Statistics & Probability Letters 53 (2001) 167–180

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

The wavelet identification for jump points of derivative in regression model☆

Yihui Luan *, Zhongjie Xie

Department of Probability and Statistics, School of Mathematical Sciences, Peking University, Beijing 100871, People's Republic of China

Received September 1999; received in revised form February 2001

Abstract

A method is proposed to detect the number, locations and heights of jump points of the derivative in the regressive model $\eta_i = f(\xi_i) + \varepsilon_i$, by checking if the empirical indirect wavelet coefficients of data have significantly large absolute values across fine scale levels. The consistency of the estimators is established and practical implementation is discussed. Some simulation examples are given to test our method.

Keywords: Wavelets; Scale function; Derivative; Stationary; Jump

1. Introduction

The detection of jump points in nonparametric models has attracted increasing interests. Since jump points usually describe sudden localised changes, they are very useful in modelling practical problems arising in fields such as quality control, economics, medicine, signal and image processing, phonetic identification, and physical sciences.

There is a great amount of statistical literature on jump detection, see for example, Basseville (1988), Basseville and Nikiforov (1993), Li and Xie (1999,2000). Yin (1988) proposed strongly consistent estimators of the number, locations, and corresponding jump heights of the jump points by using the one-sided moving average method. Müller (1992) estimated the location of a jump and its size by boundary kernels. Wu and Chu (1993) gave strongly consistent estimators of jumps based on the kernel method. Wang (1995) showed interesting results on the jump point detection and gave a detecting procedure by wavelets. All the above results are based on the assumption of uncorrelated white noise or Gaussian processes. Ogden and Parzen (1996) transformed the change point problem into a nonparametric regression problem, but their starting point

☆ This research work was supported by the National Natural Science Foundation of China (Grant No. 19971006).

* Corresponding author.
E-mail address: lyh@math.pku.edu.cn (Y. Luan).

0167-7152/01/$ - see front matter
PII: S0167-7152(01)00070-0

was a set of independent observations and Gaussian properties were used. As pointed out in Johnstone and Silverman (1997), the case of correlated noise has not been studied in great detail within the context of wavelets, though Johnstone (1999) considered the wavelet threshold estimation for data with correlated noise. They discussed both short and long range dependent noises and showed several interesting examples. However, their theoretical framework in handling the correlated noise still needs the Gaussian assumption.

Wang (1999) assumes the fractional Gaussian noise model

$$Y(\mathrm{d}x) = (Kf)(x)\,\mathrm{d}x + \tau^{2-2H}B_H(\mathrm{d}x), \quad 0 \leqslant x \leqslant 1, \tag{1.1}$$

where $f(x)$ is a deterministic function defined on [0,1] with several jumps and sharp cusps, K is a linear transformation, τ is the noise level, and $B_H(\mathrm{d}x)$ is a fractional Gaussian noise. With the total observation $Y(x)$ from model (1.1), Wang gave consistent estimators for the number and locations of jumps and sharp cusps of $f(x)$ via wavelet-vaguelette decomposition.

However, in many practical situations the observed data are obtained through a random design manner. More precisely, the observation model can be written as follows:

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1.2}$$

where $f(t)$ is a deterministic function defined on interval $[0, 1]$; x_i are random-signed sampling points; $\{\varepsilon_i\}$ is a random noise; $\{Y_i, x_i\}$ are the observed data and n is the data length. Obviously, model (1.2) is not covered by the discrete version of model (1.1). In this paper we will focus on model (1.2). Besides establishing consistent estimators for the number and locations of jumps we also give estimators for heights of jumps which were not considered in Wang (1999).

The rest of this paper is organised as follows. Section 2 introduces the basic model and some notations and assumptions which are necessary in the sequel. Section 3 gives the main results of this paper and numerical simulations are given in Section 4. All the proofs are collected in the appendix.

2. Preliminaries

In this paper, we consider the following observation model:

$$\eta_i = f(\xi_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{2.1}$$

For our discussions, we need the following assumptions.

(A1) $f(x)$ is continuous on $[0, 1]$. With exception to q points in (0,1), i.e., $\exists 0 < t_1 < t_2 < \cdots < t_q < 1$, the derivatives of $f(x)$ exist. And if we write $s(x) = (\mathrm{d}/\mathrm{d}x)f(x)$, the derivative $s(x)$ has jumps at t_k $(k = 1, 2, \ldots, q)$, and satisfies

$$\sup_{\substack{x \neq t_j \\ j=1,\ldots,q}} \left| \frac{\mathrm{d}}{\mathrm{d}x} s(x) \right| < +\infty. \tag{2.2}$$

(A2)

- $\{\varepsilon_i\}$ is a weakly stationary sequence with zero mean and its autocovariance functions γ_k are summable: $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$ where $\gamma_k = E(\xi_i \xi_{i+k})$.
- $\{\xi_i\}$ is a strictly stationary sequence with zero mean. The p.d.f. of ξ_1 $g(x)$ is bounded away from zero and infinity on some open set which contains interval [0,1], i.e., there exist an open set $\mathscr{U}$ and a positive number M_1 such that $\mathscr{U} \supset [0,1]$, and

$$\frac{1}{M_1} \leqslant g(x) \leqslant M_1, \quad x \in \mathscr{U}.$$

- There exists a positive number M_2 such that under $\xi_1 = x$, the conditional p.d.f. of ξ_l, $g_l(y|x)$, satisfies

$$\frac{1}{M_2} \leqslant g_l(y|x) \leqslant M_2, \quad y,x \in \mathscr{U}.$$

- $\{\xi_i\}$ is strongly mixing with $\alpha_u = O(\rho^u)$, $(0 < \rho < 1)$ where

$$\alpha_u = \sup_{A\in\mathscr{F}_t, B\in\mathscr{F}^{t+u}} \{|P(AB) - P(A)P(B)|\},$$

$$\mathscr{F}_t = \mathscr{F}(\xi_k; k \leqslant t), \quad \mathscr{F}^{t+u} = \mathscr{F}(\xi_k; k \geqslant t+u).$$

Here $\mathscr{F}(\xi_k;\ k \leqslant t)$ and $\mathscr{F}(\xi_k;\ k \geqslant t+u)$ represent the σ-algebras generated by $\{\xi_k;\ k \leqslant t\}$ and $\{\xi_k;\ k \geqslant t+u\}$, respectively.
- $\{\xi_i\}$ and $\{\varepsilon_i\}$ are mutually independent.
We choose wavelet $\psi(x)$ and scale function $\phi(x)$ satisfying the following conditions.
(A3)

- Both $\psi(x)$ and $\phi(x)$ have finite supports, say, $[-\sigma, \sigma]$, $\sigma \geqslant 1$. And both have derivatives with bounded variation. For convenience, we use other symbols to denote the derivatives:

$$\Psi(x) \triangleq \frac{\mathrm{d}}{\mathrm{d}x}\psi(x), \quad \Phi(x) \triangleq \frac{\mathrm{d}}{\mathrm{d}x}\phi(x). \tag{2.3}$$

- $\psi(x)$ and $\phi(x)$ satisfy the usual conditions

$$\int_{-\infty}^{+\infty} \psi(x)\,\mathrm{d}x = 0, \qquad \int_{-\infty}^{+\infty} \phi(x)\,\mathrm{d}x = 1. \tag{2.4}$$

- We require $\psi(x)$ to satisfy the following additional constraint:

$$\inf_{0<a<1/2} \left\{ \left| \int_{-\infty}^{-a} \psi(x)\,\mathrm{d}x \right|, \left| \int_{a}^{+\infty} \psi(x)\,\mathrm{d}x \right| \right\} = b_0 > 0. \tag{2.5}$$

As Wang (1995) pointed out, the localised information of trend function is provided by its wavelet coefficients at fine scales. More precisely, at fine scales, the wavelet coefficients nearby the jumps are significantly larger than those a little farther away from the jumps. Therefore, the jumps can be detected by checking the absolute values of wavelet coefficients. When the resolution level gets large, there may be several large wavelet coefficients near a single jump. So how to combine those wavelet coefficients which have large absolute values into suitable subgroups is a key point. Each subgroup should be regarded as the result from a single jump. Because the model considered by Wang (1995) is different from ours so his methods cannot be used directly to our model (2.1). Consequently, we give a special kind of division for integer sets which is used to determine whether nearby large coefficients belong to different jumps or not.

Let G be an integer set consisting of finite elements and τ be a positive number, $G = \{g_1, g_2, \dots, g_m\}$, where $g_1 < g_2 < \cdots < g_m$. Put $m_1 = \max\{k\colon 1 \leqslant k \leqslant m,\ g_k \leqslant g_1 + \tau\}$; if $m_1 < m$, then put $m_2 = \max\{k\colon m_1 < k \leqslant m,\ g_k \leqslant g_{m_1+1} + \tau\}$; if $m_2 < m$, then define m_3 in a similar way, and so on. At the end we can get a series of integers $\{m_k\colon 1 \leqslant m_1 < m_2 < \cdots < m_q = m\}$.

Let $G_1 = \{g_k\colon 1 \leqslant k \leqslant m_1\}$, $G_2 = \{g_k\colon m_1 < k \leqslant m_2\}, \dots,$ $G_q = \{g_k\colon m_{q-1} < k \leqslant m_q\}$ and put

$$G = \bigcup_{k=1}^{q} G_k. \tag{2.6}$$

For example, $G = \{2, 5, 6, 8, 11, 12, 13, 15\}$, $\tau = 3$, then $m_1 = 2$, $m_2 = 4$, $m_3 = 7$, $m_4 = 8$. Thus, we have $G_1 = \{2, 5\}$, $G_2 = \{6, 8\}$, $G_3 = \{11, 12, 13\}$, $G_4 = \{15\}$ and the 3-division of G is $G = \bigcup_{k=1}^{4} G_k$.

We will call (2.6) in the sequel the τ-division of the set G.

3. Main results

Suppose $\{(\eta_i, \xi_i),\ 1 \leqslant i \leqslant n\}$ is the observed data from model (2.1). Before establishing estimators we need to introduce some additional symbols and notations.

For every positive number x $(0 < x < 1)$, we write

$$N \triangleq [n^{1/3}], \quad \delta_n \triangleq n^{-1/3}, \quad D_n(x) \triangleq \{i\colon 1 \leqslant i \leqslant n,\ |\xi_i - x| \leqslant \delta_n\}, \tag{3.1}$$

where $[\,\cdot\,]$ presents the integer part. Let n_x denote the number of elements contained in $D_n(x)$, and for convenience we write $n_{k/N}$ as n_k $(1 \leqslant k \leqslant N)$.

Now, we construct the indirect empirical wavelet and scale coefficients as follows:

$$\begin{aligned}\alpha_{j,k} &= \frac{1}{N}\sum_{i=1}^{N}\Psi_{j,k}\left(\frac{i}{N}\right)\frac{1}{n_i}\sum_{l\in D_n(i/N)}\eta_l \\ &= \frac{1}{N}\sum_{i=1}^{N}\Psi_{j,k}\left(\frac{i}{N}\right)\frac{1}{n_i}\sum_{l\in D_n(i/N)}(f(\xi_l)+\varepsilon_l) \\ &\triangleq \alpha_{j,k}^{(1)} + \alpha_{j,k}^{(2)},\end{aligned} \tag{3.2}$$

$$\begin{aligned}\beta_{j,k} &= \frac{1}{N}\sum_{i=1}^{N}\Phi_{j,k}\left(\frac{i}{N}\right)\frac{1}{n_i}\sum_{l\in D_n(i/N)}\eta_l \\ &= \frac{1}{N}\sum_{i=1}^{N}\Phi_{j,k}\left(\frac{i}{N}\right)\frac{1}{n_i}\sum_{l\in D_n(i/N)}(f(\xi_l)+\varepsilon_l) \\ &\triangleq \beta_{j,k}^{(1)} + \beta_{j,k}^{(2)}.\end{aligned} \tag{3.3}$$

Here we use $\theta_{j,k}(x)$ to denote the dialation and translation of any function $\theta(x)$:

$$\theta_{j,k}(x) = 2^{j/2}\theta(2^j x - k). \tag{3.4}$$

The aim in this section is to establish estimators for the jump number q, locations t_l $(1 \leqslant l \leqslant q)$ and jump heights h_l $(1 \leqslant l \leqslant q)$, where

$$h_l \triangleq s(t_l + 0) - s(t_l - 0) \triangleq \lim_{x\downarrow t_l} s(x) - \lim_{x\uparrow t_l} s(x). \tag{3.5}$$

The coefficients set $\{\alpha_{j,k}\}$ contains the key information about jumps of $s(x)$, i.e., t_l $(l = 1,\dots,q)$. When j is large enough and fixed, the absolute value of $\alpha_{j,k}$ at such k that $k/2^j$ near t_l is relatively larger than that of $\alpha_{j,k}$ at other k whereas $k/2^j$ is farther away from t_l. This has led us to propose estimators for the number and locations of jump points by examining the properties of the set $\{\alpha_{j,k}\}$. Based on the location estimators we can then establish jump heights by using the coefficient set $\{\beta_{j,k}\}$. The main results of this section are summarized in the following two theorems.

Theorem 3.1. *Assume* (A1)–(A3) *are true, for any fixed number* $c > 0$, $j = j_n$ *is chosen in such a way that* $j_n \to +\infty$ $(n \to +\infty)$ *and* $2^{3j_n} \leqslant N$, *where* N *is given by* (3.1). *Write*

$$G(j) = \{k\colon 2^{j/4} \leqslant k \leqslant 2^j - 2^{j/4} \text{ and } |\alpha_{j,k}| \geqslant c2^{-(13/8)j}\}. \tag{3.6}$$

If $G(j)$ *is not empty, by the* $2^{j/2}$*-division* $G(j)$ *is partitioned into* $\hat{p}$ *parts*

$$G(j) = \bigcup_{l=1}^{\hat{p}} G_l(j). \tag{3.7}$$

Define

$$\hat{q} = \begin{cases} \hat{p} & \textit{if } G(j) \textit{ is not empty}, \\ 0 & \textit{if } G(j) \textit{ is empty}. \end{cases} \tag{3.8}$$

If k_l *denotes such an element of* $G_l(j)$ *that* $|\alpha_{j,k_l}| = \max_{k \in G_l(j)}\{|\alpha_{j,k}|\}$, *set*

$$\hat{t}_l = \frac{k_l}{2^j}, \quad l = 1, \ldots, \hat{q}. \tag{3.9}$$

Then
(1) $\hat{q} = q + o_{\mathrm{p}}(1)$ $(n \to +\infty)$, *i.e.*, $\lim_{n\to\infty} P(\hat{q} = q) = 1$,
(2) $\lim_{n\to\infty} P(|\hat{t}_l - t_l| > 2^{-j/2}) = 0$ $(1 \leqslant l \leqslant q)$.

Theorem 3.2. *Under the conditions and notations of Theorem* 3.1, *write*

$$k_l^{(+)} = k_l + 2^{j/2}, \quad k_l^{(-)} = k_l - 2^{j/2} \tag{3.10}$$

set

$$\hat{h}_l = 2^{(3/2)j}(\beta_{j,k_l^{(-)}} - \beta_{j,k_l^{(+)}}). \tag{3.11}$$

Then $\hat{h}_l = h_l + o_{\mathrm{p}}(1)$ $(n \to +\infty)$, *i.e.*, *for every positive number* $\tau > 0$, *we have*

$$\lim_{n\to\infty} P(|\hat{h}_l - h_l| > \tau) = 0 \quad (1 \leqslant l \leqslant q). \tag{3.12}$$

4. Numerical simulations

In this section we carry on some Monte Carlo simulations to investigate the validity of the theorems in the previous section. It should be pointed out that Theorem 3.1 cannot be applied directly due to the constant c. But Theorem 3.1 does provide the key information which can be used in practical implementation. Theorem 3.1 asserts that the coefficients near jumps are significantly larger than those a little far away from them across fine scales. Based on this fact, we determine the jump points, in practice, by observing those coefficients whose absolute values are prominently large at fine scales. We only focus on the number and locations of jump points in our simulations.

The simulation model is as follows:

$$\eta_i = f(\xi_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{4.1}$$

where the regression function $f(t)$ (when $t > 1$ or < 0 define $f(t) = 0$) is given by either (4.2) or (4.3):

$$f(t) = \begin{cases} 0.5\cos(2\pi t), & 0 \leqslant t < 0.4, \\ 0.5\cos(4\pi t) + 0.5\cos(0.8\pi) - 0.5\cos(1.6\pi), & 0.4 \leqslant t \leqslant 1, \end{cases} \tag{4.2}$$

$$f(t) = \begin{cases} -2t, & 0 \leqslant t < 0.3, \\ 3t - 1.5, & 0.3 \leqslant t < 0.7, \\ -2t + 2, & 0.7 \leqslant t \leqslant 1. \end{cases} \tag{4.3}$$

$\{\varepsilon_i\}$ is the AR(1) noise generated by the following mechanism:

$$\varepsilon_i = 0.5\varepsilon_{i-1} + \sigma e_i, \tag{4.4}$$

in which $\{e_i\}$ is an i.i.d. N(0,1) sequence. And the random designed sampling $\{\xi_i\}$ is chosen as i.i.d. N(0.5, 0.25), independent with $\{\varepsilon_i\}$.

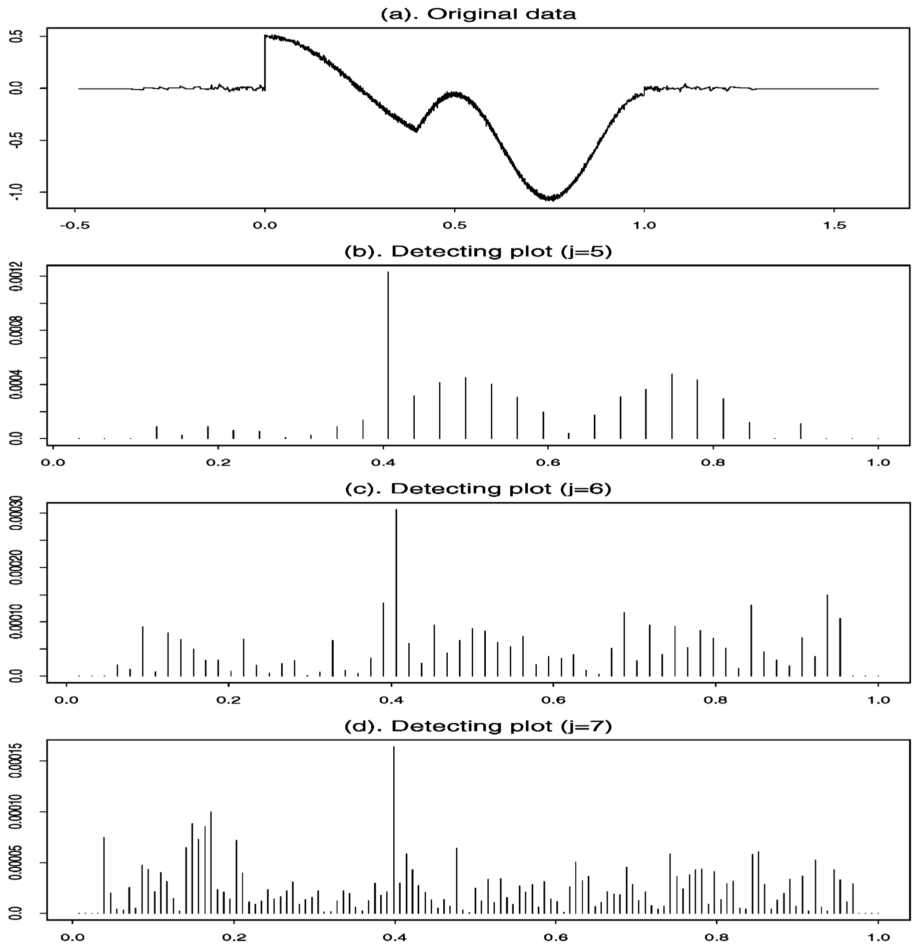

Fig. 1.

The wavelet $\psi(t)$ and the scale function $\phi(t)$ are given as follows:

$$\psi(t) = \begin{cases} t^4 - 2t^3 + t^2, & 0 \leqslant t \leqslant 1, \\ -t^4 - 2t^3 - t^2, & -1 \leqslant t < 0, \\ 0, & |t| > 1, \end{cases} \tag{4.5}$$

$$\phi(t) = \begin{cases} \frac{15}{16}(t^4 - 2t^2 + 1), & -1 \leqslant t \leqslant 1, \\ 0, & |t| > 1. \end{cases} \tag{4.6}$$

It is easily verified that $\psi(t)$ and $\phi(t)$ satisfy assumption (A3) Consequently, the “indirect” wavelet $\Psi(t)$ and scale function $\Phi(t)$ are obtained from (2.3).

Fig. 1 displays a typical simulation result for model (4.1) with regression function (4.2), where $\sigma = 0.01$ and data length $n = 4096$. In which (a) is the plot of data, (b), (c) and (d) are plots (we will call it “detecting plot” for short) of the absolute indirect empirical wavelet coefficients ($|\alpha_{j,k}|$) at resolution levels $j = 5, 6$ and 7, respectively. From Fig. 1 we find that the coefficients with significantly large values across levels

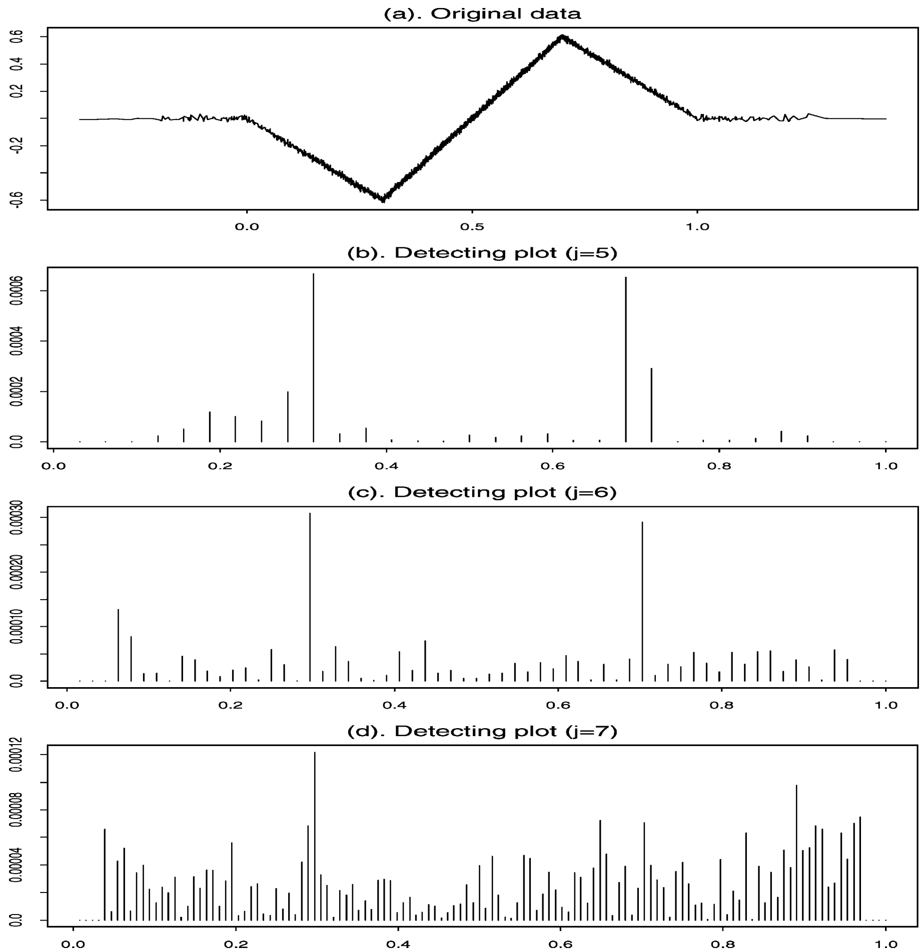

Fig. 2.

$j = 5, 6, 7$ correspond to almost the same location, near 0.4 which is the real jump. More precisely, at level $j = 5$ the maximum value of coefficients is that located at $13/2^5 = 0.406$, and that located at $26/2^6 = 0.406$ and $51/2^7 = 0.398$ at levels $j = 6$ and 7, respectively. Furthermore, Fig. 1 indicates that there is only one jump point, which coincides the real situation.

Similarly, a typical simulation result is shown in Fig. 2, where the regression function is chosen as (4.3) and with $\sigma = 0.01$ data length $n = 4096$. It is easy to see that detecting plots with $j = 5, 6$ draw comparatively the same conclusion: there are two jumps with locations near 0.3 and 0.7. At level $j = 5$ the jump locations are $10/2^5 = 0.313$ and $22/2^5 = 0.688$, and at level $j = 6$ the locations are $19/2^6 = 0.299$ and $45/2^6 = 0.703$. But at level $j = 7$ it seems that no clear result can be drawn from the detecting plot.

In our simulation studies we found that the noise level (σ) affected detecting plots badly. When noise levels increase the detecting plot become worse (i.e., dimmer to draw conclusions).

In practical implementation, we agree with Wang (1995). The wavelet coefficients should be checked at several resolution levels. Under common data lengths, resolution levels $j = 5, 6, 7$ are usually considered, and only when different resolution levels show similar results can inferences be drawn. As for jump heights, we found in simulations that the estimators were quite unstable so the results are not given here.

Appendix

This section is devoted to the proofs of Theorems 3.1 and 3.2, the main results of this paper. We proceed in stages through a series of lemmas to the proofs. Before presenting lemmas, we introduce some necessary symbols and notations which are needed in this section. For arbitrary $x \in (0,1)$, $\delta > 0$ and positive integer j, write

$$I(x;\delta,j) \triangleq \left\{k\colon \left|\frac{k}{2^j}-x\right| \leqslant \frac{\delta}{2},\ k=0,1,2,\dots,2^j-1\right\},$$

$$\Gamma^{(+)}(x;\delta,j) \triangleq \left\{k\colon x+\frac{\delta}{2} \leqslant \frac{k}{2^j} \leqslant x+\frac{3}{2}\delta,\ k=0,1,2,\dots,2^j-1\right\},$$

$$\Gamma^{(-)}(x;\delta,j) \triangleq \left\{k\colon x-\frac{3}{2}\delta \leqslant \frac{k}{2^j} \leqslant x-\frac{\delta}{2},\ k=0,1,2,\dots,2^j-1\right\}. \tag{A.1}$$

Lemma A.1. *Assume $\{\xi_i\}$ satisfies Assumption* (A2), $D_n(x)$ *and* n_x *are defined as* (3.1). *Then there exists a positive number $c_0 > 0$ such that*

$$\lim_{n\to\infty} P(\Omega_n) = 1, \tag{A.2}$$

where $\Omega_n = \{n_x \geqslant c_0 n^{2/3},\ \forall x \in [0,1]\}$.

This lemma is a slight modification of Lemma 7 in Truong and Stone (1992), so its proof is omitted here. Lemma A.1 shows that when sample length n is very large there will be enough ξ_i's in the neighborhood of every point $x \in [0,1]$.

Lemma A.2. *Suppose $h(x)$ is a function defined on* $[0,1]$ *with bounded variation, m is a positive integer. Then we have*

$$\left|\int_0^1 h(x)\,\mathrm{d}x - \frac{1}{m}\sum_{i=1}^{m} h\left(\frac{i}{m}\right)\right| \leqslant \frac{v}{m}, \tag{A.3}$$

where $v = V_0^1(h)$ *is the total variation of* $h(x)$ *on* $[0,1]$.

Proof. This is just Lemma P5.1 in Brillinger (1981). □

Lemma A.3. *Assume* (A1) *and* (A3) *are true. With the notations of* (3.4) *and* (A.1), *there exist positive constants* c_1, c_2 *so that when j is large enough the following are valid*:

(1)

$$\inf_{k\in\bigcup_{l=1}^{q} I(t_l;2^{-j},j)} \left\{\left|\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x\right|\right\} \geqslant c_1 2^{-j/2}, \tag{A.4}$$

(2)

$$\sup_{\substack{k\notin\bigcup_{l=1}^{q} I(t_l;2^{-j/2},j)\\ 2^{j/4}\leqslant k\leqslant 2^j-2^{j/4}}} \left\{\left|\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x\right|\right\} \leqslant c_2 2^{-3j/2}. \tag{A.5}$$

Proof. By variable change in integration, we have

$$\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x = 2^{-j/2}\int_{-k}^{2^j-k} s\left(\frac{x+k}{2^j}\right)\psi(x)\,\mathrm{d}x.$$

(1) If $k \in \bigcup_{l=1}^{q} I(t_l; 2^{-j}, j)$, say, $k \in I(t_l; 2^{-j}, j)$ i.e., $|k/2^j - t_l| \leqslant 2^{-j}/2$. When j is sufficiently large, then $[-k, 2^j - k] \supset [-\sigma, \sigma]$, with (2.2) and (2) we have

$$\begin{aligned}\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x &= 2^{-j/2}\int_{-\sigma}^{\sigma} s\left(\frac{x+k}{2^j}\right)\psi(x)\,\mathrm{d}x\\ &= 2^{-j/2}\int_{-\sigma}^{2^j t_l-k} s\left(\frac{x+k}{2^j}\right)\psi(x)\,\mathrm{d}x + 2^{-j/2}\int_{2^j t_l-k}^{\sigma} s\left(\frac{x+k}{2^j}\right)\psi(x)\,\mathrm{d}x\\ &= 2^{-j/2}\,(s(t_l+0)-s(t_l-0))\int_{2^j t_l-k}^{\sigma}\psi(x)\,\mathrm{d}x + O(2^{-(3/2)j}).\end{aligned}$$

With (2.5), we deduce that

$$\left|\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x\right| \geqslant c_1 2^{-j/2} \quad (j \text{ sufficiently large}).$$

For instance, c_1 can be chosen as $c_1 = (b_0/2)\inf_{1\leqslant l\leqslant q}|s(t_l+0)-s(t_l-0)| > 0$. So (A.4) is proved.

(2) If $k \notin \bigcup_{l=1}^{q} I(t_l; 2^{-j/2}, j)$ and $2^{j/4} \leqslant k \leqslant 2^j - 2^{j/4}$, when j is large enough, then there are no jump points between $(x+k)/2^j$ and $k/2^j$ for every x $(-\sigma \leqslant x \leqslant \sigma)$.

$$\begin{aligned}\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x &= 2^{-j/2}\int_{-\sigma}^{\sigma} s\left(\frac{x+k}{2^j}\right)\psi(x)\,\mathrm{d}x\\ &= 2^{-j/2}\int_{-\sigma}^{\sigma}\left(s\left(\frac{x+k}{2^j}\right) - s\left(\frac{k}{2^j}\right)\right)\psi(x)\,\mathrm{d}x\\ &= 2^{-3j/2}\int_{-\sigma}^{\sigma} s'(\theta)\psi(x)\,\mathrm{d}x.\end{aligned}$$

Set $c_2 = (\int_{-\sigma}^{\sigma}|x\psi(x)|\,\mathrm{d}x)(\sup_{x\neq t_l}|(\mathrm{d}/\mathrm{d}x)s(x)|) > 0$, then $|\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x| \leqslant c_2 2^{-3j/2}$. So, the lemma is established. □

Lemma A.4. *Assume* (A1) *and* (A3) *are true. With the notations of* (A.1), *when j is large enough we have*

(1) *If* $k \in \Gamma^{(+)}(t_l; 2^{-j/2}, j)$ $(1 \leqslant l \leqslant q)$, *the equation below is valid uniformly* (*with respect to* k).

$$\int_0^1 s(x)\phi_{j,k}(x)\,\mathrm{d}x = 2^{-j/2}s(t_l+0) + O(2^{-j}). \tag{A.8}$$

(2) *If* $k \in \Gamma^{(-)}(t_l; 2^{-j/2}, j)$ $(1 \leqslant l \leqslant q)$, *the equation below is valid uniformly* (*with respect to* k).

$$\int_0^1 s(x)\phi_{j,k}(x)\,\mathrm{d}x = 2^{-j/2}s(t_l-0) + O(2^{-j}). \tag{A.9}$$

Proof. $k \in \Gamma^{(+)}(t_l; 2^{-j/2,j})$ means $t_l + (\frac{1}{2})2^{-j/2} \leqslant k/2^j \leqslant t_l + (\frac{3}{2})2^{-j/2}$. When j is sufficiently large, we have

$$\begin{aligned}\int_0^1 s(x)\phi_{j,k}(x)\,\mathrm{d}x &= 2^{-j/2}\int_{-k}^{2^j-k} s\left(\frac{x+k}{2^j}\right)\phi(x)\,\mathrm{d}x\\ &= 2^{-j/2}\int_{-\sigma}^{\sigma}\left(s\left(\frac{x+k}{2^j}\right) - s(t_l+0)\right)\phi(x)\,\mathrm{d}x + 2^{-j/2}s(t_l+0)\\ &= 2^{-j/2}s(t_l+0) + O(2^{-j}).\end{aligned}$$

In the above we have used the fact that $\int_{-\sigma}^{\sigma} \phi(x)\,\mathrm{d}x = 1$. Therefore (A.8) is valid and (A.9) can be proved exactly in the same way. □

Lemma A.5. *Assume* (A1)–(A3) *are true. With the notations of* (3.2) *and* (3.3), $j = j_n$ *is chosen in such a way that* $j_n \to +\infty$ $(n \to +\infty)$ *and* $2^{3j_n} \leqslant N$. *Then, there exist positive constants* d_1, d_2 *such that the following are valid*:

$$\lim_{n\to\infty} P\left(\inf_{k\in\bigcup_{l=1}^{q} I(t_l;2^{-j},j)} |\alpha_{j,k}^{(1)}| \geqslant d_1 2^{-(3/2)j}\right) = 1, \tag{A.10}$$

$$\lim_{n\to\infty} P\left(\sup_{k\notin\bigcup_{l=1}^{q} I(t_l;2^{-j/2},j),\ 2^{j/4}\leqslant k\leqslant 2^j-2^{j/4}} |\alpha_{j,k}^{(1)}| \leqslant d_2 2^{-(5/2)j}\right) = 1. \tag{A.11}$$

Proof. Lemma A.1 and the notations of (3.2) imply that on Ω_n,

$$\begin{aligned}
\alpha_{j,k}^{(1)} &= \frac{1}{N}\sum_{i=1}^{N}\Psi_{j,k}\left(\frac{i}{N}\right)\frac{1}{n_i}\sum_{l\in D_n(i/N)} f(\xi_l)\\
&= \frac{1}{N}\sum_{i=1}^{N}\Psi_{j,k}\left(\frac{i}{N}\right) f\left(\frac{i}{N}\right) + \frac{1}{N}\sum_{i=1}^{N}\Psi_{j,k}\left(\frac{i}{N}\right)\frac{1}{n_i}\sum_{l\in D_n(i/N)}\left(f(\xi_l) - f\left(\frac{i}{N}\right)\right).
\end{aligned}$$

With Lemma A.2, the notations of (3.1) and the finite support of $\psi(x)$, we have

$$\begin{aligned}
\alpha_{j,k}^{(1)} &= \int_0^1 f(x)\Psi_{j,k}(x)\,\mathrm{d}x + O\left(\frac{2^{j/2}}{N}\right)\\
&= \int_0^1 f(x)(2^{-j}\psi_{j,k}(x))'\,\mathrm{d}x + O\left(\frac{2^{j/2}}{N}\right)\\
&= -2^{-j}\int_0^1 s(x)\psi_{j,k}(x)\,\mathrm{d}x + O\left(\frac{2^{j/2}}{N}\right).
\end{aligned}$$

Lemma A.3 and the above equations then imply that

$$\inf_{k\in\bigcup_{l=1}^{q} I(t_l;2^{-j},j)} |\alpha_{j,k}^{(1)}| \geqslant c_1 2^{-(3/2)j} - \left|O\left(\frac{2^{j/2}}{N}\right)\right| = c_1 2^{-(3/2)j} - \left|O(2^{-(5/2)j})\right| \geqslant d_1 2^{-(3/2)j},$$

$$\sup_{\substack{k\notin\bigcup_{l=1}^{q} I(t_l;2^{-j/2},j)\\ 2^{j/4}\leqslant k\leqslant 2^j-2^{j/4}}} |\alpha_{j,k}^{(1)}| \leqslant c_2 2^{-(5/2)j} + \left|O\left(\frac{2^{j/2}}{N}\right)\right| \leqslant c_2 2^{-(5/2)j} + \left|O\left(2^{-(5/2)j}\right)\right| \leqslant d_2 2^{-(5/2)j}.$$

The above two relations and Lemma A.1 yield the validity of (A.10) and (A.11). □

Lemma A.6. *Under the conditions of Lemma* A.5, *there exists positive constant* c_3 *such that the following are true*:

$$\lim_{n\to\infty} P\left(\bigcap_{l=1}^{q}\bigcap_{k\in\Gamma^{(+)}(t_l;2^{-j/2},j)} |\beta_{j,k}^{(1)} + 2^{-(3/2)j}s(t_l+0)| \leqslant c_3 2^{-2j}\right) = 1, \tag{A.12}$$

$$\lim_{n\to\infty} P\left(\bigcap_{l=1}^{q}\bigcap_{k\in\Gamma^{(-)}(t_l;2^{-j/2},j)} |\beta_{j,k}^{(1)} + 2^{-(3/2)j}s(t_l-0)| \leqslant c_3 2^{-2j}\right) = 1. \tag{A.13}$$

Proof. With Lemma A.4 and following the same arguments in the proof of Lemma A.5, there exists a positive constant $c_3 > 0$, when n is quite large the following relation holds:

$$\left\{\bigcap_{l=1}^{q}\bigcap_{k\in\Gamma^{(+)}(t_l;2^{-j/2},j)} |\beta_{j,k}^{(1)} + 2^{-(3/2)j}s(t_l+0)| \leqslant c_3 2^{-2j}\right\} \supset \Omega_n.$$

Lemma A.1 yields (A.12) and (A.13) can be proved similarly. □

Lemma A.7. *Assume* (A2) *and* (A3) *are true. With notations of* (3.2) *and* (3.3) $j = j_n$ *is chosen in such a way that* $j_n \to +\infty$ $(n \to +\infty)$ *and* $2^{3j_n} \leqslant N$. *Then, we have*

$$\max_{0\leqslant k\leqslant 2^j-1}(|\alpha_{j,k}^{(2)}|, |\beta_{j,k}^{(2)}|) = o_{\rm p}(2^{-(7/4)j}) \quad (n\to\infty). \tag{A.14}$$

Proof. With (3.2) we have $\alpha_{j,k}^{(2)} = (1/N)\sum_{i=1}^{N}\Psi_{j,k}(i/N)1/n_i\sum_{l\in D_n(i/N)}\varepsilon_l$. With notations of (3.1), (3.2), and the boundness of the p.d.f. $g(x)$ of ξ_1, there exists a positive constant $c_4 > 0$ such that

$$\begin{aligned}
E\left(\sum_{l\in D_n(i/N)}\varepsilon_l\right)^2 &= E\left(\sum_{l=1}^{n} I_{\{|\xi_l - i/N|\leqslant\delta_n\}}\varepsilon_l\right)^2 \\
&= \sum_{l=1}^{n}\sum_{m=1}^{n} E(I_{\{|\xi_l-i/N|\leqslant\delta_n\}}I_{\{|\xi_m-i/N|\leqslant\delta_n\}})E(\varepsilon_l\varepsilon_m) \\
&\leqslant \sum_{l=1}^{n}\sum_{m=1}^{n} E(I_{\{|\xi_l-i/N|\leqslant\delta_n\}})|\gamma_{l-m}| \\
&= n\left(\sum_{m=-\infty}^{\infty}|\gamma_m|\right)P(|\xi_1 - i/N| \leqslant \delta_n) \leqslant c_4 n^{2/3}.
\end{aligned}$$

Noting that $n_i \geqslant c_0 n^{2/3}$ $(i = 1,\dots,N)$ on set Ω_n and the finite support of $\Psi(x)$, so

$$\begin{aligned}
E(I_{\Omega_n}\alpha_{j,k}^{(2)})^2 &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{N}\Psi_{j,k}\left(\frac{i}{N}\right)\Psi_{j,k}\left(\frac{m}{N}\right)E\left(I_{\Omega_n}\frac{1}{n_i}\sum_{l\in D_n(i/N)}\varepsilon_l\frac{1}{n_m}\sum_{l\in D_n(m/N)}\varepsilon_l\right) \\
&\leqslant \left(\frac{1}{c_0 n^{2/3}}\right)^2\frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{N}\left|\Psi_{j,k}\left(\frac{i}{N}\right)\Psi_{j,k}\left(\frac{m}{N}\right)\right| \\
&\quad \left(E\left(\sum_{l\in D_n(i/N)}\varepsilon_l\right)^2\right)^{1/2}\left(E\left(\sum_{l\in D_n(m/N)}\varepsilon_l\right)^2\right)^{1/2}
\end{aligned}$$

$$\leqslant \left(\frac{1}{c_0 n^{2/3}}\right)^2 \frac{1}{N^2}\sum_{i=1}^{N}\sum_{m=1}^{N}\left|\Psi_{j,k}\left(\frac{i}{N}\right)\Psi_{j,k}\left(\frac{m}{N}\right)\right| c_4 n^{2/3}$$
$$= \left(\frac{c_4}{c_0^2} n^{-2/3}\right)\frac{1}{N^2}\left(\sum_{i=1}^{N}\left|\Psi_{j,k}\left(\frac{i}{N}\right)\right|\right)^2 = O(2^{-j/2}n^{-2/3}).$$

So, there exists a positive constant $c_5 > 0$ such that $\max_{0\leqslant k\leqslant 2^j-1}(E(I_{\Omega_n}\alpha_{j,k}^{(2)})^2) \leqslant c_5 n^{-2/3}$.

For arbitrary $\tau > 0$,

$$P\left(\max_{0\leqslant k\leqslant 2^j-1}(2^{(7/4)j}|\alpha_{j,k}^{(2)}|) \geqslant \tau\right) \leqslant P(\Omega_n^{\rm c}) + P\left(\max_{0\leqslant k\leqslant 2^j-1}(2^{(7/4)j}|\alpha_{j,k}^{(2)} I_{\Omega_n}|) \geqslant \tau\right)$$
$$\leqslant P(\Omega_n^{\rm c}) + \sum_{k=0}^{2^j-1}\frac{2^{(7/2)j}E(I_{\Omega_n}\alpha_{j,k}^{(2)})^2}{\tau^2}$$
$$\leqslant P(\Omega_n^{\rm c}) + \frac{c_5}{\tau^2}2^{(9/2)j}2^{-6j} \to 0 \quad (n\to\infty).$$

This means $\max_{0\leqslant k\leqslant 2^j-1}(|\alpha_{j,k}^{(2)}|) = o_{\rm p}(2^{-(7/4)j})$ $(n\to\infty)$. Similarly we can verify that $\max_{0\leqslant k\leqslant 2^j-1}(|\beta_{j,k}^{(2)}|) = o_{\rm p}(2^{-(7/4)j})$ $(n\to\infty)$. $\square$

Proof of Theorem 3.1. We introduce several events as follows:

$$B_1 = \left\{\inf_{k\in\bigcup_{l=1}^q I(t_l;2^{-j},j)}|\alpha_{j,k}^{(1)}| \geqslant d_1 2^{-(3/2)j}\right\},$$
$$B_2 = \left\{\sup_{k\notin\bigcup_{l=1}^q I(t_l;2^{-j/2},j),\ 2^{j/4}\leqslant k\leqslant 2^j-2^{j/4}}|\alpha_{j,k}^{(1)}| \leqslant d_2 2^{-(5/2)j}\right\},$$
$$B_3 = \left\{\max_{0\leqslant k\leqslant 2^j-1}(|\alpha_{j,k}^{(2)}|, |\beta_{j,k}^{(2)}|) < 2^{-(7/4)j}\right\}.$$

From Lemmas A.5 and A.7, it is seen that $\lim_{n\to\infty}P(B_1B_2B_3) = 1$. On $B_1B_2B_3$, $k\in\bigcup_{l=1}^q I(t_l;2^{-j},j)$ implies $|\alpha_{j,k}| = |\alpha_{j,k}^{(1)} + \alpha_{j,k}^{(2)}| \geqslant d_1 2^{-(3/2)j} - 2^{-(7/4)j} > c2^{-(13/8)j}$, and similarly, $k\notin\bigcup_{l=1}^q I(t_l;2^{-j/2},j)$ and $2^{j/4}\leqslant k\leqslant 2^j-2^{j/4}$ imply $|\alpha_{j,k}| < c2^{-(13/8)j}$. So

$$G(j)\neq\emptyset \quad\text{and}\quad \bigcup_{l=1}^q I(t_l;2^{-j},j)\subset G(j)\subset\bigcup_{l=1}^q I(t_l;2^{-j/2},j).$$

By the $2^{j/2}$-division of $G(j)$, it is easy to see that $G(j)$ is partitioned exactly into q parts:

$$G(j) = \bigcup_{l=1}^q G_l(j) \quad\text{and}\quad I(t_l;2^{-j},j)\subset G_l(j)\subset I(t_l;2^{-j/2},j) \quad (1\leqslant l\leqslant q).$$

Therefore, we have that $\hat{p} = q$ and $|\hat{t}_l - t_l| \leqslant 2^{-j/2}$, and so

$$\{\hat{q} = q\}\bigcap_{l=1}^q\{|\hat{t}_l - t_l| \leqslant 2^{-j/2}\} \supset B_1B_2B_3.$$

The above relation asserts that

$$\lim_{n\to\infty} P(\hat{q}=q)=1, \quad \lim_{n\to\infty} P(|\hat{t}_l - t_l| > 2^{-j/2})=0 \quad (1 \leqslant l \leqslant q).$$

That finishes the proof. $\square$

Proof of Theorem 3.2. We introduce several events as follows:

$$B_4=\left\{\bigcap_{l=1}^{q}\bigcap_{k\in\Gamma^{(+)}(t_l;2^{-j/2},j)} |\beta_{j,k}^{(1)}+2^{-(3/2)j}s(t_l+0)| \leqslant c_3 2^{-2j}\right\},$$

$$B_5=\left\{\bigcap_{l=1}^{q}\bigcap_{k\in\Gamma^{(-)}(t_l;2^{-j/2},j)} |\beta_{j,k}^{(1)}+2^{-(3/2)j}s(t_l-0)| \leqslant c_3 2^{-2j}\right\},$$

$$B_6=\left\{\max_{0\leqslant k\leqslant 2^j-1}(|\alpha_{j,k}^{(2)}|,|\beta_{j,k}^{(2)}|) < 2^{-(7/4)j}\right\}.$$

On $B_4B_5B_6$, $k\in\Gamma^{(+)}(t_l;2^{-j/2},j)$ implies

$$|\beta_{j,k}+2^{-(3/2)j}s(t_l+0)| = |\beta_{j,k}^{(1)}+2^{-(3/2)j}s(t_l+0)+\beta_{j,k}^{(2)}| \leqslant c_3 2^{-2j}+2^{-(7/4)j} < 2^{-(13/8)j}.$$

Similarly $k\in\Gamma^{(-)}(t_l;2^{-j/2},j)$ implies $|\beta_{j,k}+2^{-(3/2)j}s(t_l-0)|<2^{-(13/8)j}$. On $B_1B_2B_3$ and when n is very large, $\hat{q}=q$. With notation of (3.10)

$$\frac{k_l^{(+)}}{2^j} \geqslant t_l+\left(\frac{1}{2}\right)2^{-j/2}, \quad \frac{k_l^{(+)}}{2^j} \leqslant t_l+\left(\frac{3}{2}\right)2^{-j/2}.$$

That means $k_l^{(+)}\in\Gamma^{(+)}(t_l;2^{-j/2},j)$. Similarly we can get $k_l^{(-)}\in\Gamma^{(-)}(t_l;2^{-j/2},j)$. Also, on event $\bigcap_{k=1}^6 B_k$ and when n sufficiently large, with (3.11)

$$\begin{aligned}\hat{h}_l &= 2^{(3/2)j}(\beta_{j,k_l^{(-)}}-\beta_{j,k_l^{(+)}})\\ &= h_l - 2^{(3/2)j}(\beta_{j,k_l^{(+)}}+2^{-(3/2)j}s(t_l+0))+2^{(3/2)j}(\beta_{j,k_l^{(-)}}+2^{-(3/2)j}s(t_l-0)).\end{aligned}$$

So

$$|\hat{h}_l-h_l| \leqslant 2^{(3/2)j}|\beta_{j,k_l^{(+)}}+2^{-(3/2)j}s(t_l+0)|+2^{(3/2)j}|\beta_{j,k_l^{(-)}}+2^{-(3/2)j}s(t_l-0)| < 2^{-j/4}.$$

That means $\{|\hat{h}_l-h_l|<2^{-j/4}\}\supset\bigcap_{k=1}^6 B_k$. With Lemmas A.5–A.7 we see that $\lim_{n\to\infty}P(\bigcap_{k=1}^6 B_k)=1$, So

$$\lim_{n\to\infty} P(|\hat{h}_l-h_l|>2^{-j/4})=0 \quad (1\leqslant l\leqslant q).$$

That ends the proof. $\square$

References

Basseville, M., 1988. Detecting changes in signals and systems—a survey. Automatica 24, 309–326.

Basseville, M., Nikiforov, I.V., 1993. Detection of Abrupt Changes: Theory and Application. Prentice Hall, Englewood Cliffs, NJ.

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

Johnstone, I.M., 1999. Wavelet shrinkage for correlated data and inverse problem: adaptivity results. Statist. Sinica 9, 51–83.

Johnstone, I.M., Silverman, B.W., 1997. Wavelet threshold estimators for data with correlated noise. J. Roy. Statist. Soc. B 59 (2), 319–351.

Li, Y., Xie, Z., 1999. Detection of jumps by wavelets in a nonlinear autoregressive model. Acta Math. Scientia 19 (3), 261–271.

Li, Y., Xie, Z., 2000. The wavelet detection of the jumps and the cusp points of a regression function. Acta Appl. Math. Sinica 16 (3), 283–291.
Müller, H.G., 1992. Change-points in nonparametric regression analysis. Ann. Statist. 20, 737–761.
Ogden, T., Parzen, E., 1996. Data dependent wavelet thresholding in nonparametric regression with change-point applications. Comput. Statist. Data Anal. 22, 53–70.
Truong, Y.K., Stone, C.J., 1992. Nonparametric function estimation involving time series. Ann. Statist. 20, 77–97.
Wang, Y., 1995. Jump and sharp cusp detection by wavelets. Biometrika 82, 385–397.
Wang, Y., 1999. Change-points via wavelets for indirect data. Statist. Sinica 9, 103–117.
Wu, J.S., Chu, C.K., 1993. Kernel-type estimators of jump points and values of regression function. Ann. Statist. 21, 1545–1566.
Yin, Y.Q., 1988. Detection of the number, locations and magnitudes of jumps. Comm. Statist. C4, 445–455.