^{(l)} = b_s^{(l+1)}$ for $s = 1, \ldots, p, s \neq d$ and $b_d^{(l)} - b_d^{(l+1)} \neq 0$.

Wavelets can catch jumps and transient phenomena easily and have good local properties. Because the $\underline{x}_l$'s are the jump or cusp points of T, we can use wavelets to identify the thresholds λ_l. In the following, we develop a simple procedure to identify the thresholds and time delay in the general case by using wavelets. We take a similar step to Chan and Tong (1986) by firstly identifying the time delay, then estimating the thresholds based on the identified time delay. All estimators are shown to be strongly consistent.

The paper is arranged as follows. Section 2 contains preliminaries of the wavelet method. Section 3 gives main results. Section 4 presents simulations of our wavelet method.

2. Preliminaries

Model (1.2) is considered in the following discussions. We assume that $-\infty < a < \lambda_1 < \lambda_2 < \cdots < \lambda_r < b < \infty$, with a and b being known constants and $1 \leq d \leq p$, where p is a known integer, but d, r and $\lambda_l, l = 1, \ldots, r$, are unknown constants. We make the following assumptions about $\{x_t\}$ and noises $\{\epsilon_t^{(l)}\}$:

(**A1**) $\{x_t\}$ is geometrically ergodic;

(**A2**) The p.d.f. g_l and f of $\epsilon_t^{(l)}$ and $\underline{x}_p = (x_p, x_{p-1}, \ldots, x_1)^\tau$ are bounded on $\mathcal{R} = (-\infty, \infty)$ and $\mathcal{R}^p$ respectively. They are also bounded away from zero on some open subsets U and $U^p$of $\mathcal{R}$ and $\mathcal{R}^p$ respectively, and for $[a, b] \subset U$:

$$\begin{aligned} &g_l(x),\ f(\underline{x}) \leq M_1,\ x \in \mathcal{R},\ \underline{x} \in \mathcal{R}^p, \\ &M_2 \leq g_l(x),\ f(\underline{x}),\ x \in U,\ \underline{x} \in U^p, \end{aligned} \tag{2.1}$$

where $M_1 > 0$ and $M_2 > 0$ are constants. Furthermore $\{\underline{\epsilon}_t\}$ are independent of $\underline{x}_p$.

Assumption (A1) is made to ensure that $\{x_t\}$ is stationary and strong mixing. Many sufficient conditions for (A1) have been proposed (see Auestad and Tjϕstheim (1990) or Chen (1996)). For example, when $p = 1$, if $\sup_{|x|>c} |\frac{T(x)}{x}| < 1$ for some constant $c > 0$, then x_t is geometrically ergodic (Auestad and Tjϕstheim (1990)). Assumption (A2) with strong mixing condition ensures that there exist enough $\underline{x}_t$ in any neighborhood of $\underline{x} \in [a, b] \times [a, b] \times \cdots \times [a, b]$ such that we can replace $T(\underline{x})$ by $T(\underline{x}_t)$.

We take a wavelet $\psi(x), x \in \mathcal{R}$, satisfying the following condition.

(**A3**) Compactly supported on $[-A, A]$ with $A > 1$, ψ is of bounded variation on $[-A, A]$, and $\psi(x) = 0,\ x \in [-1, 1]$. Furthermore,

$$\int_{-A}^{A} \psi(x)dx = 0,\ \int_{-A}^{A} x\psi(x)dx = 0,\ \int_{-A}^{A} \psi^2(x)dx < \infty, \tag{2.2}$$

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

Condition (2.2) is the usual assumption for a wavelet with the first moment vanishing. Obviously there are many functions on $\mathcal{R}$ satisfying (A3).

From the wavelet ψ and scale function ϕ, we can obtain an orthogonal wavelet basis on $L_2[a, b]$:

$$\{\phi^{per}_{l,k},\ k \in I_l,\ \psi^{per}_{j,k},\ k \in I_j,\ j \geq l\}, \tag{2.3}$$

where

$$\begin{aligned}\phi^{per}_{l,k}(x) &= \sum_n (b-a)^{-1/2}\phi_{l,k}\Big(\frac{x-a}{b-a} + n\Big),\\ \psi^{per}_{j,k}(x) &= \sum_n (b-a)^{-1/2}\psi_{j,k}\Big(\frac{x-a}{b-a} + n\Big),\end{aligned} \tag{2.4}$$

with $\phi_{l,k}(x) = 2^{l/2}\phi(2^l x - k),\ \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ and $I_j = \{0, 1, 2, \ldots, 2^j - 1\}$.

For $f \in L_2[a, b]$, we have

$$f(x) = \sum_{k \in I_l} \alpha_{l,k}\phi^{per}_{l,k}(x) + \sum_{j \geq l}\sum_{k \in I_j} \beta_{j,k}\psi^{per}_{j,k}(x), \tag{2.5}$$

where

$$\alpha_{l,k} = \int_a^b \phi^{per}_{l,k}(x)f(x)dx,\ \beta_{j,k} = \int_a^b \psi^{per}_{j,k}(x)f(x)dx. \tag{2.6}$$

The expression (2.5) is called the wavelet expansion of f on $[a, b]$ and $\beta_{j,k}$ is called a wavelet coefficient of f.

It is easily shown that

$$\beta_{j,k} = (b-a)^{1/2}2^{-j/2}\int_{-A}^{A} f(\frac{x+k}{2^j}(b-a) + a)\psi(x)dx \tag{2.7}$$

For details, see Cohen et al. (1993).

3. Main Results

Suppose that $x_t, 1 \le t \le n$, are sampled from model (1.2). Let

$$N_n(\underline{s}) = \{l : 1 \le l \le n,\ ||\underline{x}_{l-1} - \underline{s}|| \le \delta_n,\ \underline{x}_{l-1} \le \underline{s}\},\ n_{\underline{s}} = \# N_n(\underline{s}),$$

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

$$R_{j^*} = \{\underline{t}_{j^*} = (t_{j^*1}, t_{j^*2}, \ldots, t_{j^*(p-1)})^\tau :\ t_{j^*l} = a + \frac{k_l}{2^{j^*}}(b-a),\ k_l \in I_{j^*}\}$$

for fixed j and j^*, where

$$\delta_n = n^{-1/(p+2)},\ \underline{x}_{l-1} = (x_{l-1}, x_{l-2}, \ldots, x_{l-p})^\tau,\ \underline{s} = (s_1, s_2, \ldots, s_p)^\tau$$

and $||.||$ is the Euclidean norm.

We construct the empirical wavelet coefficient as follows:

$$W_{j,k}^{(m)}(\underline{t}_{j^*}) = \frac{1}{N}\sum_{i=1}^{N} \psi_{j,k}^{per}(s_i)\frac{1}{n_{m,i}}\sum_{l \in N_n(\underline{s}_{m,i})} x_l, \tag{3.1}$$

$m = 1, \ldots, p$, where $\underline{t}_{j^*} \in R_{j^*},\ N = [n^{1/(p+2)}]$, and

$$\underline{s}_{m,i} = (t_{j^*1}, t_{j^*2}, \ldots, t_{j^*(m-1)}, s_i, t_{j^*m}, \ldots, t_{j^*(p-1)})^\tau,$$

with $s_i = a + i\frac{b-a}{N}$, and $n_{m,i} = \# N_n(\underline{s}_{m,i})$. Here $[x]$ denotes the integral part of x.

Our idea to construct (3.1) follows from the following fact: for fixed $\underline{t} \in [a,b]^{p-1}$, the wavelet coefficients of $T(t_1, \ldots, t_{m-1}, x, t_m, \ldots, t_{p-1}), x \in [a,b]$ are

$$\beta_{j,k}^{(m)}(\underline{t}) = \int_a^b T(t_1, \ldots, t_{m-1}, x, t_m, \ldots, t_{p-1})\psi_{j,k}^{per}(x)dx. \tag{3.2}$$

It follows from simple computation that when $m \ne d, \beta_{j,k}^{(m)}(\underline{t}) = 0$ for all $\underline{t} \in [a,b]^{p-1}$, and when $m = d$, there exists $\underline{t}_0 \in [a,b]^{p-1}$ such that the $\beta_{j,k}^{(d)}(\underline{t}_0)$ have large absolute values across fine scale levels. We discretize the right side of (3.2) and replace $T(\underline{x}_{t-1})$ by x_t. Then the sample version $W_{j,k}^{(m)}(\underline{t})$ of $\beta_{j,k}^{(m)}(\underline{t})$, called empirical wavelet coefficients, are obtained. In Theorem 3.1, we show that $W_{j,k}^{(m)}(\underline{t})$ have the same properties as $\beta_{j,k}^{(m)}(\underline{t})$. It is these properties of $W_{j,k}^{(m)}(\underline{t})$ that we use to identify the time delay d and thresholds λ_l.

Theorem 3.1. *Assume* (A1), (A2) *and* (A3) *are true. If* $\lim_{\substack{j\to\infty \\ n\to\infty}} \frac{2^{3j}}{N} = 0$, *then there exists a large enough* $j^* > 0$ *such that as* $j \longrightarrow \infty$,

(1) *there exist* $\underline{t}_{j^*}^0 \in R_{j^*}$ *and a constant* $c_0 > 0$, *independent of* j,k *and* j^*, *such that*

$$|W_{j,k}^{(d)}(\underline{t}_{j^*}^0)| \ge c_0 2^{-3j/2} \quad a.s. \tag{3.3}$$

for $k \in I(\lambda_l, 2^{-j})$.

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

$$W_{j,k}^{(d)}(\underline{t}_{j^*}^0) = o(2^{-3j/2}) \qquad a.s. \tag{3.4}$$

(3) *When* $m \neq d$,

$$\max_{k \in I_j, \underline{t}_{j^*} \in R_{j^*}} |W_{j,k}^{(m)}(\underline{t}_{j^*})| = o(2^{-3j/2}) \qquad a.s. \tag{3.5}$$

Proof. $W_{j,k}^{(m)}(\underline{t}_{j^*})$ can be written as

$$W_{j,k}^{(m)}(\underline{t}_{j^*}) \stackrel{\triangle}{=} W_{j,k}^{(m,T)}(\underline{t}_{j^*}) + W_{j,k}^{(m,R)}(\underline{t}_{j^*}) + W_{j,k}^{(m,\epsilon)}(\underline{t}_{j^*}), \tag{3.6}$$

where

$$W_{j,k}^{(m,T)}(\underline{t}_{j^*}) = \frac{1}{N}\sum_{i=1}^N \psi_{j,k}^{per}(s_i) T(\underline{s}_{m,i}),$$

$$W_{j,k}^{(m,R)}(\underline{t}_{j^*}) = \frac{1}{N}\sum_{i=1}^N \psi_{j,k}^{per}(s_i) \frac{1}{n_{m,i}} \sum_{l \in N_n(\underline{s}_{m,i})} [T(\underline{x}_{l-1}) - T(\underline{s}_{m,i})],$$

$$W_{j,k}^{(m,\epsilon)}(\underline{t}_{j^*}) = \frac{1}{N}\sum_{i=1}^N \psi_{j,k}^{per}(s_i) \frac{1}{n_{m,i}} \sum_{l \in N_n(\underline{s}_{m,i})} \underline{\alpha}(\underline{x}_{l-1})^\tau \underline{\epsilon_l}.$$

We can show that when $j \to \infty$,

$$\max_{k \in I_j, \underline{t}_{j^*} \in R_{j^*}} |W_{j,k}^{(m,R)}(\underline{t}_{j^*})| = o(2^{-3j/2}), \max_{k \in I_j, \underline{t}_{j^*} \in R_{j^*}} |W_{j,k}^{(m,\epsilon)}(\underline{t}_{j^*})| = o(2^{-3j/2}). \tag{3.7}$$

It is easily seen from Lemma P5.1 of Brillinger (1981) that

$$W_{j,k}^{(m,T)}(\underline{t}_{j^*}) = \int_a^b T(t_{j^*1}, \ldots, t_{j^*(m-1)}, x, t_{j^*m}, \ldots, t_{j^*(p-1)}) \psi_{j,k}^{per}(x) dx + O(\frac{2^{j/2}}{N}),$$

where $O(\cdot)$ holds uniformly for $k \in I_j$ and $\underline{t}_{j^*} \in R_{j^*}$.

(1) When $k \in I(\lambda_l, 2^{-j})$ and j is large enough, we have

$$\begin{aligned} W_{j,k}^{(d,T)}(\underline{t}_{j^*}) = {} & 2^{-j/2}(b-a)^{1/2} \int_1^A \psi(x) dx \Big[b_0^{(l+1)} - b_0^{(l)} \\ & + (b_1^{(l+1)} - b_1^{(l)}) t_{j^*1} + \cdots + (b_{d-1}^{(l+1)} - b_{d-1}^{(l)}) t_{j^*(d-1)} \\ & + (b_d^{(l+1)} - b_d^{(l)})(a + \frac{k}{2^j}(b-a)) + (b_{d+1}^{(l+1)} - b_{d+1}^{(l)}) t_{j^*d} \\ & + \cdots + (b_p^{(l+1)} - b_p^{(l)}) t_{j^*(p-1)} \Big] \\ & + 2^{-3j/2}(b-a)^{3/2} \int_1^A x\psi(x) dx (b_d^{(l+1)} - b_d^{(l)}) + O(\frac{2^{j/2}}{N}). \end{aligned} \tag{3.8}$$

Since λ_l is a threshold of model (1.2), λ_l is a jump point or a cusp point of T. For the former, we know from Section 1 that there exists $\underline{t}_0 \in [a,b]^{p-1}$ such that

$$b_0^{(l+1)} - b_0^{(l)} + (b_1^{(l+1)} - b_1^{(l)})t_1^0 + \cdots + (b_{d-1}^{(l+1)} - b_{d-1}^{(l)})t_{d-1}^0 + (b_d^{(l+1)} - b_d^{(l)})\lambda_l + (b_{d+1}^{(l+1)} - b_{d+1}^{(l)})t_d^0 + \cdots + (b_p^{(l+1)} - b_p^{(l)})t_{p-1}^0 \neq 0.$$

For the latter, we have $b_s^{(l+1)} - b_s^{(l)} = 0,\ s = 1, \ldots, p,\ s \neq d$, and $b_d^{(l+1)} - b_d^{(l)} \neq 0$. Then we know from (3.6),(3.7) and (3.8) that for both cases, there exist a large enough $j^* > 0$, $\underline{t}_{j^*}^0 \in R_{j^*}$ (which approximates $\underline{t}_0$), and a constant $c_0 > 0$ independent of $j,\ k,\ j^*$ and $\underline{t}_{j^*}$ such that

$$|W_{j,k}^{(d)}(\underline{t}_{j^*}^0)| \geq c_0 2^{-3j/2} \qquad a.s.$$

(2) For all $k \notin \cup_{l=1}^p I(\lambda_l, 2^{-j/2})$,

$$|a + \frac{k}{2^j}(b-a) - \lambda_l| > 2^{-j/2},\ l = 1, \ldots, r.$$

Since

$$2^j|a + \frac{k}{2^j}(b-a) - \lambda_l| \longrightarrow \infty\ (j \to \infty),\ \int_{-A}^{A} \psi(x)dx = 0,\ \int_{-A}^{A} x\psi(x)dx = 0,$$

it is trival to show that when j is large enough,

$$\int_{-A}^{A} I_{(\lambda_{l-1},\lambda_l]}(a + \frac{x+k}{2^j}(b-a))\psi(x)dx = 0,$$
$$\int_{-A}^{A} I_{(\lambda_{l-1},\lambda_l]}(a + \frac{x+k}{2^j}(b-a))x\psi(x)dx = 0,$$

for all $k \notin \cup_{l=1}^r I(\lambda_l, 2^{-j/2})$. Then it follows from (3.6) and (3.7) that $W_{j,k}^{(d)}(\underline{t}_{j^*}^0) = o(2^{-3j/2})$ a.s. for all $k \notin \cup_{l=1}^r I(\lambda_l, 2^{-j/2})$.

(3) When $m \neq d$, since

$$\int_a^b T(t_{j^*1}, \ldots, t_{j^*(m-1)}, x, t_{j^*m}, \ldots, t_{j^*(p-1)})\psi_{j,k}^{per}(x)dx = 0$$

for all $\underline{t}_{j^*} \in R_{j^*}$, it follows from (3.6) and (3.7) that

$$\max_{k \in I_j, \underline{t}_{j^*} \in R_{j^*}} |W_{j,k}^{(m)}(\underline{t}_{j^*})| = o(2^{-3j/2}).$$

For j^* in Theorem 3.1, let

$$E^0(j) = \{m : \text{there exist } k \in I_j \text{ and } \underline{t}_{j^*} \in R_{j^*} \text{ such that}$$

$$|W_{j,k}^{(m)}(\underline{t}_{j^*})| \geq c_0 2^{-3j/2},\ 1 \leq m \leq p\}, \tag{3.9}$$

where c_0 is defined in (3.3).

Let

$$\hat{d} = \begin{cases} \text{any element of } E^0(j), & \text{if } E^0(j) \text{ is not empty,} \\ 0, & \text{if } E^0(j) \text{ is empty.} \end{cases} \tag{3.10}$$

Theorem 3.2. *Under the conditions of Theorem* 3.1, $\hat{d} = d$ *a.s., for* j *large enough.*

For $\underline{t}_{j^*}^0$ of Theorem 3.1, let

$$E(j) = \{k : |W_{j,k}^{(\hat{d})}(\underline{t}_{j^*}^0)| \geq c_0 2^{-3j/2}, k \in I_j\}. \tag{3.11}$$

We know from Theorem 3.2 that $\hat{d} = d$ a.s., for j large enough. So

$$E(j) = \{k : |W_{j,k}^{(d)}(\underline{t}_{j^*}^0)| \geq c_0 2^{-3j/2}, k \in I_j\} \tag{3.12}$$

a.s., for j large enough.

It follows from Theorem 3.1 that when j is large enough, $E(j)$ is not empty. Suppose that $E(j) = \{e_1, \ldots, e_m\}$ where $e_1 < e_2 < \cdots < e_m$. Let $\rho = 2 \times 2^{j/2}$ and $m_1 = \max\{l :\ 1 \leq l \leq m,\ e_l \leq e_1 + \rho\}$. If $m_1 < m$, then define $m_2 = \max\{k : m_1 < l \leq m,\ e_l \leq e_{m_1} + \rho\}$; if $m_2 < m$, then define m_3 in a similar way, and so on. At last we can get a series of integers $\{m_l :\ 1 \leq m_1 < m_2 < \cdots < m_q \leq m\}$.

Let $E_1(j) = \{e_l :\ 1 \leq l \leq m_1\},\ E_2(j) = \{e_l :\ m_1 < l \leq m_2\}, \ldots, E_q(j) = \{e_l :\ m_{q-1} < l \leq m_q\}$. Then we have

$$E(j) = \bigcup_{l=1}^{q} E_l(j). \tag{3.13}$$

Equation (3.13) is called the $\rho-$ division of $E(j)$. For details, see Luan and Xie (1995). Let

$$\hat{r} = \begin{cases} q, & \text{if } E(j) \text{ is not empty,} \\ 0, & \text{if } E(j) \text{ is empty.} \end{cases} \tag{3.14}$$

Take k_l, satisfying

$$|W_{j,k_l}^{(\hat{d})}(\underline{t}_{j^*}^0)| = \max_{k \in E_l(j)} |W_{j,k}^{(\hat{d})}(\underline{t}_{j^*}^0)|, \quad l = 1, \ldots, \hat{r}, \tag{3.15}$$

and let

$$\hat{\lambda}_l = a + \frac{k_l}{2^j}(b - a). \tag{3.16}$$

Theorem 3.3. *Under the conditions of Theorem* 3.1, *when* $j \longrightarrow \infty$,

(1) $\hat{r} = r$ *a.s.;*
(2) *for* $r > 0$, $\hat{\lambda}_l \longrightarrow \lambda_l$ *a.s.,* $l = 1, \ldots, r$.

Proof. We know from Theorem 3.2 that $\hat{d} = d$ a.s., for j large enough, so we assume $\hat{d} = d$ in the following. Then $W_{j,k}^{(\hat{d})}(\underline{t}_{j^*}^0) = W_{j,k}^{(d)}(\underline{t}_{j^*}^0)$ for j large enough.

In the case of $r = 0$, T is a linear function of $\underline{x}$. It is easily seen that

$$\int_a^b \psi_{j,k}^{per}(x_i) T(\underline{x}) dx_i = 0$$

for $i = 1, \ldots, p$. Then $W_{j,k}^{(d)}(\underline{t}_{j^*}^0) = o(2^{-3j/2})$. Therefore $E(j)$ is empty a.s. Hence $\hat{r} = 0$ a.s.

In the case of $r > 0$, we know from Theorem 3.1 that when j is large enough, $E(j)$ is not empty a.s. Then

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

From Theorem 3.1, it can also be shown that

$$\cup_{l=1}^{r} I(\lambda_l, 2^{-j}) \subset \cup_{l=1}^{\hat{r}} E_l(j) \subset \cup_{l=1}^{r} I(\lambda_l, 2^{-j/2}).$$

Due to $\rho-$ division ($\rho = 2 \times 2^{j/2}$), we have

$$E(j) = \cup_{l=1}^{r} E_l(j), I(\lambda_l, 2^{-j}) \subset E_l(j) \subset I(\lambda_l, 2^{-j/2}).$$

Therefore $\hat{r} = r$ a.s. and $\hat{\lambda}_l \in I(\lambda_l, 2^{-j/2})$ a.s., i.e., $\hat{\lambda}_l \xrightarrow{a.s.} \lambda_l$.

4. Numerical Simulations

To test the proposed method in the present paper, we carry out simulations for the Sun spots data and the models

$$x_t = \begin{cases} 0.4x_{t-1} + 0.26x_{t-2}, & \text{if } x_{t-2} \le 0.35, \\ 0.2x_{t-1} - 4.2x_{t-2}, & \text{if } 0.35 < x_{t-2} \le 0.5, \\ 0.3x_{t-1} + 0.6x_{t-2}, & \text{if } 0.5 < x_{t-2}, \end{cases} \quad + \quad \sigma\, \epsilon_t \tag{4.1}$$

with $\sigma = 0.2$, 0.4 and 0.5, and

$$x_t = \begin{cases} 0.4x_{t-1} + 0.3x_{t-2}, & \text{if } x_{t-1} \le 0.4, \\ -4.6x_{t-1} + 0.9x_{t-2}, & \text{if } 0.4 < x_{t-1} \le 0.55, \\ 0.5x_{t-1} + 0.2x_{t-2}, & \text{if } 0.55 < x_{t-1} \le 0.7, \\ -2.8x_{t-1} + 2.7x_{t-2}, & \text{if } 0.7 < x_{t-1} \le 0.85, \\ 0.6x_{t-1} + 0.3x_{t-2}, & \text{if } 0.85 < x_{t-1}, \end{cases} \quad + \quad \sigma\, \epsilon_t \tag{4.2}$$

with $\sigma = 0.4$ or 0.6, where $\{\epsilon_t\}$ are i.i.d. $N(0,1)$ white noises.

We take ψ as follows:

$$\psi(x)=\begin{cases}5(x-1)^4, & \text{if } 1\le x\le 2,\\ \frac{20}{3}(x+1)^3+2(x+1)^2, & \text{if } -2\le x\le -1,\\ 0, & \text{otherwise.}\end{cases} \tag{4.3}$$

Figure 1 and Figure 2 show the detection of the thresholds and time delay by wavelets for model (4.1) with $\sigma=0.2$. Here we carry out simulations with the resolution level j from 1 to 10. At level $j=7$, we find that the empirical wavelet coefficients $W^{(1)}_{j,k}$ are significantly smaller than the threshold value $w=0.13$, and $W^{(2)}_{j,k}$ are significantly larger than w only near 0.35 and 0.5. Therefore $\hat d=2$. The estimated thresholds are $\hat\lambda_1=0.35$ and $\hat\lambda_2=0.50$, with $\hat r=2$.

If we increase σ respectively to 0.4 and 0.5, then we see from Figure 3 and Figure 4 that wavelets work well for $\sigma=0.4$, but fail for $\sigma=0.5$, which reflects the influence of the noise on the empirical wavelet coefficients. Figure 5 and Figure 6 show the work for model (4.2) with $\sigma=0.4$ and $j=7$, and Figure 7 shows the work for model (4.2) with $\sigma=0.6$ and $j=7$. It can be seen that wavelets work well for both cases.

For the Sun spots data ($n=176$), we make the transformation: $x_t=((w_t)^{1/2}-1)/10$. Take $a=0.25$, $b=1.2$ and wavelet ψ in (4.3). For $p=11$ and $\underline{t_0}=(0.5,0.5,0.6,0.7,0.55,0.4,0.45,0.65,0.75,0.5)^\tau$, we calculate $W^{(m)}_{j,k}(\underline{t_0})$, $m=1,2,\ldots,11$ at $j=7$. Figures 8-18 show that the wavelet works for the Sun spots data (the abscissas in Figures 8-18 corresponds to the Box-Cox transformation $x_t=2((w_t)^{1/2}-1)$). It is easily seen from Figures 8-18 that only the $W^{(8)}_{j,k}(\underline{t_0})$ have large absolute values near 12.15. So the estimated time delay $\hat d=8$, and there is only one threshold $\hat\lambda=12.15$. Therefore the wavelet results are basically consistent with (1.3) where $d=8$, $\lambda=11.9824$.

Acknowledgements

The authors thank the referees for a very careful reading of the the paper and for their valuable comments and suggestions which led to improvement on earlier version of this paper. They also thank the referees for corrections to the English writing. The research was supported by NNSF of China and the doctoral grant of China.

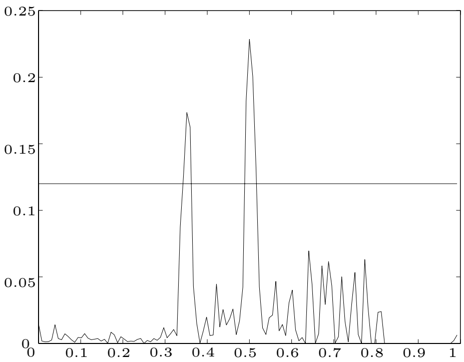

Figure 1. The second wavelet for model (4.1) with sigma=0.2.

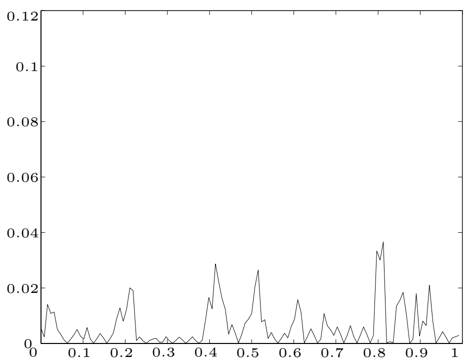

Figure 2. The first wavelet for model (4.1) with sigma=0.2.

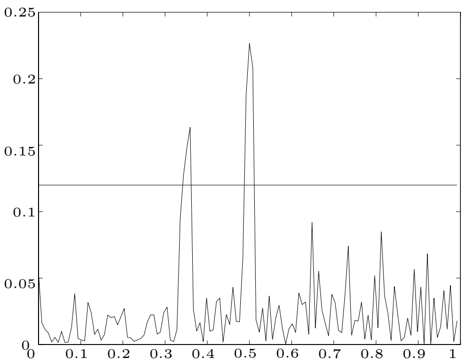

Figure 3. The second wavelet for model (4.1) with sigma=0.4.

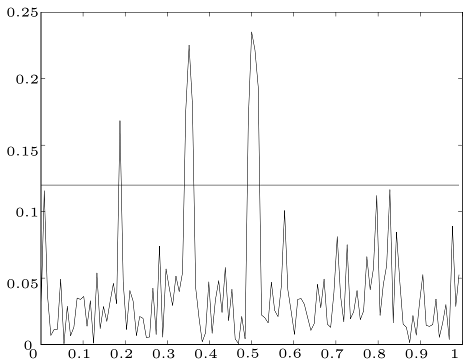

Figure 4. The second wavelet for model (4.1) with sigma=0.5.

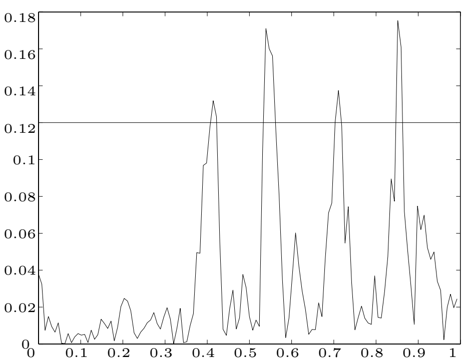

Figure 5. The first wavelet for model (4.2) with sigma=0.4.

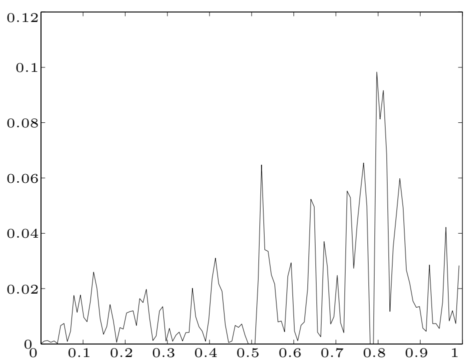

Figure 6. The second wavelet for model (4.2) with sigma=0.4.

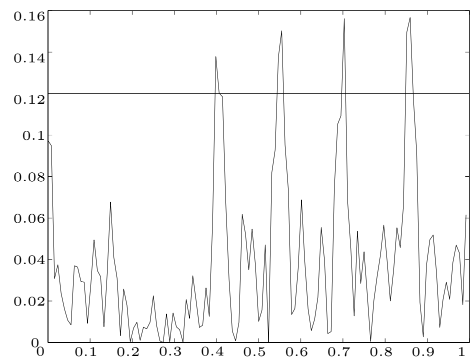

Figure 7. The first wavelet for model (4.2) with sigma=0.6.

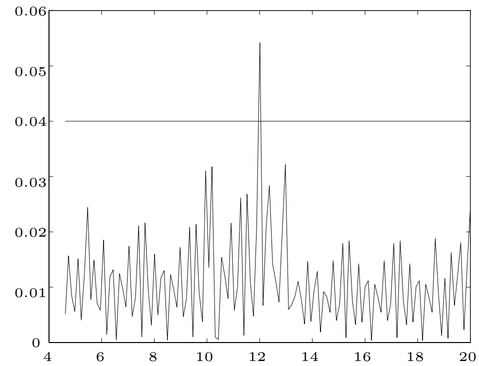

Figure 8. The eighth wavelet for the Sun spots data.

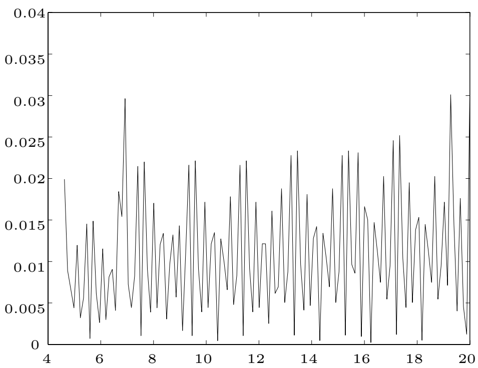

Figure 9. The first wavelet for the Sun spots data.

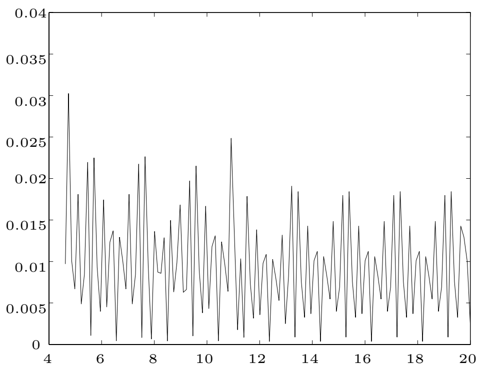

Figure 10. The second wavelet for the Sun spots data.

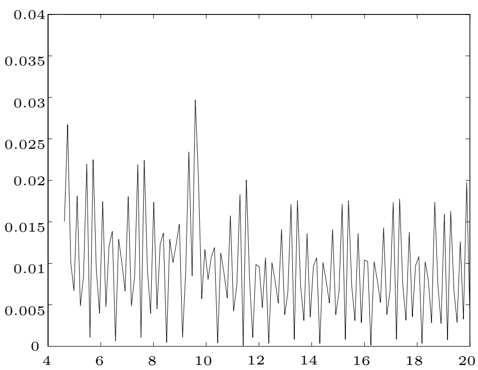

Figure 11. The third wavelet for the Sun spots data.

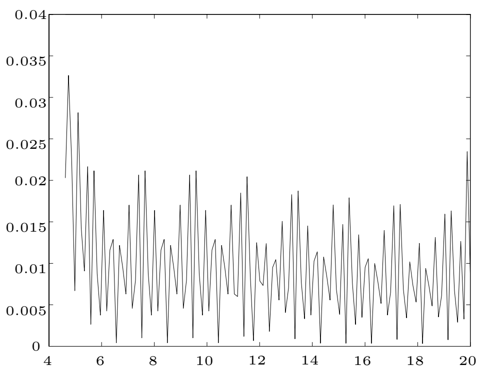

Figure 12. The fourth wavelet for the Sun spots data.

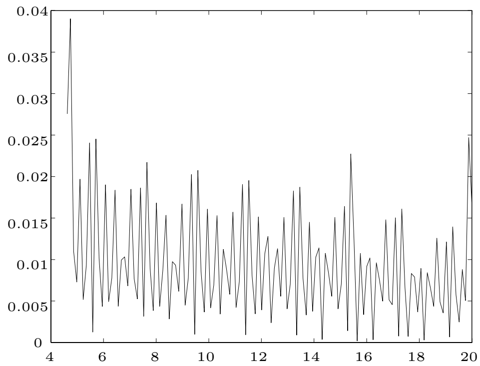

Figure 13. The fifth wavelet for the Sun spots data.

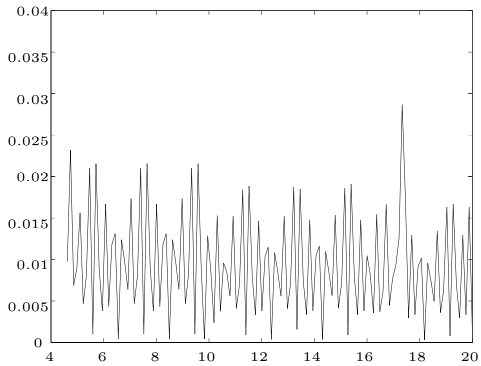

Figure 14. The sixth wavelet for the Sun spots data.

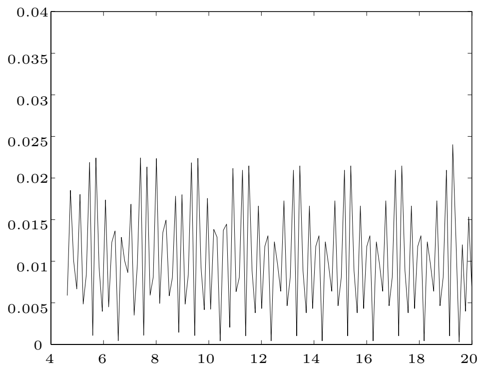

Figure 15. The seventh wavelet for the Sun spots data.

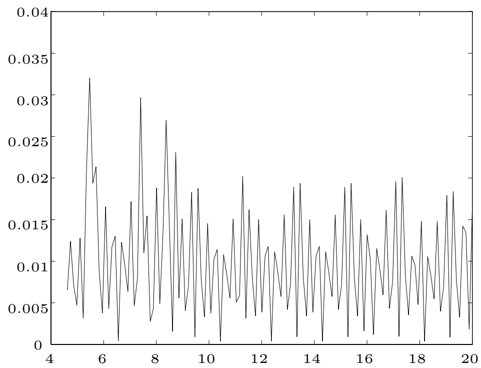

Figure 16. The ninth wavelet for the Sun spots data.

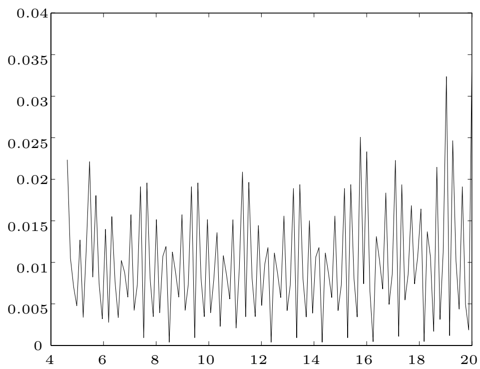

Figure 17. The tenth wavelet for the Sun spots data.

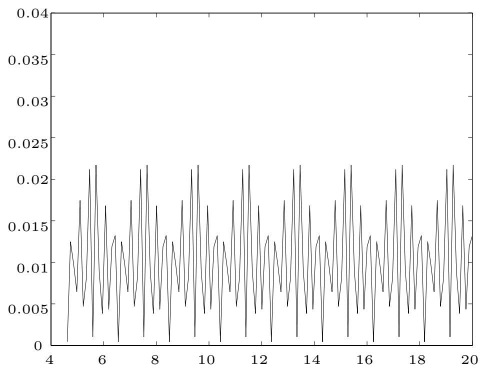

Figure 18. The eleventh wavelet for the Sun spots data.

References

Auestad, B. and Tjϕstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika* **77**, 669-687.

Brillinger, D. R. (1981). *Time Series-Data Analysis and Theory.* Holden-Day, San Francisco.

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

Chan, K. S. (1988). Consistency and limiting distribution of least squares estimation of a threshold autoregressive model. Technical Report No. 245, Department of Statistics, University of Chicago.

Chen, C. W. S. and Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *J. Time Ser. Anal.* **16**, 483-492.

Chen, M. (1996). Probabilistic properties and statistical analysis for time series models with conditional heteroscedasticity. Ph.D. Dissertation, Institute of Applied Mathematics, Academia Sinica.

Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1**, 54-81.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B* **57**, 301-369.

Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *J. Time Ser. Anal.* **14**, 441-454.

Li, Y. and Xie, Z. (1997). The wavelet detection of hidden periodicities in time series. *Statist. Probab. Lett.* **35**, 9-23.

Luan, Y. and Xie, Z. (1995). The wavelet detection for the jump points of signals with additive stationary noises. Tech. Rept. No. 29, Inst. of Math. Statist., Peking University.

Neumann, M. H. (1994). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. Statistics Research Report No. SRR 028-94, CMA, Australian National University, Canberra.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, NATO ASI Series E: applied SC. No. 29 (Edited by C. H. Chen). Sijthoff & Noordhoff, Groninaen.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. Ser. B* **42**, 245-292.

Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis.* Springer lecture notes in statistics, 21. Springer, New York.

Truong, Y. K. and Stone, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20**, 77-97.

Truong, Y. K. and Stone, C. J. (1994). Semiparametric time series regression. *J. Time Ser. Anal.* **15**, 405-428.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385-397.

Department of Applied Mathematics, University of Petroleum, Dongying, Shandong 257062.

E-mail: liyuan@hdpu.edu.cn

Department of Probability and Statistics, Peking University, Beijing 100871.

(Received March 1997; accepted May 1998)