Jump Detection In A Regression Curve And Its Derivative

Jong-Hoon Joo and Peihua Qiu
School of Statistics
University of Minnesota
313 Ford Hall
224 Church St. S.E.
Minneapolis, MN 55455

Abstract

Curve estimation from observed noisy data has broad applications. In certain applications, the underlying regression curve may have singularities, including jumps and roofs/valleys (i.e., jumps in the first order derivative of the regression curve), at some unknown positions, representing structural changes of the related process. Detection of such singularities is important for understanding the structural changes. In the literature, a number of jump detection procedures have been proposed, most of which are based on estimation of the (one-sided) first order derivatives of the true regression curve. In this paper, motivated by certain related research in image processing, we propose an alternative jump detection procedure. Besides the first order derivatives, we suggest using helpful information about jumps in the second order derivatives as well. Theoretical justifications and numerical studies show that this jump detector works well in applications. This procedure is then extended for detecting roofs/valleys of the regression curve. Using detected jumps/roofs/valleys, a curve estimation procedure is also proposed, which can preserve possible jumps/roofs/valleys when removing noise. *Note that the printed paper only provides outlines of proofs of several theorems in the paper. Complete proofs of the theorems are available online as supplemental materials.*

Key Words: Curve estimation; Derivatives; Discontinuities; Function estimation; Jump preserving; Jumps; Local polynomial kernel smoothing; Nonparametric regression; Roof edges.

1 Introduction

Curve estimation from observed noisy data has broad applications. Nonparametric regression analysis provides a major statistical tool for handling this problem (cf., e.g., Fan and Gijbels 1996). In some applications, the underlying regression curve may have jumps and roofs/valleys

(i.e., jumps in the first-order derivative of the regression curve) at certain unknown positions, representing structural changes of the related process. For instance, a quality index of a product could have a downward or upward shift after an unknown time point when the production line goes out-of-control (cf., e.g., Hawkins and Olwell 1998). It has been demonstrated that the sea-level pressures observed by a Bombay weather station in India have a jump around the year 1960 (Qiu and Yandell 1998). In these examples, jumps/roofs/valleys are an important part of the related process. This paper focuses on accurate detection of them, and on estimation of the regression curve with these important structures preserved.

In the literature, several jump detection procedures have been proposed. Some of them are under the assumption that the number of jumps is known (cf., e.g., Shiau 1987, Qiu *et al.* 1991, Müller 1992, Wu and Chu 1993, Eubank and Speckman 1994, Wang 1995, Loader 1996, Gijbels *et al.* 1999, Grégoire and Hamrouni 2002, Müller 2002). Some others allow the number of jumps to be unknown (cf., e.g., Hall and Titterington 1992, Qiu 1994, Qiu and Yandell 1998). Wu and Zhao (2007) discussed the jump detection problem in the context of time series analysis. Most of these procedures detect jumps based on estimation of the first order derivative (or, the difference between the right-sided and left-sided limits) of the underlying regression curve at a given point, using the properties that the first order derivative at a given point would be relatively large if that point is a jump point and relatively small if it is a certain distance away from any jump points. See Chapter 3 in Qiu (2005) for a detailed introduction of these procedures.

When detecting jumps in regression curves, availability of prior information about the number of jumps plays an important role. When the number of jumps is known, jump detection becomes easier to handle. For instance, if we know that there is a single jump point in the entire design interval, then we can simply locate the maximizer of a jump detection criterion for estimating the jump location. However, in many applications, it is difficult to know the number of jumps beforehand. In such cases, a typical jump detector (cf., e.g., Hall and Titterington 1992, Qiu 1994, Qiu and Yandell 1998) would flag a design point as a detected jump point if the value of the related jump detection criterion is larger than a threshold value. By doing so, there would be deceptive jumps detected. Some deceptive jumps are located around true jumps, due to the fact that the related jump detection criterion would have large values at some (instead of just one) design points close to a true jump point and they could be detected simultaneously. Other deceptive jumps are scattered in the entire design space, due to the nature of thresholding, which is similar to type-I

errors in the hypothesis testing context. Although a modification step is often adopted in a typical jump detector mentioned above to delete these deceptive jumps, such modifications are often ad hoc in nature.

In the image processing literature, besides first order derivatives of the image intensity function, certain edge detectors also use estimated second order derivatives for detecting edges (e.g., Marr and Hildreth 1980, Clark 1989, Fleck 1992, Sun and Qiu 2007). A major reasoning behind these methods is that the directional second order derivative in the direction orthogonal to an edge segment would have the following *zero-crossing* properties: (i) it is zero at the edge location, and (ii) it changes its sign from positive to negative when the image intensity jumps from a lower level to a higher level across the edge segment. See Figure 1 for a demonstration. Compared to the first order derivatives, it is found that edge detectors based on the second order derivatives would have better localization in their detected edges (i.e., their detected edges would be thinner). For detailed discussion about this topic, see Chapter 6 in Qiu (2005).

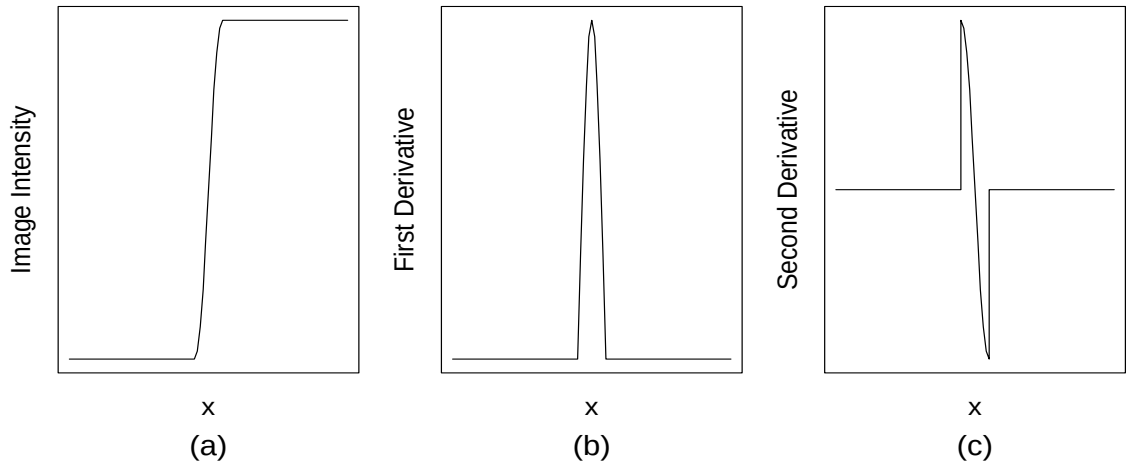

Figure 1: (a) 1-D profile of an image intensity function around an edge. (b) First-order derivative of the 1-D profile. (c) Second-order derivative of the 1-D profile.

Motivated by the above mentioned research in image processing, in this paper, we demonstrate that the jump detection problem when the number of jumps is unknown can also be handled efficiently if we combine useful information about the jumps in both the first order and the second order derivatives of the regression curve. Compared the 1-dimensional (1-D) jump detection problem considered here and the 2-dimensional (2-D) edge detection problem considered in image processing, we notice some substantial differences between the two problems. For instance, in edge detection, detected edge pixels are often scattered in the whole design space, and the total number

of edge pixels is usually not our major concern. But, in the current 1-D problem, the number of jumps is an important parameter to estimate, besides the jump positions and jump sizes. Also, in 1-D cases, the right and left sides of a given point in the design interval is well defined, and thus jump detection should make use of one-sided features of the underlying regression curve, which are not readily available in 2-D problems. In this paper, besides detection of jumps, we also discuss detection of roofs/valleys and estimation of regression curves with possible jumps/roofs/valleys preserved. The corresponding problems in 2-D image analysis are still under research and have not been well addressed yet, due to many challenges involved. See Qiu (2007) for related discussion. In the literature on 1-D regression curve estimation, there are a number of existing procedures for estimating curves with possible jumps preserved (cf., e.g., McDonald and Owen 1992, Chu *et al.* 1998, Qiu 2003, Gijbels *et al.* 2007). However, we are not aware of any existing curve estimation procedures which can also accommodate potential roofs/valleys. Theoretically speaking, our proposed approach can be generalized for detecting/preserving jumps in derivatives of any order. In practice, jumps in the second order or higher order derivatives are hardly visible. So, they are not discussed here.

We can find some existing research related to the jump/roof/valley detection problem considered here. For instance, Marron and co-authors suggest a graphical approach called SiZer for displaying potential structures in curves, such as peaks and valleys, using a SiZer map (Chaudhuri and Marron 1999). That map is constructed using estimated first order derivative of the regression curve alone. It displays regions in which the first order derivative is significantly positive, significantly negative, and insignificant from 0, respectively, and traces these regions when the bandwidth used in their method changes. Kim and Marron (2006) show that SiZer is also helpful in displaying jumps in curves; but, they admit that "This type of jump detection is intended as exploratory." They suggest using a conventional jump detection approach to further investigate the existence of jumps after jumps are found by SiZer. Another research that is closely related to the topic of this paper concerns estimation of a jump point in a parametric regression curve when the sampling scheme is adaptive (Lan *et al.* 2009). Lan *et al.* show that their proposed multistage procedure has certain good properties.

The remaining part of the paper is organized as follows. Our proposed jump detection procedure is introduced in detail in Section 2. Its generalization for detecting roofs/valleys and the proposed curve estimation procedure are also described there. Some theoretical properties of these

procedures are discussed in Section 3. A simulation study for evaluating their numerical performance is presented in Section 4. A real-data example is discussed in Section 5. Some concluding remarks are given in Section 6. Proofs of some theorems are outlined in Appendix.

2 Jump/Roof/Valley Detection and Curve Estimation

This section consists of four parts. In Subsection 2.1, our proposed procedure for detecting jumps in f is described. Its extention for detecting roofs/valleys is discussed in Subsection 2.2. In Subsection 2.3, we discuss curve estimation with possible jumps/roofs/valleys preserved. Selection of procedure parameters is discussed in Subsection 2.4.

2.1 Detection of jumps

Assume that $\{(x_i, Y_i), i = 1, ..., n\}$ are observed data from the regression model

$$Y_i = f(x_i) + \varepsilon_i, \text{ for } i = 1, ..., n, \tag{1}$$

where f is the unknown regression function, $\{x_i = i/n, i = 1, \ldots, n\}$ are the equally spaced design points in $[0, 1]$, and $\{\varepsilon_i, i = 1, \ldots, n\}$ are i.i.d. random errors with mean 0 and unknown variance σ^2. We further assume that the regression function f has the following expression:

$$f(x) = g(x) + \sum_{j=1}^{p} d_j I(x > s_j) + \sum_{k=1}^{q} \tilde{d}_k (x - \tilde{s}_k)_+, \text{ for } x \in [0, 1], \tag{2}$$

where g is a continuous function with continuous first order derivative g' in the entire design interval, p is the number of jumps in f, $\{s_j, j = 1, \ldots, p\}$ and $\{d_j, j = 1, \ldots, p\}$ are the corresponding jump positions and jump sizes, q is the number of jumps in f' (i.e., roofs/valleys of f), and $\{\tilde{s}_k, k = 1, \ldots, q\}$ and $\{\tilde{d}_k, k = 1, \ldots, q\}$ are the corresponding jump positions and jump sizes. Note that, in (2), the two sets $\{s_j, j = 1, \ldots, p\}$ and $\{\tilde{s}_k, k = 1, \ldots, q\}$ could have certain points in common. At those common points, both f and f' have jumps, although our visual perception of the jumps in f usually dominates that of the jumps in f'. In model (1), the assumption that the design points are equally spaced is made for simplicity of presentation only. Our proposed procedure should still perform well when they are unequally spaced but satisfy certain homogeneity conditions. In expression (2), both the continuity part g and the terms in the jump part, i.e., p, $\{s_j, j = 1, \ldots, p\}$, $\{d_j, j = 1, \ldots, p\}$, q, $\{\tilde{s}_k, k = 1, \ldots, q\}$, and $\{\tilde{d}_k, k = 1, \ldots, q\}$, are assumed unknown.

At a given point $x \in [h_n, 1-h_n]$, we consider the following local quadratic kernel smoothing procedure:

$$\min_{a,b,c\in R} \sum_{i=1}^{n} \left\{ Y_i - [a + b(x_i - x) + c(x_i - x)^2/2] \right\}^2 K\left(\frac{x_i - x}{h_n}\right), \tag{3}$$

where K is a non-negative kernel function with support $[-1, 1]$, and h_n is a bandwidth parameter. The solution to a, b, and c of the minimization problem (3) is denoted as $\widehat{a}(x), \widehat{b}(x)$, and $\widehat{c}(x)$. They have the following expressions:

$$\begin{aligned}
\widehat{b}(x) &= \sum_{i=1}^{n} \frac{(r_2 r_3 - r_1 r_4) + (r_0 r_4 - r_2^2)(x_i - x) + (r_1 r_2 - r_0 r_3)(x_i - x)^2}{r_0 r_2 r_4 - r_2^3 + 2 r_1 r_2 r_3 - r_1^2 r_4 - r_0 r_3^2} K\left(\frac{x_i - x}{h_n}\right) Y_i, \\
\widehat{c}(x) &= 2 \sum_{i=1}^{n} \frac{(r_1 r_3 - r_2^2) + (r_1 r_2 - r_0 r_3)(x_i - x) + (r_0 r_2 - r_1^2)(x_i - x)^2}{r_0 r_2 r_4 - r_2^3 + 2 r_1 r_2 r_3 - r_1^2 r_4 - r_0 r_3^2} K\left(\frac{x_i - x}{h_n}\right) Y_i, \\
\widehat{a}(x) &= \sum_{i=1}^{n} \frac{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h_n}\right) Y_i - r_1 \hat{b}(x) - 2 r_2 \hat{c}(x)}{r_0},
\end{aligned} \tag{4}$$

where $r_k = \sum_{i=1}^{n} (x_i - x)^k K\left(\frac{x_i - x}{h_n}\right)$, for $k = 0, 1, 2, 3, 4$. Then, $\widehat{a}(x)$, $\widehat{b}(x)$, and $\widehat{c}(x)$ are the local quadratic kernel estimators of $f(x)$, $f'(x)$, and $f''(x)$, respectively (cf., e.g., Fan and Gijbels 1996).

In (3), a quadratic function is fitted in the local neighborhood $[x - h_n, x + h_n]$ of the given point x, using the weighted least squares procedure, with the weights determined by the kernel function K. Intuitively, if x is a jump point, then observations on two different sides of x would be quite different in their values, the fitted quadratic curve by (3) would be steep, and consequently $|\widehat{b}(x)|$ would be large. Furthermore, in such cases, $\widehat{c}(x)$ would be close to 0, and the values of $\widehat{c}$ would change signs on two different sides of x, as discussed in Section 1 about the zero-crossing properties of f''. By combining such helpful information in both $\widehat{b}(x)$ and $\widehat{c}(x)$ about the jumps in f, we suggest the following jump detection procedure:

Proposed Procedure for Detecting Jumps in f

Step 1 For a given design point $x_i \in [h_n, 1-h_n]$, it is flagged as a detected jump point if (i) $|\widehat{b}(x_i)| \geq u_n$, (ii) $|\widehat{c}(x_i)| \leq v_n$, and (iii) there exist two design points x_{i_1} and x_{i_2} in $[x_i - h_n, x_i + h_n]$ such that $|\widehat{c}(x_{i_1})| > v_n, |\widehat{c}(x_{i_2})| > v_n$, and $\widehat{c}(x_{i_1})\widehat{c}(x_{i_2}) < 0$, where u_n and v_n are two non-negative threshold values.

Step 2 Assume that the detected jumps from Step 1 are $\{x_i^*, i = 1, 2, \ldots, m\}$. If there are integers $1 \leq t_1 < t_2 \leq m$ such that $x_j^* - x_{j-1}^* \leq h_n$, for $j = t_1 + 1, \ldots, t_2$, $x_{t_1}^* - x_{t_1-1}^* > h_n$, and

$x^*_{t_2+1} - x^*_{t_2} > h_n$, then the set $\{x^*_{t_1}, x^*_{t_1+1}, \ldots, x^*_{t_2}\}$ forms *a tie* in $\{x^*_i, i = 1, 2, \ldots, m\}$ and the entire tie set is replaced by its central point $(x^*_{t_1} + x^*_{t_2})/2$ for estimating the jump positions.

In the above procedure, condition (i) in Step 1 is on the estimated first order derivative of f, and conditions (ii) and (iii) are based on the zero-crossing properties of the estimated second order derivatives. Note that the estimated second order derivatives are random. Therefore, theoretically speaking, the chance that their values are exactly zero is zero, and estimated second order derivatives may change signs on two different sides of a continuity point of f as well. Inclusion of the threshold value v_n in conditions (ii) and (iii) is mainly for handling these practical issues. Step 2 is for deleting certain deceptive jump points. See Qiu (1994) for related discussion.

To demonstrate the above jump detection procedure, let us consider a simple regression function shown in Figure 2(a) by the solid line, which has two jumps at $x = 0.25$ and $x = 0.75$. A set of observations generated from model (1) with $\varepsilon_1 \sim N(0, 0.25^2)$ are also shown in that plot by small circles. The estimated first order and second order derivatives $\{\widehat{b}(x_i)\}$ and $\{\widehat{c}(x_i)\}$ by (3) are shown in Figures 2(b) and 2(c), respectively, using $h_n = 0.1$ and the Epanechnikov kernel function defined in Section 4. When $u_n = \widehat{\sigma}_n\sqrt{\chi^2_{1,0.9}(\widehat{\delta}_n)}$ and $v_n = \widehat{\sigma}^*_n\sqrt{\chi^2_{1,0.7}}$, the detected jumps by Step 1 of the above jump detection procedure are shown in Figure 2(d), where $\widehat{\sigma}_n$ and $\chi^2_{1,0.9}(\widehat{\delta}_n)$ are defined in Section 2.4 below, $\widehat{\sigma}^*_n$ is the standard error of $\widehat{c}(x)$ when x is a continuity point of f, defined similarly to $\widehat{\sigma}_n$, and $\chi^2_{1,0.7}$ is the 0.7-th quantile of the χ^2_1 distribution. The detected jumps shown in Figure 2(d) form two ties. After Step 2 of the jump detection procedure, all points in each tie are replaced by the central point of the tie. The modified detected jumps are shown in Figure 2(e). It can be seen that they estimate the true jumps well.

2.2 Detection of roofs/valleys

The jump detection procedure described in the previous subsection can be extended for detecting roofs/valleys of f. A natural idea is to replace estimators of f' and f'' in the above jump detection procedure by estimators of f'' and $f^{(3)}$, respectively, since roofs/valleys of f are actually jumps in f'. In the local quadratic kernel smoothing procedure (3), the bandwidth h_n used for jump detection could be replaced by another bandwidth $\tilde{h}_n$ for roof/valley detection. Then, for a given point $x \in [\tilde{h}_n, 1 - \tilde{h}_n]$, $f''(x)$ is estimated by $\widehat{c}(x)$. But, $f^{(3)}$ is not estimated in (3). One possibility is to replace (3) with the local cubic kernel smoothing, which would add much extra computation to the roof/valley detection procedure. As an alternative, we suggest using the zero-crossing properties

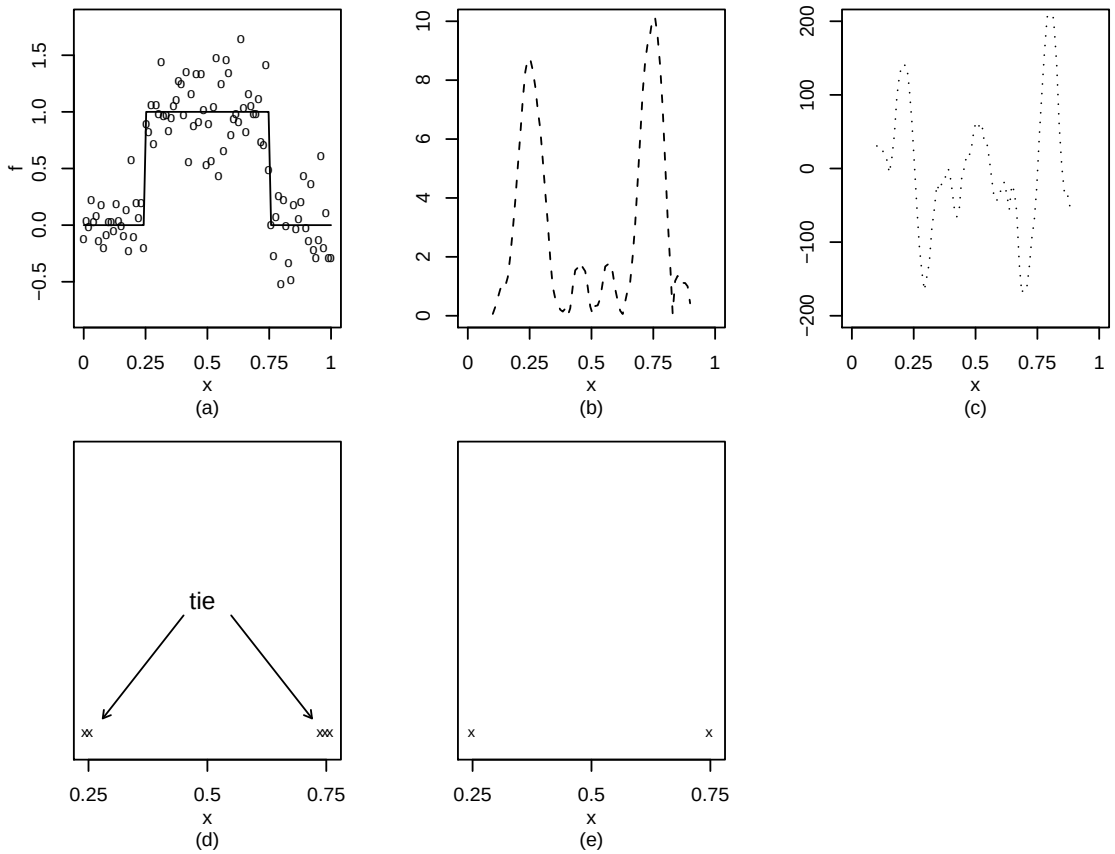

Figure 2: (a) True regression function (solid line) and a set of observations (small circles). (b) $\{\widehat{b}(x_i)\}$. (c) $\{\widehat{c}(x_i)\}$. (d) Detected jumps by Step 1 of the jump detection procedure. (e) Modified detected jumps by Step 2 of the jump detection procedure.

of

$$\widehat{f}^{(3)}(x) = [\widehat{c}_+(x) - \widehat{c}_-(x)]/\tilde{h}_n,$$

where $\widehat{c}_+(x)$ and $\widehat{c}_-(x)$ are solutions to c of the minimization problem (3) after $\sum_{i=1}^n$ is replaced by $\sum_{x_i \in (x, x+\tilde{h}_n]}$ and $\sum_{x_i \in [x-\tilde{h}_n, x)}$, respectively. Therefore, $\widehat{f}^{(3)}(x)$ is relatively convenient to obtain. Then, roofs/valleys of f can be detected by the following procedure:

Proposed Procedure for Detecting Roofs/Valleys of f

Step 1 For a given design point $x_i \in [\tilde{h}_n, 1 - \tilde{h}_n]$, it is flagged as a detected roof/valley point if (i) $|\widehat{c}(x_i)| \geq \tilde{u}_n$, (ii) $|\widehat{f}^{(3)}(x_i)| \leq \tilde{v}_n$, and (iii) there exist two design points x_{i_1} and x_{i_2} in $[x_i - \tilde{h}_n, x_i + \tilde{h}_n]$ such that $|\widehat{f}^{(3)}(x_{i_1})| > \tilde{v}_n, |\widehat{f}^{(3)}(x_{i_2})| > \tilde{v}_n$, and $\widehat{f}^{(3)}(x_{i_1})\widehat{f}^{(3)}(x_{i_2}) < 0$, where $\tilde{u}_n$ and $\tilde{v}_n$ are two non-negative threshold values.

Step 2 Assume that the detected roofs/valleys from Step 1 are $\{x_i^*, i = 1, 2, \ldots, m\}$. If there are

integers $1 \le t_1 < t_2 \le m$ such that $x^*_j - x^*_{j-1} \le \tilde{h}_n$, for $j = t_1+1, \ldots, t_2$, $x^*_{t_1} - x^*_{t_1-1} > \tilde{h}_n$, and $x^*_{t_2+1} - x^*_{t_2} > \tilde{h}_n$, then the set $\{x^*_{t_1}, x^*_{t_1+1}, \ldots, x^*_{t_2}\}$ forms *a tie* in $\{x^*_i, i = 1, 2, \ldots, m\}$ and the entire tie set is replaced by its central point $(x^*_{t_1} + x^*_{t_2})/2$ for estimating the roof/valley positions.

2.3 Curve estimation with possible jumps/roofs/valleys preserved

After all the jumps/roofs/valleys are detected, we have estimators $\widehat{p}$, $\{\widehat{s}_j, j = 1, \ldots, \widehat{p}\}$, $\widehat{q}$, and $\{\widehat{\tilde{s}}_k, k = 1, \ldots, \widehat{q}\}$ of the true numbers and positions of jumps/roofs/valleys. The jump sizes $\{d_j, j = 1, \ldots, p\}$ and $\{\tilde{d}_k, k = 1, \ldots, q\}$ can be estimated by

$$\widehat{d}_j = \widehat{a}_+(\widehat{s}_j) - \widehat{a}_-(\widehat{s}_j), \quad \widehat{\tilde{d}}_k = \widehat{b}_+(\widehat{\tilde{s}}_k) - \widehat{b}_-(\widehat{\tilde{s}}_k), \tag{5}$$

where $\widehat{a}_+(x)$ and $\widehat{a}_-(x)$ are the local quadratic kernel estimators of $f(x)$ obtained in the right-sided neighborhood $(x, x + h_{n,1}]$ and the left-sided neighborhood $[x - h_{n,1}, x)$, respectively. Similarly, $\widehat{b}_+(x)$ and $\widehat{b}_-(x)$ are the one-sided local quadratic kernel estimators of $f'(x)$, obtained from the neighborhoods $(x, x + \tilde{h}_{n,1}]$ and $[x - \tilde{h}_{n,1}, x)$, respectively. The two bandwidths $h_{n,1}$ and $\tilde{h}_{n,1}$ used here could be different from h_n and $\tilde{h}_n$ used in detecting jumps/roofs/valleys. Then, we define

$$Y^*_i = Y_i - \sum_{j=1}^{\widehat{p}} \widehat{d}_j I(x_i > \widehat{s}_j) - \sum_{k=1}^{\widehat{q}} \widehat{\tilde{d}}_k (x_i - \widehat{\tilde{s}}_k)_+, \text{ for } i = 1, ..., n.$$

From models (1) and (2), $\{(x_i, Y^*_i), i = 1, 2, \ldots, n\}$ can be regarded as observations from a model with regression function g. A natural estimator of $g(x)$, for $x \in [0, 1]$, is then the local linear kernel estimator, denoted as $\widehat{g}(x)$ and defined to be the solution to a^* of the following minimization problem:

$$\min_{a^*, b^* \in R} \sum_{i=1}^{n} \{Y^*_i - [a^* + b^*(x_i - x)]\}^2 K\left(\frac{x_i - x}{h^*_n}\right),$$

where h^*_n is a bandwidth that could be different from h_n and $\tilde{h}_n$ used in detecting jumps/roofs/valleys. Finally, our proposed curve estimator is defined by

$$\widehat{f}(x) = \widehat{g}(x) + \sum_{j=1}^{\widehat{p}} \widehat{d}_j I(x > \widehat{s}_j) + \sum_{k=1}^{\widehat{q}} \widehat{\tilde{d}}_k (x - \widehat{\tilde{s}}_k)_+, \text{ for } x \in [0, 1]. \tag{6}$$

2.4 Selection of procedure parameters

In the proposed procedures for detecting jumps, for detecting roofs/valleys, and for estimating the regression curve, there are three sets of parameters $\{h_n, u_n, v_n)$, $(\tilde{h}_n, \tilde{u}_n, \tilde{v}_n)$, and $(h_{n,1}, \tilde{h}_{n,1}, h^*_n)$ to

choose. From description about the related procedures given above, the first two sets of parameters can be chosen separately while the last one is chosen after the first two sets of parameters are determined.

We first discuss selection of h_n, u_n and v_n used in the proposed jump detection procedure. The parameter u_n is a threshold value for $|\widehat{b}(x_i)|$. We notice that $\widehat{b}(x)$ is a linear combination of all observations. Therefore, it has an asymptotic Normal distribution. When the kernel function is chosen to be symmetric and h_n satisfies the condition that $1/(nh_n) = o(1)$, it is not difficult to check that, at a continuity point x of f, the asymptotic mean of $\widehat{b}(x)$ is $f'(x)$ and its variance is

$$\sigma_n^2 = \sigma^2 \sum_{i=1}^{n} \frac{(x_i - x)^2}{r_2^2} K^2\left(\frac{x_i - x}{h_n}\right) + o\left(\frac{1}{nh_n}\right).$$

Thus, we have

$$\lim_{n\to\infty} \frac{\left(\widehat{b}(x) - f'(x)\right)^2}{\widehat{\sigma}_n^2} \stackrel{\mathcal{D}}{=} \chi_1^2,$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution, and $\widehat{\sigma}_n^2$ is an estimator of σ_n^2, obtained by replacing σ^2 in its expression above by a consistent estimator $\widehat{\sigma}^2$ and by ignoring the term $o\left(\frac{1}{nh_n}\right)$. Therefore, a natural choice for u_n is

$$u_n = \widehat{\sigma}_n \sqrt{\chi^2_{1,1-\alpha_n}(\widehat{\delta}_n)}, \tag{7}$$

where $\chi^2_{1,1-\alpha_n}(\widehat{\delta}_n)$ is the $(1-\alpha_n)$-th quantile of the χ_1^2 distribution with non-centrality parameter $\widehat{\delta}_n = (\widehat{f}'(x))^2/\widehat{\sigma}_n^2$, $\widehat{\sigma}_n^2 = \widehat{\sigma}^2 \sum_{i=1}^n \frac{(x_i-x)^2}{r_2^2} K^2\left(\frac{x_i-x}{h_n}\right)$, and $\widehat{f}'(x)$ is an estimator of $f'(x)$. In this paper, we compute $\widehat{\sigma}^2$ and $\widehat{f}'(x)$ based on the idea of one-sided function estimation originally suggested by Qiu (2003), described briefly below. We first make local linear kernel smoothing in the left-sided neighborhood $[x - h_n, x)$ and the right-sided neighborhood $(x, x + h_n]$, respectively; then, $\widehat{\sigma}^2$ is defined by the smaller one of the two weighted residual mean squares obtained from the two one-sided local fits, and $\widehat{f}'(x)$ is defined by the slope estimator obtained from the selected one-sided local fit. This procedure is chosen due to its simplicity. Otherwise, alternative procedures, including the ones by Chu *et al.* (1998) and Gijbels *et al.* (2007), can also be used. To choose u_n by (7), we still need to choose α_n. We found from numerical studies that the default level 0.05 often gives reasonably good results. Therefore, we suggest using $\alpha_n = 0.05$ in applications, although α_n can always be chosen together with h_n and v_n.

To choose parameters h_n and v_n properly, we should first choose a performance measure of the detected jumps. Let $J = \{s_1, s_2, \ldots, s_p\}$ and $\widehat{J}_n$ be the sets of true and detected jump points,

respectively. Then, a reasonable measure that is commonly used in the literature is the following Hausdorff distance:

$$d_H(\widehat{J}_n, J; h_n, v_n) = \max\left\{\max_{x\in\widehat{J}_n}\min_{x'\in J}|x-x'|,\ \max_{x\in J}\min_{x'\in\widehat{J}_n}|x-x'|\right\}.$$

Because J is unknown, the Hausdorff distance $d_H(\widehat{J}_n, J; h_n, v_n)$ can not be computed directly. To overcome this difficulty, we suggest using the following bootstrap procedure. Suppose that there are B bootstrap samples, and the set of detected jumps from the k-th bootstrap sample is $\widehat{J}_n^{(k)}$, for $k = 1, 2, \ldots, B$. Then, $d_H(\widehat{J}_n, J; h_n, v_n)$ can be estimated by

$$\widehat{d}_H(\widehat{J}_n, J; h_n, v_n) = \frac{1}{B}\sum_{k=1}^{B} d_H(\widehat{J}_n^{(k)}, \widehat{J}_n; h_n, v_n), \tag{8}$$

and the optimal values of h_n and v_n are approximated by the solution of the problem

$$\min_{h_n>0}\ \min_{v_n>0}\ \widehat{d}_H(\widehat{J}_n, J; h_n, v_n). \tag{9}$$

To generate a bootstrap sample, the following two-step procedure can be used:

Step 1 Obtain a jump-preserving estimator $\widehat{f}(x)$ of the true regression function $f(x)$, using the procedure by Qiu (2003), as described several paragraphs earlier for estimating $f'(x)$, except that $\widehat{f}(x)$ is defined by the intercept estimator obtained from the selected one-sided local fit. Then, define residuals $\widehat{\varepsilon}_i = Y_i - \widehat{f}(x_i)$, for $i = 1, 2, \ldots, n$.

Step 2 Draw with replacement a sample $\{\widehat{\varepsilon}_i^*, i = 1, 2, \ldots, n\}$ from the set of residuals $\{\widehat{\varepsilon}_i, i = 1, 2, \ldots, n\}$. Then, the bootstrap sample is defined by $\{(x_i, Y_i^*), i = 1, 2, \ldots, n\}$, where $Y_i^* = \widehat{f}(x_i) + \widehat{\varepsilon}_i^*$, for $i = 1, 2, \ldots, n$.

Based on our numerical experience, this scheme for obtaining bootstrap samples would provide better jump detection results, compared to the natural scheme to resample the original data $\{(x_i, Y_i), i = 1, 2, \ldots, n\}$ directly.

Parameters $\tilde{h}_n, \tilde{u}_n$, and $\tilde{v}_n$ can be chosen similarly. After studying the asymptotic distribution of $\widehat{c}(x)$, $\tilde{u}_n$ can be chosen to be

$$\tilde{u}_n = \widehat{\widetilde{\sigma}}_n\sqrt{\chi^2_{1,1-\alpha_n}(\widehat{\widetilde{\delta}}_n)},$$

where $\widehat{\widetilde{\sigma}}_n^2 = 4\widehat{\sigma}^2\sum_{i=1}^n \frac{\{r_0(x_i-x)^2-r_2\}^2}{(r_0r_4-r_2^2)^2}K^2\left(\frac{x_i-x}{\tilde{h}_n}\right)$, $\chi^2_{1,1-\alpha_n}(\widehat{\widetilde{\delta}}_n)$ is the $(1-\alpha_n)$-th quantile of the χ_1^2 distribution with non-centrality parameter $\widehat{\widetilde{\delta}}_n = (\widehat{f''}(x))^2/\widehat{\widetilde{\sigma}}_n^2$, and $\widehat{f''}(x)$ is an estimator of $f''(x)$

obtained similarly to $\widehat{f}'(x)$ used in (7). As in (7), we suggest using $\alpha_n = 0.05$. The same bootstrap procedure described above for choosing h_n and v_n can be used for choosing $\tilde{h}_n$ and $\tilde{v}_n$, except that J and $\widehat{J}_n$ should be replaced by $R = \{s'_1, s'_2, \ldots, s'_q\}$ and its estimator $\widehat{R}_n$, and $\widehat{J}_n^{(k)}$ should be replaced by $\widehat{R}_n^{(k)}$ which is the set of detected roofs/valleys from the k-th bootstrap sample.

When computing the curve estimator $\widehat{f}(x)$ defined in (6), the jumps/roofs/valleys have been detected beforehand. So, the bandwidths $h_{n,1}$, $\tilde{h}_{n,1}$ and h_n^* used in (5) and (6) can be determined simply by minimizing the following cross validation (CV) score:

$$CV(h_{n,1}, \tilde{h}_{n,1}, h_n^*) = \frac{1}{n}\sum_{i=1}^{n} [Y_i^* - \widehat{g}_{-i}(x_i)]^2 ,$$

where $\widehat{g}_{-i}(x_i)$ is the leave-one-out local linear kernel estimator of $g(x_i)$, computed from all modified observations $\{(x_j, Y_j^*), j = 1, 2, \ldots, n\}$, except the i-th one (x_i, Y_i^*).

3 Some Statistical Properties

In this section, we provide some statistical properties of the proposed jump/roof/valley detection and curve estimation procedures. Theorem 1 below is concerned about the asymptotic properties of the local quadratic estimators (cf., expressions in (4)) in regions where the true regression function f is continuous.

Theorem 1. Assume that the regression function f has continuous third-order derivatives in $[0, 1]$ except at the jump/roof/valley points in $J \bigcup R$, the kernel function K is a Lipschitz continuous and symmetric density function in $[-1, 1]$, and the bandwidth h_n satisfies the conditions that $h_n = o(1)$ and $1/(nh_n^4) = o(1)$. Then,

$$\begin{aligned}
\| \widehat{a} - f \|_{\Omega_n} &= o\left(h_n^2\right) + O\left(\sqrt{\frac{ln(n)}{nh_n}}\right), \quad a.s., \\
\| \widehat{b} - f' \|_{\Omega_n} &= o(h_n) + O\left(\sqrt{\frac{ln(n)}{nh_n^3}}\right), \quad a.s., \\
\| \widehat{c} - f'' \|_{\Omega_n} &= o(h_n) + O\left(\sqrt{\frac{ln(n)}{nh_n^5}}\right), \quad a.s.,
\end{aligned}$$

where $\Omega_n = \{x : x \in [h_n, 1 - h_n]$ and $|x - x'| \geq h_n$ for any $x' \in J \bigcup R\}$, and $\| f \|_{\Omega_n} = \max_{x \in \Omega_n} |f(x)|$.

By Theorem 1, in continuity regions Ω_n of f, $\widehat{a}, \widehat{b}$, and $\widehat{c}$ are almost surely consistent estimators of f, f', and f'', respectively, if we further assume that $\sqrt{ln(n)/(nh_n^5)} = o(1)$. It is clear that

one-sided estimators $\widehat{a}_+, \widehat{a}_-, \widehat{b}_+, \widehat{b}_-, \widehat{c}_+$, and $\widehat{c}_-$, used in detecting roofs/valleys and in estimating f (cf., expressions of $\widehat{f}^{(3)}(x), \widehat{d}_j$, and $\widehat{\tilde{d}}_k$ in Subsections 2.2 and 2.3), would have the same asymptotic properties as those described in Theorem 1. Around the jump/roof/valley points, we have the following results in Theorems 2 and 3.

Theorem 2. Let s be a given jump point, x be a point in $(s-h_n, s+h_n)$, and the Euclidean distance between x and s is τh_n, where $0 \le \tau < 1$ is a constant. (Note that x depends on n although it is not made explicit in notation.) Let $Q_n^{(1)}(x)$ and $Q_n^{(2)}(x)$ be the left and right parts of $[x-h_n, x+h_n]$, separated by s. For $j=1,2$, $Q^{(j)} = \{u : u \in [-1,1] \text{ and } x+uh_n \in Q_n^{(j)}(x)\}$ are the corresponding parts of the support of K. The jump size of f at s is denoted as d. Then, under the conditions in Theorem 1 and the extra condition that $ln(n)/(nh_n^3) = o(1)$, we have

(i) $\widehat{b}(x) = \frac{d}{h_n \nu_2} \int_{Q^{(2)}} uK(u)du + o(1/h_n)$, a.s., and

(ii) $$\widehat{c}(x) = \begin{cases} O(1/h_n), \ a.s., & \text{if } \tau = 0, \\ C/h_n^2 + O(1/h_n), \ a.s., & \text{if } 0 < \tau < 1 \text{ and } x \in Q_n^{(1)}(x), \\ -C/h_n^2 + O(1/h_n), \ a.s., & \text{if } 0 < \tau < 1 \text{ and } x \in Q_n^{(2)}(x), \end{cases}$$

where $\nu_2 = \int_{-1}^{1} u^2 K(u)du$ and C is a positive constant.

Theorem 3. Let $\tilde{s}$ be a given roof/valley point, x be a point in $(\tilde{s}-\tilde{h}_n, \tilde{s}+\tilde{h}_n)$, and the Euclidean distance between x and $\tilde{s}$ be $\tau \tilde{h}_n$, where $0 \le \tau < 1$ is a constant. Let $\tilde{d}$ be the jump size of f' at $\tilde{s}$, and $\tilde{Q}_n^{(1)}(x), \tilde{Q}_n^{(2)}(x), \tilde{Q}^{(1)}$ and $\tilde{Q}^{(2)}$ are defined similarly to $Q_n^{(1)}(x)$, $Q_n^{(2)}(x)$, $Q^{(1)}$ and $Q^{(2)}$ in Theorem 2. Then, if f is continuous in $[x-\tilde{h}_n, x+\tilde{h}_n]$ and $\tilde{h}_n$ satisfies the conditions that $1/(n\tilde{h}_n^2) = o(1)$ and $ln(n)/(n\tilde{h}_n^5) = o(1)$, we have

(i) $\widehat{c}(x) = \frac{2\tilde{d}}{\tilde{h}_n(\nu_4 - \nu_2^2)} \int_{\tilde{Q}^{(2)}} (u^3 - \nu_2 u)K(u)du + o(1/\tilde{h}_n)$, a.s., and

(ii) $$\widehat{c}_+(x) - \widehat{c}_-(x) = \begin{cases} o(1/\tilde{h}_n), \ a.s., & \text{if } \tau = 0, \\ \tilde{C}/\tilde{h}_n + o(1/\tilde{h}_n), \ a.s., & \text{if } 0 < \tau < 1 \text{ and } x \in \tilde{Q}_n^{(1)}(x), \\ -\tilde{C}/\tilde{h}_n + o(1/\tilde{h}_n), \ a.s., & \text{if } 0 < \tau < 1 \text{ and } x \in \tilde{Q}_n^{(2)}(x), \end{cases}$$

where $\nu_k = \int_{-1}^{1} u^k K(u)du$, for $k = 2, 4$, and $\tilde{C}$ is a positive constant.

Based on results in Theorems 1 and 2, we have the following consistency result for the detected jumps of the proposed jump detection procedure.

Theorem 4. Given the conditions in Theorems 1 and 2, and the extra conditions that (i) $\alpha_n = o(1)$, (ii) $Z_{1-\alpha_n/2}/\sqrt{nh_n} = o(1)$, (iii) $1/(v_n h_n) = o(1)$, and (iv) $v_n h_n^2 = o(1)$, we have $d_H(\widehat{J}_n, J; h_n, v_n) = O(h_n)$, a.s.

Theorem 4 says that, under some generality conditions, the Hausdorff distance between the set of detected jump points and the set of true jumps points would converge to zero almost surely. Notice that, when n is large enough, $\widehat{J}_n \subseteq O_n(J)$ and $J \subseteq \widehat{J}_n$ (cf., expressions (31) and (32) in Appendix), and that there would not be more than one detected jumps in $[s_j - h_n, s_j + h_n]$, for each $j = 1, 2, \ldots, p$, after the modification step (i.e., Step 2) of the jump detection procedure. Therefore, when n is large enough, the number of detected jump points would equal the true number of jumps, and the positions of detected jump points would be within h_n of the true jump positions, almost surely. For the detected roofs/valleys, we have similar results summarized in the following theorem.

Theorem 5. Under the conditions in Theorems 1 and 3, and the extra conditions that (i) $\alpha_n = o(1)$, (ii) $Z_{1-\alpha_n/2}/\sqrt{n\tilde{h}_n^3} = o(1)$, (iii) $1/(\tilde{v}_n \tilde{h}_n) = o(1)$, and (iv) $\tilde{v}_n \tilde{h}_n = O(1)$, we have $d_H(\widehat{R}_n, R; \tilde{h}_n, \tilde{v}_n) = O(\tilde{h}_n)$, a.s.

Similar to the explanations given after Theorem 4, by Theorem 5, when n is large enough, the number of detected roof/valley points would equal the true number of roof/valley points, and the positions of detected roof/valley points would be within $\tilde{h}_n$ of the true roof/valley positions, almost surely. Regarding the estimated jump sizes in both f and f', we have the following results.

Theorem 6. Under the conditions in Theorems 1–5, and the extra conditions that (i) $h_n/h_{n,1} = o(1)$ and (ii) $\tilde{h}_n/\tilde{h}_{n,1} = o(1)$, we have, for $j = 1, 2, \ldots, p$,

$$\begin{aligned} |\widehat{d}_j - d_j| &= O\left(\frac{h_n}{h_{n,1}}\right) + o(h_{n,1}^2) + O\left(\sqrt{\frac{ln(n)}{nh_{n,1}}}\right), a.s., \\ |\widehat{\tilde{d}}_k - \tilde{d}_k| &= O\left(\frac{\tilde{h}_n}{\tilde{h}_{n,1}}\right) + o(\tilde{h}_{n,1}) + O\left(\sqrt{\frac{ln(n)}{n\tilde{h}_n^3}}\right), a.s. \end{aligned}$$

Finally, based on Theorems 1–6, we have the following result for the estimated regression curve $\widehat{f}$. Proofs of all the theorems are outlined in Appendix.

Theorem 7. Under the conditions in Theorems 1–6, and the extra conditions that (i) $h_n^* = o(1)$, and (ii) $1/(nh_n^*) = o(1)$, we have

$$\int_{H_n} \left(\widehat{f}(x) - f(x)\right)^2 dx = O\left(h_n + \tilde{h}_n + h_n^*\right) + O\left(\frac{1}{n^2 (h_n^*)^2}\right) +$$

$$O\left(\frac{ln(n)}{nh_{n,1}}+\frac{ln(n)}{n\tilde{h}_{n,1}}+\frac{ln(n)}{nh_n^*}\right)+$$
$$O\left(\frac{h_n^2}{h_{n,1}^2}+\frac{\tilde{h}_n^2}{\tilde{h}_{n,1}^2}\right)+o\left(h_{n,1}^4+\tilde{h}_{n,1}^4\right), a.s.,$$

where $H_n=[h_n^*, 1-h_n^*]$.

4 A Simulation Study

In this section, we present some simulation results regarding the numerical performance of the proposed jump/roof/valley detection procedures and the jump/roof/valley-preserving curve estimation procedure. In the simulation study, the following two true regression functions are considered:

$$f_1(x)=\begin{cases}4x, & \text{if } 0\le x<0.25,\\ 4x+1, & \text{if } 0.25\le x<0.5,\\ -4x+5, & \text{if } 0.5\le x<0.75,\\ 4x, & \text{if } 0.75\le x\le 1;\end{cases} \quad f_2(x)=\begin{cases}1, & \text{if } 0\le x<0.25,\\ 4x-1, & \text{if } 0.25\le x<0.5,\\ -4x+3, & \text{if } 0.5\le x<0.75,\\ 1-\exp\{-15(x-0.75)\}, & \text{if } 0.75\le x\le 1.\end{cases}$$

The function f_1 has two jumps at 0.25 and 0.75, and two jumps in its first-order derivative (i.e., roofs/valleys) at 0.5 and 0.75. The function f_2 has one jump at 0.25, and three jumps in its first-order derivative at 0.25, 0.5, and 0.75 with different magnitudes. The two functions cover several different patterns of jumps/roofs/valleys. With either regression function, n observations are generated from model (1), in which $\{\varepsilon_i, i=1,2,\ldots,n\}$ are i.i.d. from $N(0,\sigma^2)$ distribution. When $n=100$ and $\sigma=0.5$, one realization of the observations from f_1 is shown in Figure 3(a), and one realization from f_2 is shown in Figure 3(b), along with the true regression functions.

To use the proposed jump/roof/valley detection procedures, the kernel function is chosen to be the Epanechnikov function $K(x)=(3/4)(1-x^2)I(|x|\le 1)$, and α_n used in defining u_n (cf., (7)) and $\tilde{u}_n$ is fixed at 0.05. Procedure parameters (h_n, v_n) and $(\tilde{h}_n, \tilde{v}_n)$ are first chosen by minimizing the Hausdorff distance between $\widehat{J}_n$ and J and the Hausdorff distance between $\widehat{R}_n$ and R, respectively. Although we can not choose parameter values in this way in practice because both J and R are often unknown, performance of the proposed procedures with such optimal parameter values (the corresponding jump/roof/valley detection procedures are denoted as PP_{OPT}) can be used as a reference in evaluating performance of the same procedures when their parameters are chosen by the suggested bootstrap procedure (the corresponding jump/roof/valley detection procedures are denoted as PP_{BST}). In PP_{BST}, the bootstrap sample size is fixed at $B=100$ (cf., equation (8)).

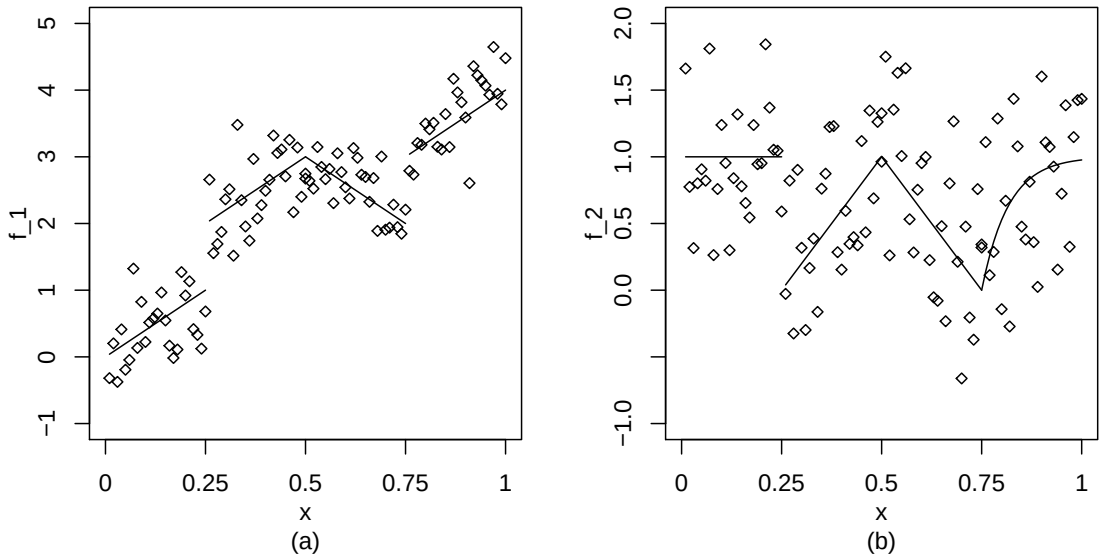

Figure 3: (a) Small diamonds denote the observations and solid lines denote the true regression function f_1. (b) Small diamonds denote the observations and solid lines denote the true regression function f_2. In both plots, $n = 100$ and $\sigma = 0.5$.

As comparisons, we also consider two existing methods. One is the jump detection method by Loader (1996), which is based on local linear kernel estimation. The original version of this method only handles cases when the number of jumps is known. Here, we make the following modifications to make it comparable with our proposed method. First, a threshold parameter is used, as in (7), for the jump detection criterion by Loader, and α_n is fixed at 0.05. Second, detected jumps are modified by the same modification procedure discussed in Section 2.1. Third, its bandwidth parameter is chosen by a bootstrap procedure similar to the one described in Section 2.4. To detect roofs/valleys by this method, we simply use $\widehat{f}'_+(x) - \widehat{f}'_-(x)$ as the detection criterion, where $\widehat{f}'_+(x)$ and $\widehat{f}'_-(x)$ are local linear kernel estimators of $f'(x)$ constructed from a right-sided and a left-sided neighborhoods, respectively. This method is denoted as Loader hereafter. The second method detects jumps using zero-crossing properties of $\widehat{c}(x)$ obtained in (4), and detects roofs/valleys using zero-crossing properties of $\widehat{f}^{(3)}(x)$, in similar ways to that described in Section 2. This procedure has two parameters to choose: the bandwidth used in local smoothing and the threshold value. It can be regarded as a modified version of the zero-crossing edge detectors used in image processing (cf., e.g., Fleck 1992), and is denoted as Zero-Crossing hereafter.

The following eight cases are considered in this study: n =100 or 200, σ =0.25 or 0.5, and $f = f_1$ or f_2. In each case, 100 replicated simulations are performed. Tables 1 and 2 present frequency tables of the numbers of detected jump/roof/valley points by each method, averaged

values of the Hausdorff distances between $\widehat{J}_n$ and J, or between $\widehat{R}_n$ and R, over 100 simulations, and the averaged values of the selected procedure parameters (i.e., (h_n, v_n) or $(\tilde{h}_n, \tilde{v}_n)$ for procedures $\mathrm{PP}_{\mathrm{OPT}}$, $\mathrm{PP}_{\mathrm{BST}}$, and Zero-Crossing, and a bandwidth for procedure Loader).

From the tables, we can have the following conclusions. (i) The performance of $\mathrm{PP}_{\mathrm{BST}}$ is a little worse than the performance of $\mathrm{PP}_{\mathrm{OPT}}$; the two sets of results are compatible in all cases considered, implying that the proposed bootstrap procedure for selecting parameter values works reasonably well in this example. As a matter of fact, selected parameter values by the bootstrap procedure are close to the optimal values in all cases. (ii) From the presented frequency tables of the numbers of detected jumps/roofs/valleys, we can see that numbers of detected jumps/roofs/valleys by both procedures equal to the correct numbers in most replicated simulations, and the results become more impressive when n gets larger or σ gets smaller, which is due to the consistency of $\widehat{J}_n$ and $\widehat{R}_n$ (cf., Theorems 4 and 5 in Section 3). (iii) Results about the averaged Hausdorff distances between $\widehat{J}_n$ and J and between $\widehat{R}_n$ and R are also reasonably good. (iv) $\mathrm{PP}_{\mathrm{BST}}$ performs uniformly better than procedures Loader and Zero-Crossing.

To further study the ability of the proposed procedure $\mathrm{PP}_{\mathrm{BST}}$ in detecting jumps/roofs/valleys, Figure 4 presents the bar charts of the frequencies of the detected jumps/roofs/valleys that are within 0.02 of the true jumps/roofs/valleys in various cases. From the bar charts, we can see that the sample size n, the noise level σ, and the jump size in f or f' all have certain impact on the performance of $\mathrm{PP}_{\mathrm{BST}}$. $\mathrm{PP}_{\mathrm{BST}}$ would perform better when n is larger, σ is smaller, or jump size is larger (e.g., the jump size in f_2' at 0.75 is larger than that at 0.5), which is intuitively reasonable. Furthermore, the first jump in f_1 at 0.25 seems easier to detect, compared to its second jump at 0.75, which implies that in finite sample cases slope of the regression curve f has an impact on detection of jumps in f. Similarly, the roof/valley in f_1 at 0.5 seems easier to detect than its roof/valley at 0.75. These results are not revealed well in the large-sample properties of the proposed procedures that are given in Section 3.

Next, we consider curve estimation with jumps/roofs/valleys preserved. Three competitive methods are considered here. One is the proposed estimator (6) based on jump/roof/valley detection by $\mathrm{PP}_{\mathrm{BST}}$. The second one is the estimator (6) based on jump/roof/valley detection by procedure Loader. These two curve estimators are denoted as $\hat{f}_{\mathrm{BST}}$ and $\hat{f}_{\mathrm{L}}$, respectively, and their parameters are chosen by the CV procedure described in Section 2.4. The third one is the method by Gijbels *et al.* (2007), which was shown to perform well in various cases, compared to

Table 1: When $f = f_1$, $n = 100$ or 200, and σ=0.25 or 0.5, this table presents frequency tables of the numbers of detected jumps, frequency tables of the numbers of detected roofs/valleys, averaged Hausdorff distances between $\widehat{J}_n$ and J, averaged Hausdorff distances between $\widehat{R}_n$ and R, and averaged values of selected procedure parameters, for various methods based on 100 replicated simulations.

			Jump Detection		Roof/Valley Detection	
n	σ	Method	Frequency Table	d_H	Frequency Table	d_H
100	0.25	PP_{OPT}	2 3 95 5	0.021 (0.11, 15.29)	2 3 87 13	0.038 (0.12, 54.15)
		PP_{BST}	2 3 93 7	0.024 (0.11, 23.81)	2 3 87 13	0.039 (0.12, 83.51)
		Loader	2 3 4 86 11 3	0.078 (0.11)	2 3 85 15	0.082 (0.11)
		Zero-Crossing	2 3 4 79 16 5	0.101 (0.12, 15.49)	2 3 4 76 23 1	0.154 (0.13, 32.78)
	0.5	PP_{OPT}	1 2 3 3 85 12	0.047 (0.11, 29.68)	1 2 3 18 74 8	0.044 (0.12, 90.67)
		PP_{BST}	1 2 3 4 9 79 11 1	0.052 (0.11, 39.96)	1 2 3 22 72 6	0.056 (0.13, 121.50)
		Loader	2 3 4 5 68 23 8 1	0.107 (0.11)	2 3 4 71 18 11	0.110 (0.11)
		Zero-Crossing	1 2 3 4 3 60 33 4	0.115 (0.12, 26.12)	2 3 4 62 23 15	0.295 (0.14, 70.02)
200	0.25	PP_{OPT}	2 3 98 2	0.015 (0.11, 14.83)	2 3 95 5	0.034 (0.11, 36.65)
		PP_{BST}	2 3 97 3	0.019 (0.11, 13.45)	2 3 93 7	0.037 (0.10, 75.18)
		Loader	1 2 3 1 92 7	0.063 (0.10)	2 3 89 11	0.057 (0.10)
		Zero-Crossing	2 3 4 5 89 5 4 2	0.081 (0.12, 11.34)	2 3 4 81 17 2	0.106 (0.13, 21.12)
	0.5	PP_{OPT}	1 2 3 7 91 2	0.034 (0.11, 25.81)	1 2 3 1 85 14	0.040 (0.12, 73.19)
		PP_{BST}	1 2 3 9 89 2	0.042 (0.11, 34.03)	1 2 3 2 83 15	0.047 (0.12, 92.13)
		Loader	1 2 3 12 83 5	0.074 (0.10)	2 3 4 78 13 9	0.066 (0.11)
		Zero-Crossing	1 2 3 19 76 5	0.075 (0.12, 21.35)	2 3 4 74 20 6	0.149 (0.13, 43.37)

Table 2: When $f = f_2$, $n = 100$ or 200, and σ=0.25 or 0.5, this table presents frequency tables of the numbers of detected jumps, frequency tables of the numbers of detected roofs/valleys, averaged Hausdorff distances between $\widehat{J}_n$ and J, averaged Hausdorff distances between $\widehat{R}_n$ and R, and averaged values of selected procedure parameters, for various methods based on 100 replicated simulations.

			Jump Detection		Roof/Valley Detection	
n	σ	Method	Frequency Table	d_H	Frequency Table	d_H
100	0.25	PP_{OPT}	1 2 94 6	0.031 (0.11, 10.72)	3 4 93 7	0.037 (0.12, 54.10)
		PP_{BST}	1 2 87 13	0.034 (0.12, 15.84)	3 4 94 6	0.041 (0.13, 69.71)
		Loader	1 2 3 85 14 1	0.097 (0.11)	3 4 5 83 12 5	0.111 (0.13)
		Zero-Crossing	1 2 3 69 26 5	0.114 (0.11, 11.89)	2 3 4 2 71 26	0.145 (0.13, 53.31)
	0.5	PP_{OPT}	1 2 3 82 12 6	0.039 (0.12, 25.81)	1 2 3 4 3 9 71 17	0.058 (0.12, 113.24)
		PP_{BST}	1 2 3 83 14 3	0.054 (0.12, 31.23)	1 2 3 4 3 6 69 22	0.077 (0.13, 132.66)
		Loader	1 2 3 70 26 4	0.134 (0.12)	1 2 3 4 1 11 58 30	0.142 (0.13)
		Zero-Crossing	1 2 3 4 62 28 8 2	0.174 (0.13, 19.97)	2 3 4 7 55 38	0.215 (0.13, 126.56)
200	0.25	PP_{OPT}	1 2 99 1	0.023 (0.11, 7.29)	3 4 98 2	0.029 (0.12, 44.36)
		PP_{BST}	1 2 96 4	0.027 (0.12, 8.42)	3 4 96 4	0.032 (0.12, 53.77)
		Loader	1 2 3 87 11 2	0.053 (0.11)	3 4 5 88 10 2	0.062 (0.12)
		Zero-Crossing	1 2 3 73 20 7	0.084 (0.10, 10.46)	2 3 4 4 74 22	0.102 (0.13, 39.75)
	0.5	PP_{OPT}	1 2 3 88 8 4	0.035 (0.12, 14.83)	2 3 4 3 84 13	0.047 (0.12, 74.51)
		PP_{BST}	1 2 3 81 13 6	0.044 (0.12, 20.01)	2 3 4 6 78 16	0.058 (0.13, 75.76)
		Loader	1 2 3 76 18 6	0.116 (0.11)	2 3 4 5 71 24	0.128 (0.13)
		Zero-Crossing	1 2 3 4 67 21 11 1	0.145 (0.12, 13.55)	2 3 4 3 64 33	0.182 (0.13, 77.39)

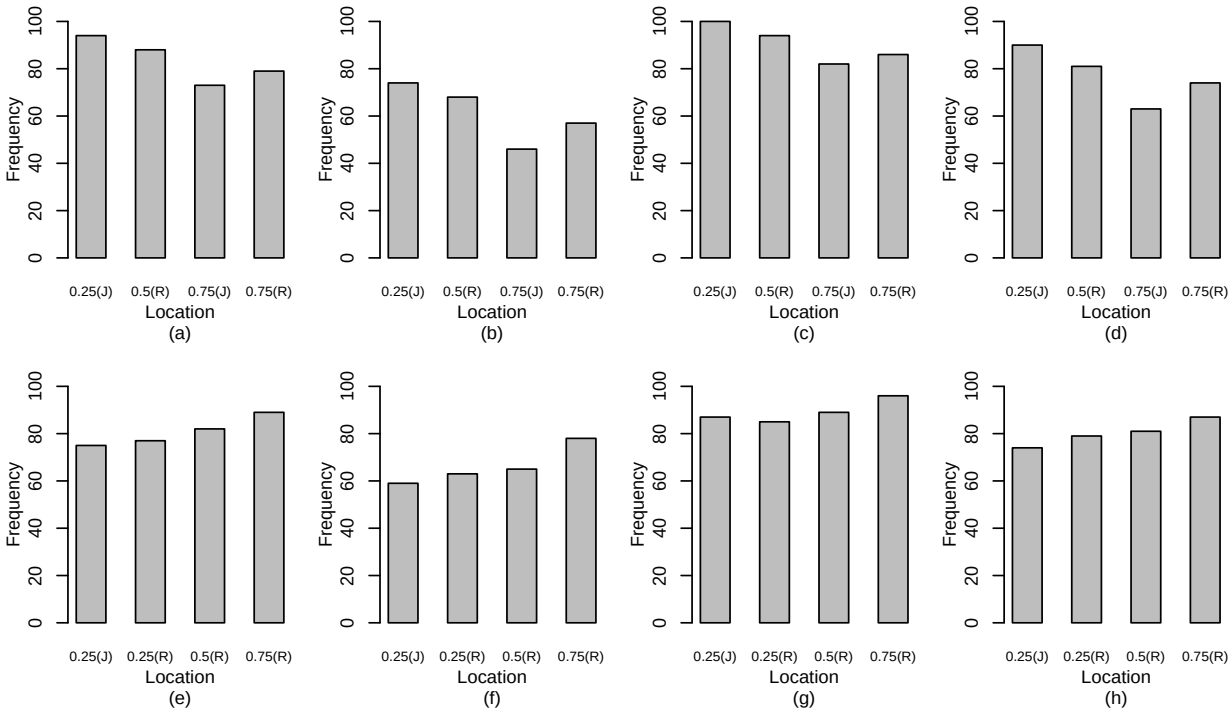

Figure 4: Bar charts of the frequencies of the detected jumps(J) and roofs/valleys(R) that are within 0.02 of the true jumps/roofs/valleys in various cases. (a) $n = 100$ and $\sigma = 0.25$. (b) $n = 100$ and $\sigma = 0.5$. (c) $n = 200$ and $\sigma = 0.25$. (d) $n = 200$ and $\sigma = 0.5$. These four plots are in cases when $f = f_1$. Plots (e)–(h) present corresponding results in cases when $f = f_2$.

the sigma filter and M-smoother by Chu *et al.* (1998). Its curve estimator is denoted as $\hat{f}_{\mathrm{G}}$. For each estimator, we compute an approximation of the mean integrated squared error (MISE), which is defined to be the sample mean of $ISE = \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{f}(x_i) - f(x_i)\right]^2$ for estimator $\widehat{f}$, based on 100 replications. To further investigate jump/roof/valley preservation, we also compute approximations of local MISE, defined in the same way as the approximations of MISE except that the former is computed within 0.05 of each true jump/roof/valley point. Related results are presented in Tables 3 and 4. From the tables, we can see that $\hat{f}_{\mathrm{BST}}$ performs uniformly better than $\hat{f}_{\mathrm{L}}$ and $\hat{f}_{\mathrm{G}}$ in both MISE and local MISE. The typical estimators by the three methods, defined to be the ones with median ISE values among 100 replications, are shown in Figure 5. From the plots, we can see that $\hat{f}_{\mathrm{BST}}$ does preserve jumps/roofs/valleys reasonably well.

5 A Real-Data Example

In this section, we demonstrate the proposed method using a real-data example about daily exchange rates between Korean currency Won and US currency Dollar between January 2, 1997 and

Table 3: MISE, local MISE and their standard errors (in parenthesis) of various curve estimators when $f = f_1$.

n	σ	Method	MISE	Local MISE		
				0.25 ± 0.05	0.5 ± 0.05	0.75 ± 0.05
100	0.25	$\hat{f}_{\mathrm{BST}}$	0.0041 (0.0003)	0.0044 (0.0005)	0.0069 (0.0008)	0.0073 (0.0006)
		$\hat{f}_{\mathrm{L}}$	0.0211 (0.0004)	0.0770 (0.0021)	0.0113 (0.0009)	0.0540 (0.0014)
		$\hat{f}_{\mathrm{G}}$	0.0189 (0.0004)	0.0572 (0.0019)	0.0125 (0.0011)	0.0660 (0.0021)
	0.5	$\hat{f}_{\mathrm{BST}}$	0.0123 (0.0009)	0.0098 (0.0011)	0.0094 (0.0011)	0.0087 (0.0012)
		$\hat{f}_{\mathrm{L}}$	0.0441 (0.0013)	0.1067 (0.0046)	0.0539 (0.0050)	0.1169 (0.0056)
		$\hat{f}_{\mathrm{G}}$	0.0444 (0.0012)	0.0947 (0.0040)	0.0354 (0.0032)	0.1239 (0.0060)
200	0.25	$\hat{f}_{\mathrm{BST}}$	0.0018 (0.0002)	0.0021 (0.0002)	0.0013 (0.0002)	0.0013 (0.0002)
		$\hat{f}_{\mathrm{L}}$	0.0122 (0.0002)	0.0460 (0.0008)	0.0060 (0.0006)	0.0415 (0.0008)
		$\hat{f}_{\mathrm{G}}$	0.0149 (0.0002)	0.0509 (0.0008)	0.0061 (0.0005)	0.0595 (0.0010)
	0.5	$\hat{f}_{\mathrm{BST}}$	0.0071 (0.0006)	0.0059 (0.0009)	0.0054 (0.0009)	0.0052 (0.0008)
		$\hat{f}_{\mathrm{L}}$	0.0245 (0.0008)	0.0667 (0.0028)	0.0286 (0.0032)	0.0625 (0.0030)
		$\hat{f}_{\mathrm{G}}$	0.0298 (0.0007)	0.0748 (0.0029)	0.0198 (0.0020)	0.0877 (0.0033)

Table 4: MISE, local MISE and their standard errors (in parenthesis) of various curve estimators when $f = f_2$.

n	σ	Method	Global MISE	Local MISE		
				0.25 ± 0.05	0.5 ± 0.05	0.75 ± 0.05
100	0.25	$\hat{f}_{\mathrm{BST}}$	0.0075 (0.0003)	0.0054 (0.0005)	0.0044 (0.0005)	0.0105 (0.0008)
		$\hat{f}_{\mathrm{L}}$	0.0156 (0.0003)	0.0789 (0.0020)	0.0075 (0.0007)	0.0175 (0.0015)
		$\hat{f}_{\mathrm{G}}$	0.0140 (0.0004)	0.0607 (0.0019)	0.0075 (0.0008)	0.0157 (0.0015)
	0.5	$\hat{f}_{\mathrm{BST}}$	0.0260 (0.0011)	0.0183 (0.0024)	0.0182 (0.0021)	0.0264 (0.0023)
		$\hat{f}_{\mathrm{L}}$	0.0375 (0.0012)	0.1146 (0.0055)	0.0287 (0.0030)	0.0384 (0.0036)
		$\hat{f}_{\mathrm{G}}$	0.0368 (0.0012)	0.1011 (0.0043)	0.0288 (0.0029)	0.0386 (0.0034)
200	0.25	$\hat{f}_{\mathrm{BST}}$	0.0042 (0.0002)	0.0031 (0.0004)	0.0029 (0.0004)	0.0054 (0.0004)
		$\hat{f}_{\mathrm{L}}$	0.0091 (0.0002)	0.0462 (0.0009)	0.0047 (0.0004)	0.0109 (0.0009)
		$\hat{f}_{\mathrm{G}}$	0.0098 (0.0002)	0.0531 (0.0011)	0.0046 (0.0004)	0.0093 (0.0007)
	0.5	$\hat{f}_{\mathrm{BST}}$	0.0147 (0.0007)	0.0103 (0.0013)	0.0090 (0.0009)	0.0129 (0.0010)
		$\hat{f}_{\mathrm{L}}$	0.0213 (0.0007)	0.0651 (0.0027)	0.0141 (0.0014)	0.0243 (0.0019)
		$\hat{f}_{\mathrm{G}}$	0.0241 (0.0007)	0.0859 (0.0029)	0.0145 (0.0014)	0.0228 (0.0018)

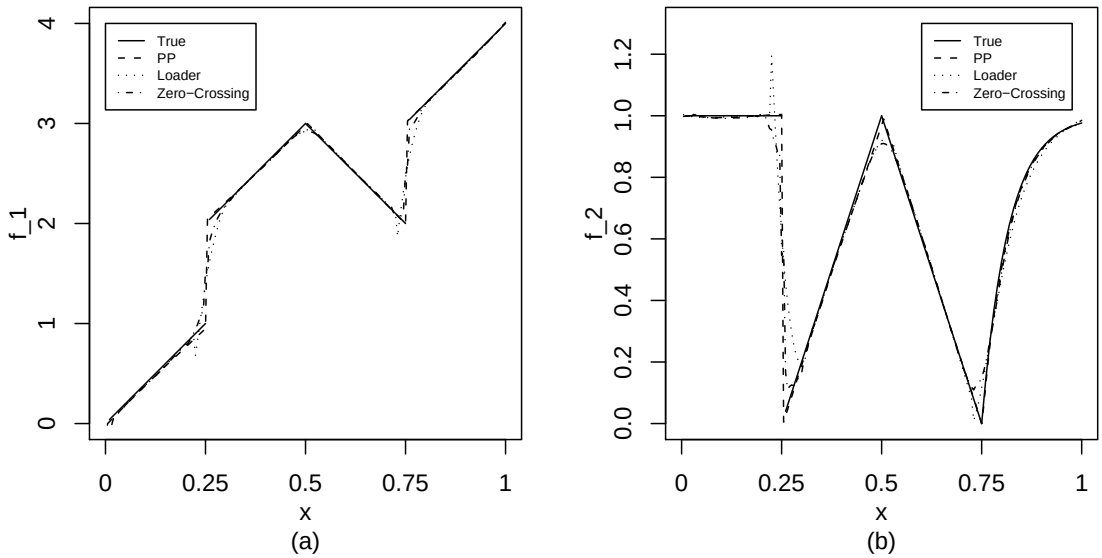

Figure 5: True regression functions (solid) and their estimators. Plot (a): f_1. Plot (b): f_2.

December 31, 1998. During this two-year period, the world financial market experienced several crises and the daily exchange rates between Won and Dollar were unstable. This can be seen from Figure 6(a) in which 498 daily exchange rates (Won/Dollar) observed in that period are shown. Our proposed methods are relevant to this example because the proposed jump/roof/valley detection procedures can identify positions where the exchange rates jump and where they change trends (either from increasing to decreasing, or from decreasing to increasing). Such information is important to financial research (cf., e.g., Devereux 1999).

In the proposed jump/roof/valley detection procedures, we choose the kernel function and the value of α_n to be the same as those used in the simulation examples in Section 4. After normalizing the design points such that the normalized design interval is [0,1], and after choosing procedure parameters properly by the proposed bootstrap procedure, our jump detection procedure identifies one jump point at the 229-th observation, which corresponds to the date December 01, 1997. This result is consistent with the fact that Korea experienced a financial crisis starting from December 1997, and the exchange rates from Dollar to Won surged at that time which resulted in later interventions from the International Monetary Fund (IMF) to stabilize the Korean currency. Our roof/valley detection procedure finds five roofs/valleys at five time points after December 1997. These results and the corresponding results by procedures Loader and Zero-Crossing are summarized in Table 5, and the results by PP_{BST} are also shown in Figure 6(a) by vertical lines. From Table 5 and Figure 6(a), we can see that the five roofs/valleys detected by PP_{BST} correspond to five places that the exchange rates clearly change their trends. As a comparison, the procedure

Table 5: Detected jumps/roofs/valleys by various methods for the daily exchange rates data.

Methods	Detected Jumps	Detected Roofs/Valleys
PP_{BST}	12/01/1997	01/09/1998, 05/01/1998, 05/15/1998, 07/29/1998, 09/24/1998
Loader	05/13/1998	01/08/1998, 02/03/1998, 04/03/1998, 05/13/1998, 07/23/1998, 09/21/1998
Zero-Crossing	05/29/1997, 12/02/1997, 05/20/1998, 08/31/1998	01/28/1998, 02/09/1998, 03/06/1998, 04/14/1998, 05/15/1998, 07/30/1998, 09/04/1998, 10/05/1998, 11/30/1998

Loader could not detect the sharp jump around December 01, 1997, several detected jumps by procedure Zero-Crossing do not seem real (e.g., the one at 05/29/1997), and some of their detected roofs/valleys do not correspond to places that the exchange rates have significant trend shifts (e.g., the ones at 03/06/1998 and 09/04/1998 by Zero-Crossing). We then apply the three curve estimation methods considered in the previous section to this data, for estimating the underlying regression curve of the daily exchange rates. The estimated curve should be useful for economists to study how the daily exchange rates change over time during that period of time and how the exchange rates react to certain monetary policies. The estimated curves are shown in Figure 6(b). From the plot, it can be seen that our proposed procedure $\widehat{f}_{\text{BST}}$ can preserve jumps/roofs/valleys well, and the two alternative procedures $\widehat{f}_{\text{L}}$ and $\widehat{f}_{\text{G}}$ would blur certain roofs/valleys.

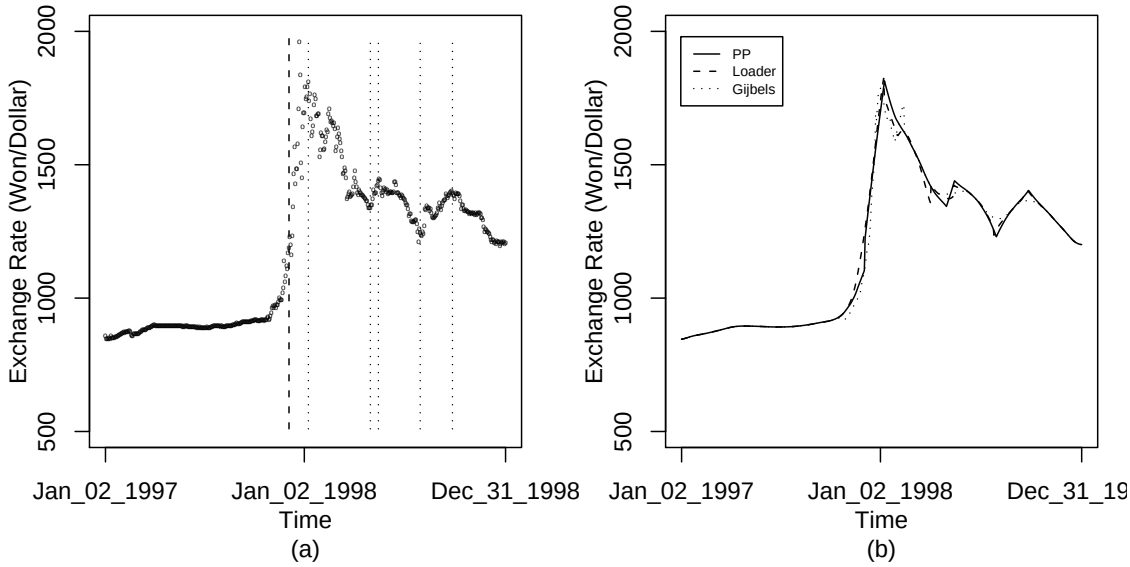

Figure 6: (a) Daily exchange rates data (small circles) and the detected jump (vertical dashed line) and detected roofs/valleys (vertical dotted lines) by PP_{BST}. (b) Curve estimators $\hat{f}_{\text{BST}}$, $\hat{f}_{\text{L}}$ and $\hat{f}_{\text{G}}$.

As a side note, jump (or, shift) detection in a sequential time series is the major research problem of statistical process control (SPC) (cf., e.g., Choi 2008, Hawkins and Olwell 1998, Qiu 2008). There are several essential differences between that problem and the jump/roof/valley detection problem considered in this paper. First, in the current problem, the sample size is fixed beforehand; the data are sequential and the sample size keeps increasing in the SPC problem. Second, the mean response is assumed to be a constant before or after a potential jump point in the SPC problem; it is considered as a nonparametric curve in the current problem. Third, the major goal of a SPC procedure is to signal a jump as soon as possible after it occurs when its false alarm rate is kept under a fixed level; the major goal in the current problem is to detect jumps/roofs/valleys accurately. Generally speaking, SPC procedures can not be used for handling the current problem because they have made use of the special data structure that is commonly assumed in the SPC problem but invalid in the current problem. On the other hand, our proposed jump/roof/valley detection procedures would not be efficient to handle the SPC problem because we have not accommodated the special data structure that is commonly assumed in the SPC problem. There is some other existing research on jump (or, change point) detection in analysing time series data with fixed sample sizes (e.g., Horváth 2001, Lin 2007), most of which assumes that the number of jumps is known and the mean response follows a parametric model. Our proposed jump/roof/valley detection procedures do not make such assumptions.

6 Concluding Remarks

We have presented procedures for detecting jumps/roofs/valleys in regression curves in cases when the number of jumps/roofs/valleys is unknown. Jump/roof/valley preserving curve estimation is also discussed, along with practical selection of procedure parameters. Some theoretical justifications and numerical studies show that these procedures work reasonably well in applications. Computer codes of the proposed procedures are available from the authors upon request.

In this paper, procedure parameters are chosen to be the same in the entire design interval, which may not be ideal for certain applications. Intuitively, in regions where the underlying regression curve has relatively large curvature, the bandwidth, for instance, should be chosen relatively small to reduce possible estimation bias, and it should be chosen relatively large in regions where the regression curve is quite flat. Selection of variable procedure parameters requires much future research. Also, we did not discuss jump/roof/valley detection in boundary regions $[0, h_n)$ and

$(1-h_n, 1]$ in this paper. When the sample size is limited, this boundary jump detection problem could be important, which also requires much future research effort.

Acknowledgments: The authors thank the editor, the associate editor, and two anonymous referees for many constructive comments and suggestions, which greatly improved the quality of the paper. This research is supported in part by an NSF grant.

Appendix

In Appendix, we outline the proofs of seven theorems discussed in Section 3. More complete proofs of these theorems are available online as supplemental materials.

A Outline of the Proof of Theorem 1

For r_k used in equation (4), it is easy to check that

$$\frac{r_k}{nh_n^{k+1}} = \nu_k + O\left(\frac{1}{nh_n}\right), \text{ for } k = 0, 1, 2, 3, 4, \tag{10}$$

where $\nu_k = \int_{-1}^{1} u^k K(u)\, du$. Since K is a symmetric density function, $\nu_0 = 1$ and $\nu_1 = \nu_3 = 0$. By (4) and (10), for a given point $x \in \Omega_n$, we have

$$\begin{aligned} E(\widehat{a}(x)) &= \int_{-1}^{1} \frac{\nu_4 - \nu_2 u^2}{\nu_4 - \nu_2^2} \left\{ f(x) + f'(x)uh_n + \frac{f''(x)}{2} u^2 h_n^2 + o(h_n^2) \right\} K(u)\, du + O\left(\frac{1}{nh_n}\right) \\ &= f(x) + o(h_n^2) + O\left(\frac{1}{nh_n}\right). \end{aligned}$$

So,

$$|E(\widehat{a}(x)) - f(x)| = o(h_n^2) + O\left(\frac{1}{nh_n}\right). \tag{11}$$

On the other hand, it is easy to check that

$$\widehat{a}(x) - E(\widehat{a}(x)) = \frac{\nu_4}{\nu_4 - \nu_2^2}\xi_0(x) - \frac{\nu_2}{\nu_2 - \nu_2^2}\xi_2(x) + O\left(\frac{1}{nh_n}\right), \tag{12}$$

where $\xi_k(x) = \sum_{i=1}^{n} (\frac{x_i - x}{h_n})^k K(\frac{x_i - x}{h_n})\varepsilon_i$, for $k = 0, 1, 2$. By (A.23) in Gijbels *et al.* (2007), we have

$$\sup_{x \in [h_n, 1-h_n]} \sqrt{\frac{nh_n}{ln(n)}} |\xi_k(x)| = O(1), \ a.s., \text{ for } k = 0, 1, 2. \tag{13}$$

So, by (12), (13), and the condition that $1/(nh_n^4) = o(1)$, we have

$$|\widehat{a}(x) - E(\widehat{a}(x))| = O\left(\sqrt{\frac{ln(n)}{nh_n}}\right), a.s.. \tag{14}$$

Notice that both (11) and (14) hold uniformly for all $x \in \Omega_n$. Therefore, we have

$$\| \widehat{a} - f \|_{\Omega_n} = o(h_n^2) + O\left(\sqrt{\frac{ln(n)}{nh_n}}\right), a.s..$$

The first equation in Theorem 1 regarding $\widehat{a}$ has been proved.

To prove the second equation in the theorem regarding $\widehat{b}$, by (4) and (10), we have

$$\begin{aligned} E(\widehat{b}(x)) &= \frac{1}{h_n}\int_{-1}^{1} \frac{u}{\nu_2} f(x+uh_n)K(u)\, du + O\left(\frac{1}{nh_n^2}\right) \\ &= f'(x) + o(h_n) + O\left(\frac{1}{nh_n^2}\right) \end{aligned} \tag{15}$$

So,

$$|E(\widehat{b}(x)) - f'(x)| = o(h_n) + O\left(\frac{1}{nh_n^2}\right). \tag{16}$$

By (4), (13) and and the condition that $1/(nh_n^4) = o(1)$, we have

$$\widehat{b}(x) - E(\widehat{b}(x)) = \frac{1}{h_n}\left[\frac{1}{\nu_2}\xi_1(x) + O\left(\frac{1}{nh_n}\right)\xi_0(x)\right] = O\left(\sqrt{\frac{ln(n)}{nh_n^3}}\right),\ a.s. \tag{17}$$

Therefore, by (16) and (17), which are uniformly true for $x \in \Omega_n$, we have

$$\| \widehat{b} - f' \|_{\Omega_n} = o(h_n) + O\left(\sqrt{\frac{ln(n)}{nh_n^3}}\right), a.s.$$

To prove the third equation in Theorem 1 regarding $\widehat{c}$, from (4), we have

$$\begin{aligned} E(\widehat{c}(x)) &= \frac{2}{h_n^2}\int_{-1}^{1} \frac{-\nu_2 + u^2}{\nu_4 - \nu_2^2} f(x+uh_n)K(u)\, du + O\left(\frac{1}{nh_n^3}\right) \\ &= f''(x) + o(h_n) + O\left(\frac{1}{nh_n^3}\right). \end{aligned} \tag{18}$$

So,

$$|E(\widehat{c}(x)) - f''(x)| = o(h_n) + O\left(\frac{1}{nh_n^3}\right). \tag{19}$$

Meanwhile, by (4), (13) and the condition that $1/(nh_n^4) = o(1)$, we have

$$\widehat{c}(x) - E(\hat{c}(x)) = \frac{1}{h_n^2}\left[\frac{-\nu_2}{\nu_4 - \nu_2^2}\xi_0(x) + \frac{1}{\nu_4 - \nu_2^2}\xi_2(x) + O\left(\frac{1}{nh_n}\right)\xi_0(x)\right] = O\left(\sqrt{\frac{ln(n)}{nh_n^5}}\right) \tag{20}$$

Therefore, by (19), (20) and the condition that $1/(nh_n^4) = o(1)$, we have

$$\| \hat{c} - f'' \|_{\Omega_n} = o(h_n) + O\left(\sqrt{\frac{ln(n)}{nh_n^5}}\right), a.s..$$

B Outline of the Proof of Theorem 2

To prove result (i), from expression (15) and the condition that $1/(nh_n) = o(1)$, we have

$$\begin{aligned} E(\widehat{b}(x)) &= \frac{1}{h_n}\left\{\int_{Q^{(1)}} + \int_{Q^{(2)}}\right\} \frac{u}{\nu_2} f(x+uh_n)K(u)\, du + O\left(\frac{1}{nh_n^2}\right) \\ &= \frac{f_-(s)\int_{Q^{(1)}} uK(u)du + f_+(s)\int_{Q^{(2)}} uK(u)du}{h_n\nu_2} + o\left(\frac{1}{h_n}\right). \end{aligned}$$

Since $\int_{-1}^{1} uK(u)\, du = 0$, $\int_{Q^{(1)}} uK(u)du = -\int_{Q^{(2)}} uK(u)du$. So, we have

$$E(\widehat{b}(x)) = \frac{d}{h_n\nu_2}\int_{Q^{(2)}} uK(u)du + o\left(\frac{1}{h_n}\right). \tag{21}$$

Therefore, by (17), (21) and the condition that $ln(n)/(nh_n^3) = o(1)$, we have

$$\widehat{b}(x) = \frac{d}{h_n\nu_2}\int_{Q^{(2)}} uK(u)du + o\left(\frac{1}{h_n}\right), a.s.$$

To prove result (ii) in Theorem 2, by (18) and the condition that $1/(nh_n) = o(1)$, we have

$$\begin{aligned} E(\widehat{c}(x)) &= \frac{2}{h_n^2}\left\{\int_{Q^{(1)}} + \int_{Q^{(2)}}\right\} \frac{-\nu_2 + u^2}{\nu_4 - \nu_2^2} f(x+uh_n)K(u)\, du\} + O\left(\frac{1}{nh_n^3}\right) \\ &= \frac{2d}{h_n^2(\nu_4 - \nu_2^2)}\left\{\int_{Q^{(2)}} u^2K(u)du - \nu_2\int_{Q^{(2)}} K(u)du\right\} + O\left(\frac{1}{h_n}\right). \end{aligned} \tag{22}$$

In the above expression, we have used the result that $1/(nh_n^2) = o(1/(nh_n^3)) = o(1)$ and the facts that $\int_{Q^{(1)}} K(u)du = 1 - \int_{Q^{(2)}} K(u)du$ and $\int_{Q^{(1)}} u^2K(u)du = \nu_2 - \int_{Q^{(2)}} u^2K(u)du$.

Next, we notice that $\nu_4 - \nu_2^2 = \int_{-1}^{1} u^4K(u)du - (\int_{-1}^{1} u^2K(u)du)^2 > 0$, by the Hölder's inequality. Also, let $\rho(\tau) = \frac{\int_0^\tau u^2K(u)du}{2\int_0^\tau K(u)du}$. Then, ρ is a strictly increasing function of τ on $(0,1)$, and it is continuous from left at 1. Consequently, $\rho(\tau) < \rho(1)$ for any $\tau \in (0,1)$, which implies that

$$\int_0^\tau u^2K(u)\, du < 2\int_0^\tau K(u)\, du \int_0^1 u^2K(u)\, du = \nu_2\int_0^\tau K(u)\, du.$$

Therefore, for any $\tau \in (0,1)$,

$$C := \frac{2d}{\nu_4 - \nu_2^2}\left[\nu_2\int_0^\tau K(u)\, du - \int_0^\tau u^2K(u)\, du\right] > 0 \tag{23}$$

By (20), (22) and (23), we have result (ii) in the theorem.

C Outline of the Proof of Theorem 3

To prove result (i) of Theorem 3, from expression (18) and the condition that $1/(n\tilde{h}_n^2) = o(1)$, we have

$$E(\widehat{c}(x)) = \frac{2}{\tilde{h}_n^2}\left\{\int_{\tilde{Q}^{(1)}} + \int_{\tilde{Q}^{(2)}}\right\} \frac{-\nu_2 + u^2}{\nu_4 - \nu_2^2} f(x+u\tilde{h}_n)K(u)\, du + O\left(\frac{1}{n\tilde{h}_n^3}\right)$$

$$= \frac{2\tilde{d}}{\tilde{h}_n(\nu_4-\nu_2^2)}\left\{\int_{\tilde{Q}^{(2)}} u^3K(u)du - \nu_2\int_{\tilde{Q}^{(2)}} uK(u)du\right\} + o\left(\frac{1}{\tilde{h}_n}\right). \tag{24}$$

By (20), (24) and the condition that $ln(n)/(n\tilde{h}_n^5) = o(1)$, we have result (i) of Theorem 3.

To prove result (ii), first, we have the following expressions for $\hat{c}_+(x)$ and $\hat{c}_-(x)$:

$$\widehat{c}_+(x) = 2\sum_{i=1}^n \frac{(r_1^+r_3^+-(r_2^+)^2)+(r_1^+r_2^+-r_0^+r_3^+)(x_i-x)+(r_0^+r_2^+-(r_1^+)^2)(x_i-x)^2}{r_0^+r_2^+r_4^+-(r_2^+)^3+2r_1^+r_2^+r_3^+-(r_1^+)^2r_4^+-r_0^+(r_3^+)^2}K_r\left(\frac{x_i-x}{\tilde{h}_n}\right)Y_i,$$

$$\widehat{c}_-(x) = 2\sum_{i=1}^n \frac{(r_1^-r_3^--(r_2^-)^2)+(r_1^-r_2^--r_0^-r_3^-)(x_i-x)+(r_0^-r_2^--(r_1^-)^2)(x_i-x)^2}{r_0^-r_2^-r_4^--(r_2^-)^3+2r_1^-r_2^-r_3^--(r_1^-)^2r_4^--r_0^-(r_3^-)^2}K_l\left(\frac{x_i-x}{\tilde{h}_n}\right)Y_i,$$

where $r_k^+ = \sum_{i=1}^n (x_i-x)^k K_r\left(\frac{x_i-x}{\tilde{h}_n}\right)$ and $r_k^- = \sum_{i=1}^n (x_i-x)^k K_l\left(\frac{x_i-x}{\tilde{h}_n}\right)$, for $k=0,1,2,3,4$, and K_r and K_l are defined by $K_r(u) = 2K(u)I(0<u\le 1)$ and $K_l(u) = 2K(u)I(-1\le u<0)$.

Similar to (10), it is easy to check that, for $k=0,1,2,3,4$,

$$\frac{r_k^+}{n\tilde{h}_n^{k+1}} = \nu_k^+ + O\left(\frac{1}{n\tilde{h}_n}\right), \quad \frac{r_k^-}{n\tilde{h}_n^{k+1}} = \nu_k^- + O\left(\frac{1}{n\tilde{h}_n}\right), \tag{25}$$

where $\nu_k^+ = \int_0^1 u^k K_r(u)\,du$ and $\nu_k^- = \int_{-1}^0 u^k K_l(u)\,du$. Also, we can check that $\nu_0^+ = \nu_0^- = 1$ and $\nu_k^- = (-1)^k\nu_k^+$ for $k=0,1,2,3,4$. As in (22), by all the results mentioned above and the condition that $1/(n\tilde{h}_n^2) = o(1)$, we have

$$\begin{aligned}
E(\widehat{c}_+(x)-\widehat{c}_-(x)) &= \frac{2\tilde{d}}{\tilde{h}_n\left[\nu_2^+\nu_4^+-(\nu_2^+)^3+2\nu_1^+\nu_2^+\nu_3^+-(\nu_1^+)^2\nu_4^+-(\nu_3^+)^2\right]} \\
&\left\{(\nu_1^+\nu_3^+-(\nu_2^+)^2)\int_{\tilde{Q}^{(2)}} u(K_r(u)-K_l(u))du + (\nu_1^+\nu_2^+-\nu_3^+)\int_{\tilde{Q}^{(2)}} u^2(K_r(u)+K_l(u))du + \right. \\
&\left.(\nu_2^+-(\nu_1^+)^2)\int_{\tilde{Q}^{(2)}} u^3(K_r(u)-K_l(u))du\right\} + o\left(\frac{1}{\tilde{h}_n}\right) =: \frac{\tilde{C}}{\tilde{h}_n} + o\left(\frac{1}{\tilde{h}_n}\right).
\end{aligned}$$

In the above expression, we have used the facts that $\int_{\tilde{Q}^{(1)}} u^k K_r(u)du = \nu_k^+ - \int_{\tilde{Q}^{(2)}} u^k K_r(u)du$ and $\int_{\tilde{Q}^{(1)}} u^k K_l(u)du = \nu_k^- - \int_{\tilde{Q}^{(2)}} u^k K_l(u)du$, for $k=0,1,2,3$. As in (23), by the Hölder's inequality, we can show that $\tilde{C}>0$ for any $\tau\in(0,1)$. Therefore, we have

$$E(\widehat{c}_+(x)-\widehat{c}_-(x)) = \begin{cases} o(\frac{1}{\tilde{h}_n}), & \text{if } \tau=0, \\ \frac{\tilde{C}}{\tilde{h}_n}+o\left(\frac{1}{\tilde{h}_n}\right), & \text{if } 0<\tau<1 \text{ and } x\in\tilde{Q}_n^{(1)}(x), \\ -\frac{\tilde{C}}{\tilde{h}_n}+o\left(\frac{1}{\tilde{h}_n}\right), & \text{if } 0<\tau<1 \text{ and } x\in\tilde{Q}_n^{(2)}(x). \end{cases} \tag{26}$$

On the other hand, as in (20), $(\widehat{c}_+(x)-\widehat{c}_-(x)) - E(\widehat{c}_+(x)-\widehat{c}_-(x))$ can be expressed as a linear combination of $\xi_k^+(x) = \frac{\sum_{i=1}^n (\frac{x_i-x}{h_n'})^k K_r(\frac{x_i-x}{h_n'})\varepsilon_i}{\sum_{i=1}^n K_r(\frac{x_i-x}{h_n'})}$ and $\xi_k^-(x) = \frac{\sum_{i=1}^n (\frac{x_i-x}{h_n'})^k K_l(\frac{x_i-x}{h_n'})\varepsilon_i}{\sum_{i=1}^n K_l(\frac{x_i-x}{h_n'})}$, for $k=0,1,2$. As in (13), it can be checked that, for $k=0,1,2$,

$$\sup_{x\in[h_n',1-h_n']}\sqrt{\frac{nh_n'}{ln(n)}}|\xi_k^+(x)| = O(1),\ a.s., \quad \sup_{x\in[h_n',1-h_n']}\sqrt{\frac{nh_n'}{ln(n)}}|\xi_k^-(x)| = O(1),\ a.s.. \tag{27}$$

Therefore, we have

$$(\widehat{c}_+(x) - \widehat{c}_-(x)) - E\left(\widehat{c}_+(x) - \widehat{c}_-(x)\right) = O\left(\sqrt{\frac{ln(n)}{n\tilde{h}_n^5}}\right), a.s.. \tag{28}$$

By (26), (28) and the condition that $ln(n)/(n\tilde{h}_n^5) = o(1)$, we have proved result (ii) of the theorem.

D Outline of the Proof of Theorem 4

First, by (4) and the condition that $1/(nh_n) = o(1)$, we have

$$\widehat{\sigma}_n^2 = \frac{\widehat{\sigma}^2}{nh_n^3}\int_{-1}^{1}\frac{u^2}{\nu_2^2}K^2(u)\,du + O\left(\frac{1}{n^2h_n^4}\right) = O\left(\frac{1}{nh_n^3}\right), a.s., \tag{29}$$

where $\widehat{\sigma}^2$ is a consistent estimator of σ^2. Second, it can be show that

$$u_n \sim u_n^* = \widehat{\sigma}_n\sqrt{\chi^2_{1,1-\alpha_n}} = \widehat{\sigma}_n Z_{1-\alpha_n/2}, a.s., \tag{30}$$

where "$\sim$" means that the two terms on its two sides have the same order.

By Theorem 1, we know that $\| \widehat{b} - f' \|_{\Omega_n} = O(\sqrt{ln(n)/(nh_n^3)})$, a.s. By this result, results in (29) and (30), and the condition that $Z_{1-\alpha_n/2}/\sqrt{nh_n} = o(1)$, we have $|\widehat{b}(x)| < u_n$, a.s., which is uniformly true for $x \in \Omega_n$. Therefore, when n is large enough,

$$\widehat{J}_n \subseteq O_n(J), a.s., \tag{31}$$

where $O_n(J) = \bigcup_{j=1}^{p}[s_j - h_n, s_j + h_n]$ is the union of the neighborhoods of all jump points.

On the other hand, when x is one of the true jump points, from result (i) of Theorem 2, we can see that $|\widehat{b}(x)| = C_1/h_n + o(1/h_n)$ where $C_1 = (d/\nu_2)\int_0^1 uK(u)\,du \neq 0$. Therefore, when n is large enough, $|\widehat{b}(x)| > u_n = o(1/h_n)$, a.s., by (29), (30), and the condition that $Z_{1-\alpha_n/2}/\sqrt{nh_n} = o(1)$. Furthermore, by result (ii) of Theorem 2 and the two conditions on v_n stated in Theorem 4, we have that, when n is large enough, $|\widehat{c}(x)| \leq v_n, |\widehat{c}(x_{i_1})| > v_n, |\widehat{c}(x_{i_2})| > v_n$, and $\widehat{c}(x_{i_1})\widehat{c}(x_{i_2}) < 0$, where x_{i_1} and x_{i_2} are two points in $[x - h_n, x + h_n]$ that satisfy the following conditions: (i) x_{i_1} is on the left side of x, (ii) x_{i_2} is on the right side, and (iii) both x_{i_1} and x_{i_2} are $\tau^* h_n$ away from x with $\tau^* > 0$ fixed. Therefore, x would be detected, which implies that, when n is large enough,

$$J \subseteq \widehat{J}_n, a.s. \tag{32}$$

By combining results (31) and (32), we have $\widehat{d}_H(\widehat{J}_n, J; h_n, v_n) = O(h_n)$.

E Proof of Theorem 5

Proof of this theorem is similar to that of Theorem 4. It is therefore omitted.

F Outline of the Proof of Theorem 6

When n is large enough, by Theorem 4, $\widehat{s}_j \in [s_j - h_n, s_j + h_n]$, a.s.. Without loss of generality, we only consider the case when $j = 1$, and assume that $\widehat{s}_1 = s_1 - \delta h_n$ with $\delta \in [0, 1]$. In such a case, by Theorem 1,

$$\begin{aligned} |\widehat{a}_-(\widehat{s}_1) - f_-(s_1)| &\leq |\widehat{a}_-(\widehat{s}_1) - f_-(\widehat{s}_1)| + |f_-(\widehat{s}_1) - f_-(s_1)| \\ &= O(h_n) + o(h_{n,1}^2) + O\left(\sqrt{\frac{ln(n)}{nh_{n,1}}}\right). \end{aligned} \tag{33}$$

On the other hand, similar to (11) and (14), we have

$$\begin{aligned} \widehat{a}_+(\widehat{s}_1) &= \frac{\left(\sum_{x_i \in [\widehat{s}_1, s_1]} + \sum_{x_i > s_1}\right) \frac{\nu_4^+ - \nu_2^+ (\frac{x_i - \widehat{s}_1}{h_{n,1}})^2}{\nu_4^+ - (\nu_2^+)^2} f(x_i) K\left(\frac{x_i - \widehat{s}_1}{h_{n,1}}\right)}{\sum_{i=1}^n K\left(\frac{x_i - \widehat{s}_1}{h_{n,1}}\right)} + O\left(\sqrt{\frac{ln(n)}{nh_{n,1}}}\right) \\ &= f_+(s_1) + o(h_{n,1}^2) + O\left(\frac{h_n}{h_{n,1}}\right) + O\left(\sqrt{\frac{ln(n)}{nh_{n,1}}}\right), \ a.s. \end{aligned} \tag{34}$$

By (33) and (34), we have

$$|\widehat{d}_1 - d_1| = O\left(\frac{h_n}{h_{n,1}}\right) + o(h_{n,1}^2) + O\left(\sqrt{\frac{ln(n)}{nh_{n,1}}}\right), \ a.s.$$

Using similar arguments, we can check the results about $\widehat{\tilde{d}}_j$, for $j = 1, 2, \ldots, p$.

G Outline of the Proof of Theorem 7

By Theorems 4 and 5, both $\widehat{p}$ and $\widehat{q}$ would converge almost surely to p and q, respectively. Therefore, when n is large enough, $\widehat{p} = p$ and $\widehat{q} = q$, a.s. By (2) and (6), for any $x \in H_n = [h_n^*, 1 - h_n^*]$, we have

$$\begin{aligned} \widehat{g}(x) &= \sum_{i=1}^n \frac{r_2 - r_1(x_i - x)}{r_0 r_2 - r_1^2} g(x_i) K\left(\frac{x_i - x}{h_n^*}\right) + \\ &\quad \sum_{i=1}^n \left[\sum_{j=1}^p d_j I(x_i > s_j) - \sum_{j=1}^p \widehat{d}_j I(x_i > \widehat{s}_j)\right] \frac{r_2 - r_1(x_i - x)}{r_0 r_2 - r_1^2} K\left(\frac{x_i - x}{h_n^*}\right) + \end{aligned}$$

$$\sum_{i=1}^{n}\left[\sum_{k=1}^{q}\tilde{d}_k I(x>\tilde{s}_k)_+ - \sum_{j=1}^{q}\widehat{\tilde{d}}_k I(x>\widehat{\tilde{s}}_k)_+\right]\frac{r_2-r_1(x_i-x)}{r_0r_2-r_1^2}K\left(\frac{x_i-x}{h_n^*}\right)+$$
$$\sum_{i=1}^{n}\frac{r_2-r_1(x_i-x)}{r_0r_2-r_1^2}\varepsilon_i K\left(\frac{x_i-x}{h_n^*}\right)$$
$$=: A_1(x)+A_2(x)+A_3(x)+A_4(x) \tag{35}$$

Similar to (11), it can be checked that $\|A_1-g\|_{H_n}=O\left((h_n^*)^2\right)+O\left(\frac{1}{nh_n^*}\right)$. So,

$$\int_{h_n^*}^{1-h_n^*}[A_1(x)-g(x)]^2\ dx \le \|A_1-g\|_{H_n}^2 = O\left((h_n^*)^4\right)+O\left(\frac{1}{(nh_n^*)^2}\right). \tag{36}$$

Regarding $A_2(x)$, we can check that

$$\|A_2\|_{H_n\setminus\bigcup_{j=1}^{p}[s_j-h_n-h_n^*,s_j+h_n+h_n^*]} \le \sum_{j=1}^{p}|d_j-\widehat{d}_j| = O\left(\frac{h_n}{h_{n,1}}\right)+o(h_{n,1}^2)+O\left(\sqrt{\frac{ln(n)}{nh_{n,1}}}\right). \tag{37}$$

On the other hand, when $x\in\bigcup_{j=1}^{p}[s_j-h_n-h_n^*,s_j+h_n+h_n^*]$, $|A_2(x)|\le\max_{j=1}^{p}(|d_j|+|\widehat{d}_j|+|d_j-\widehat{d}_j|)$. Therefore,

$$\|A_2\|_{\bigcup_{j=1}^{p}[s_j-h_n-h_n^*,s_j+h_n+h_n^*]}=O(1),\ a.s. \tag{38}$$

By (37) and (38), we have

$$\int_{H_n}[A_2(x)]^2\ dx = \left[O\left(\frac{h_n}{h_{n,1}}\right)+o(h_{n,1}^2)+O\left(\sqrt{\frac{ln(n)}{nh_{n,1}}}\right)\right]^2+O(h_n+h_n^*) \tag{39}$$

Similarly, we have

$$\int_{H_n}[A_3(x)]^2\ dx = (O\left(\frac{\tilde{h}_n}{\tilde{h}_{n,1}}\right)+o(\tilde{h}_{n,1}^2)+O\left(\sqrt{\frac{ln(n)}{n\tilde{h}_{n,1}}}\right))^2+O(\tilde{h}_n+h_n^*). \tag{40}$$

Similar to (14), we have $\|A_4\|_{H_n}=O(\sqrt{ln(n)/(nh_n^*)})$. Therefore,

$$\int_{H_n}[A_4(x)]^2\ dx = O\left(\frac{ln(n)}{nh_n^*}\right). \tag{41}$$

By (36), (39)–(41), and the fact that $(\alpha+\beta)^2\le 2\alpha^2+2\beta^2$, for any $\alpha,\beta\in R$, we have the result in Theorem 7.

H Supplemental Materials

Proofs.pdf: This pdf file contains complete proofs of theorems 1–7 in the paper.

References

Chaudhuri, P., and Marron, J.S. (1999), "SiZer for exploration of structures in curves," *Journal of the American Statistical Association,* **94**, 807–823.

Choi, H., Ombao, H., and Ray, B. (2008), "Sequential Change-Point Detection Methods for Nonstationary Time Series," *Technometrics*, **50**, 40–52.

Chu, C.K., Glad, I.K., Godtliebsen, F., and Marron, J.S. (1998), "Edge-preserving smoothers for image processing (with discussion)," *Journal of the American Statistical Association*, **93**, 526–556.

Clark, J.J. (1989), "Authenticating edges produced by zero-crossing algorithms," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **11**, 43–57.

Devereux, M.B. (1999), "Real exchange rate trends and growth: a model of east asia," *Review of International Economics,* **7**, 509–521.

Eubank, R.L., and Speckman, P.L. (1994), "Nonparametric estimation of functions with jump discontinuities," IMS Lecture Notes, **23**, *Change-Point Problems* (E. Carlstein, H.G. Müller and D. Siegmund eds.), 130–144.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications,* London: Chapman & Hall.

Fleck, M.M. (1992), "Multiple widths yield reliable finite differences," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **14**, 412–428.

Grégoire, G., and Hamrouni, Z. (2002), "Change-point estimation by local linear smoothing," *Journal of Multivariate Analysis*, **83**, 56–83.

Gijbels, I., and Goderniaux, A.-C. (2004), "Bandwidth selection for change point estimation in nonparametric regression," *Technometrics*, **46**, 76–86.

Gijbels, I., Hall, P., and Kneip, A. (1999), "On the estimation of jump points in smooth curves," *The Annals of the Institute of Statistical Mathematics*, **51**, 231–251.

Gijbels, I., Lambert, A., and Qiu, P. (2007), "Jump-preserving regression and smoothing using local linear fitting: a compromise," *Annals of the Institute of Statistical Mathematics*, **59**, 235–272.

Hall, P., and Titterington, M. (1992), "Edge-preserving and peak-preserving smoothing," *Technometrics*, **34,** 429–440.

Hawkins, D.M., and Olwell, D.H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement,* New York: Springer-Verlag.

Horváth, L. (2001), "Change-point detection in long-memory processes," *Journal of Multivariate Analysis,* **78**, 218–234.

Kim, C.S., and Marron, J.S. (2006), "SiZer for jump detection," *Journal of Nonparametric Statistics,* **18**, 13–20.

Lan, Y., Banerjee, M., and Michailidis, G. (2009), "Changepoint estimation under adaptive sampling," *The Annals of Statistics*, in press.

Lin, S. (2007), "Testing for change points in time series models and limiting theorems for NED sequences," *The Annals of Statistics*, **35,** 1213–1237.

Loader, C.R. (1996), "Change point estimation using nonparametric regression," *The Annals of Statistics*, **24,** 1667–1678.

Marr, D., and Hildreth, E. (1980), "Theory of edge detection," *Proceedings of the Royal Society of London. Series B*, **207**, 187–217.

McDonald, J.A., and Owen, A.B.(1986), "Smoothing with split linear fits," *Technometrics*, **28** 195–208.

Müller, Ch.H. (1992), "Robust estimators for estimating discontinuous functions," *Metrika*, **55**, 99–109.

Müller, H.G. (1992), "Change-points in nonparametric regression analysis," *The Annals of Statistics*, **20,** 737–761.

Qiu, P. (1994), "Estimation of the number of jumps of the jump regression functions," *Communications in Statistics-Theory and Methods*, **23,** 2141–2155.

Qiu, P. (1999), "Comparisons of several local smoothing jump detectors in one-dimensional nonparametric regression," *The ASA Proceedings of the Statistical Computing Section*, 150–155.

Qiu, P. (2003), "A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation," *Journal of Nonparametric Statistics,* **15**, 437–453.

Qiu, P. (2005), *Image Processing and Jump Regression Analysis*, New York: John Wiley & Sons.

Qiu, P. (2007), "Jump surface estimation, edge detection, and image restoration," *Journal of the American Statistical Association*, **102**, 745–756.

Qiu, P. (2008), "Distribution-free multivariate process control based on log-linear modeling," *IIE Transactions*, **40**, 664–677.

Qiu, P., Asano, Chi., and Li, X. (1991), "Estimation of jump regression functions," *Bulletin of Informatics and Cybernetics*, **24,** 197–212.

Qiu, P., and Yandell, B. (1998), "A local polynomial jump detection algorithm in nonparametric regression," *Technometrics,* **40,** 141–152.

Scott, D.W. (1992). *Multivariate Density Estimation. Theory, Practice and Visualization*, New York: John Wiley & Sons.

Shiau, J. (1987), "A note on MSE coverage intervals in a partial Spline model," *Communications in Statistics – Theory and Methods*, **16**, 1851–1866.

Sun, J., and Qiu, P. (2007), "Jump detection in regression surfaces using both first-order and second-order derivatives," *Journal of Computational and Graphical Statistics*, **16**, 289–311.

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

Wu, J.S., and Chu, C.K. (1993), "Kernel type estimators of jump points and values of a regression function," *The Annals of Statistics*, **21**, 1545–1566.

Wu, W.B., and Zhao, Z. (2007), "Inference of trends in time series," *Journal of the Royal Statistical Society (Series B)*, **69**, 391–410.