Pak. J. Statist.
200x, Vol. xx(x), xx-xx

ON JUMP DETECTION IN REGRESSION CURVES USING LOCAL POLYNOMIAL KERNEL ESTIMATION

Bo Zhang[1], **Zhihua Su**[2], and **Peihua Qiu**[3]

[1] School of Statistics, University of Minnesota, Minneapolis, USA
Email: dzhang@stat.umn.edu
[2] School of Statistics, University of Minnesota, Minneapolis, USA
Email: suzhihua@stat.umn.edu
[3] School of Statistics, University of Minnesota, Minneapolis, USA
Email: qiu@stat.umn.edu

ABSTRACT

Regression analysis when the underlying regression function has jumps is a research problem with many applications. In practice, jumps often represent structure changes of a related process. Hence, it is important to detect them accurately from observed noisy data. In the literature, there are some jump detectors proposed, most of which are based on local constant or local linear kernel smoothing. For a given application, which method is more appropriate to use? Will local quadratic or local higher-order polynomial kernel smoothing provide a better jump detector in certain cases? All these practical questions have not been well addressed yet. To answer these questions, in this paper, we study both theoretical and numerical properties of jump detectors based on various local polynomial kernel smoothing, and provide certain guidelines on their practical use. Besides a simulation study, two real data examples are presented for demonstrating cases when two specific jump detectors are more appropriate to use, compared to other methods.

KEYWORDS

Bandwidth selection; Bootstrap resampling; Consistency; Curve estimation; Discontinuities; Jumps; Local smoothing; Nonparametric regression.

2000 Mathematics Subject Classification: 62G08

1 INTRODUCTION

Regression analysis provides a major statistical tool for building a functional relationship between response variables and explanatory variables. In certain applications, such a functional relationship has jumps at some unknown positions, representing structural changes of a related process. For instance, stock indices would have abrupt changes after certain unexpected events of great social or economic impact. It has been demonstrated that 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 are an important part of the underlying regression function; accurate detection of them is important for understanding the structural changes of the process and for estimating the regression function properly. This paper focuses on jump detection in regression curves.

In the literature, a number of jump detectors have been proposed. Many of them are based on local constant kernel smoothing (e.g., Qiu 1991, Qiu *et al.* 1991, Müller 1992, Wu and Chu 1993, Qiu 1994, Gijbels *et al.* 1999, Qiu 1999), or local linear kernel smoothing (e.g., Loader 1996, Qiu and Yandell 1998, Grégoire and Hamrouni 2002). A recent jump detector by Joo and Qiu (2009) is based on local quadratic kernel smoothing. Other jump detectors include the partial smoothing spline method by Shiau (1987), the method based on comparison of three local linear estimators by Hall and Titterington (1992), the semi-parametric method by Eubank and Speckman (1994), the wavelet transformation method by Wang (1995), the robust jump detector by Müller (2002), the method for handling time series data by Wu and Zhao (2007), among others. Jump-preserving curve estimation based on local linear kernel smoothing is discussed by Qiu (2003), Gijbels *et al.* (2007), and the references cited therein. See Chapter 3 of Qiu (2005) for a detailed introduction about jump detection and jump-preserving curve estimation.

For a real application, should we use a jump detector based on local constant kernel smoothing, or a method based on local linear kernel smoothing? Would local quadratic or local higher-order polynomial kernel smoothing provide a better jump detector in certain cases? These practical questions have not been well addressed in the literature. In this paper, we study jump detection based on local polynomial kernel smoothing systematically. By investigating both theoretical properties and numerical performance of such edge detectors, certain practical guidelines are provided about their use. Basically, we conclude that (i) it depends on the curvature of the true regression curve to choose local constant, local linear, or local quadratic kernel smoothing for jump detection, (ii) lower order local polynomial kernel smoothing should be used when the curvature is smaller, and (iii) local polynomial kernel smoothing of order 3 or higher would hardly provide good results in most cases.

The rest part of the paper is organized as follows. In next section, jump detection based on general local polynomial kernel smoothing is described in detail. Theoretical properties of the corresponding jump detectors are discussed in Section 3. Selection of procedure parameters is discussed in Section 4. A simulation study is presented in Section 5, where we compare various jump detectors in many different cases. Two real data examples are presented in Section 6. Some concluding remarks are given in Section 7. Proofs of several theorems are provided in the Appendix.

2 Jump Detection by Local Polynomial Kernel Smoothing

Let $\mathcal{D} = \{(X_i,Y_i), i = 1,2,\cdots,n\}$ be n observations from the following regression model:

$$Y_i|\, X_i = f(X_i) + \sigma(X_i)\, \varepsilon_i, \qquad i = 1,2,\cdots,n, \tag{2.1}$$

where f is the unknown regression function, $\sigma(X_i)$ is the standard deviation of the response variable Y at X_i, $\{\varepsilon_i\}$ are independent and identically distributed random errors with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = 1$, $\{X_i\}$ are $[0,1]$-valued design points, and $\{X_i\}$ and $\{\varepsilon_i\}$ are independent of each other. The regression function f is assumed to be

$$f(x) = g(x) + \sum_{j=1}^{J} d_j I(x > s_j), \qquad x \in [0,1], \tag{2.2}$$

where g is a continuous function in the design interval $[0,1]$, $I(u)$ is an indicator function taking the value of 1 when $u =$"true" and 0 otherwise, J denotes the number of jumps in f, $\{s_j, j = 1,2,\cdots,J\}$ are jump locations, and $\{d_j, j = 1,2,\cdots,J\}$ are the corresponding jump magnitudes. In model (2.1) and (2.2), σ, g, J, $\{s_j, j = 1,2,\cdots,J\}$ and $\{d_j, j = 1,2,\cdots,J\}$ are all assumed unknown.

To detect a jump at a given point x, we consider two one-sided neighborhoods $(x, x+h_n]$ and $[x-h_n, x)$ with bandwidth h_n. Then, we construct estimators of the right and left limits of f at x (denoted as $f_+(x)$ and $f_-(x)$) by the following local polynomial kernel smoothing in $(x, x+h_n]$ and $[x-h_n, x)$, respectively:

$$\min_{\beta_0,\beta_1,\cdots,\beta_p} \sum_{i=1}^{n} \left[Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right]^2 K_l\left(\frac{X_i - x}{h_n}\right), \qquad l = 1,2, \tag{2.3}$$

where p is the order of the local polynomial, β_js are coefficients, and K_1 and K_2 are two density kernel functions with supports $(0,1]$ and $[-1,0)$, respectively. The solutions to β_0 of the minimization problem (2.3) when $l=1,2$ are denoted as $\widehat{f}^{+}_{p,h_n}(x)$ and $\widehat{f}^{-}_{p,h_n}(x)$. They are the local p-th-order polynomial kernel estimators of $f_+(x)$ and $f_-(x)$.

By some routine algebraic manipulations, $\widehat{f}^{+}_{p,h_n}(x)$ and $\widehat{f}^{-}_{p,h_n}(x)$ have the following expressions:

$$\begin{aligned}
\widehat{f}^{+}_{p,h_n}(x) &= \sum_{i=1}^{n} \frac{\sum_{j=0}^{p} (w^{(1)}_{j,\,h_n})^{*}(X_i-x)^j}{|\mathbf{W}^{(1)}|} Y_i K_1\left(\frac{X_i-x}{h_n}\right), \\
\widehat{f}^{-}_{p,h_n}(x) &= \sum_{i=1}^{n} \frac{\sum_{j=0}^{p} (w^{(2)}_{j,\,h_n})^{*}(X_i-x)^j}{|\mathbf{W}^{(2)}|} Y_i K_2\left(\frac{X_i-x}{h_n}\right),
\end{aligned} \tag{2.4}$$

where

$$\mathbf{W}^{(l)} = \begin{pmatrix} w^{(l)}_{0,\,h_n} & w^{(l)}_{1,\,h_n} & \cdots & w^{(l)}_{p,\,h_n} \\ w^{(l)}_{1,\,h_n} & w^{(l)}_{2,\,h_n} & \cdots & w^{(l)}_{p+1,\,h_n} \\ \vdots & \vdots & \ddots & \vdots \\ w^{(l)}_{p,\,h_n} & w^{(l)}_{p+1,\,h_n} & \cdots & w^{(l)}_{2p,\,h_n} \end{pmatrix},$$

$$\left(w^{(l)}_{j,\,h_n}\right)^{*} = (-1)^{(j+1)} \begin{vmatrix} w^{(l)}_{1,\,h_n} & \cdots & w^{(l)}_{j,\,h_n} & w^{(l)}_{j+2,\,h_n} & \cdots & w^{(l)}_{p+1,\,h_n} \\ w^{(l)}_{2,\,h_n} & \cdots & w^{(l)}_{j+1,\,h_n} & w^{(l)}_{j+3,\,h_n} & \cdots & w^{(l)}_{p+2,\,h_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ w^{(l)}_{p,\,h_n} & \cdots & w^{(l)}_{j+p-1,\,h_n} & w^{(l)}_{j+p+1,\,h_n} & \cdots & w^{(l)}_{2p,\,h_n} \end{vmatrix},$$

for $l=1,2$, $j=0,1,\cdots,p$, and

$$w^{(l)}_{j,\,h_n} = \sum_{i=0}^{n} (X_i-x)^j K_l\left(\frac{X_i-x}{h_n}\right), \qquad \text{for } j=0,1,2,\ldots,2p,\ \ l=1,2.$$

A natural jump detection criterion is then defined by

$$M_p(x) = \widehat{f}^{+}_{p,h_n}(x) - \widehat{f}^{-}_{p,h_n}(x). \tag{2.5}$$

If x is a jump point, then $|M_p(x)|$ would be relatively large. Otherwise, it would be relatively small. When $p=0$, $M_0(x)$ (or, similar quantities) is the jump detection criterion discussed in references, such as Qiu *et al.* (1991), Müller (1992), Wu and Chu (1993), Qiu (1994), and Gijbels *et al.* (1999), which detect jumps based on local constant kernel smoothing. When $p = 1$, $M_1(x)$ (or, similar quantities) is the jump detection criterion discussed by several authors, including Loader (1996), Qiu and Yandell (1998), and Grégoire and Hamrouni (2002), who suggest detecting jumps based on local linear kernel smoothing. Because $|M_p(x)|$ is based on one-sided local p-th-order polynomial kernel smoothing, k-th-order trend in f, for $k = 1, 2, \ldots, p$, would have little effect on $|M_p(x)|$. For instance, when f is continuous at x with a large slope, $|M_1(x)|$ would still be small, which is one major advantage of jump detection based on local linear kernel smoothing, compared to jump detection based on local constant kernel smoothing (cf., Qiu 1999). Similarly, jump detection based on $|M_2(x)|$ might be more appropriate when f has large curvature at certain places. In the next two sections, we will investigate both theoretically and numerically whether jump detection based on $|M_p(x)|$ would always be improved when p increases.

When the number of jumps J is known, jump positions can be estimated as follows. Let s_j^* be the maximizer of $|M_p(x)|$ over the range $x \in [h_n, 1-h_n] \backslash (\bigcup_{r=1}^{j-1}[s_r^* - h_n, s_r^* + h_n])$, for $j = 1, 2, \ldots, J$. The order statistics of $\{s_j^*, j = 1, 2, \cdots, J\}$ are denoted as $s_{(1)}^* < s_{(2)}^* < \cdots < s_{(J)}^*$. Then, estimators of the jump positions and jump magnitudes can be defined by

$$\widehat{s}_j = s_{(j)}^*, \qquad \widehat{d}_j = M_p(s_{(j)}^*), \qquad \text{for } j = 1, 2, \cdots, J.$$

When the number of jumps J is unknown, jump detection becomes much more challenging. In such cases, some existing methods (e.g., Qiu 1994) use the strategy that only those design points whose signal-to-noise ratio values are above a threshold C_n are flagged as jump points. Based on the fact that $M_p(x)$ is a linear combination of independent observations and on Theorem 1 in Section 3, we know that $M_p(x)$ has the following asymptotic distribution

$$\mathrm{N}\left(f_+(x) - f_-(x), \; \frac{2\sigma^2 \sum_{j=0}^{2p} K_{j,2} \tilde{K}_{j,1}^*}{n h_n |\mathbf{K}|^2} \right),$$

where $K_{j,2}$, $\tilde{K}_{j,1}^*$, and $|\mathbf{K}|$ are constants depending on the kernel function K. Then, it is natural to choose C_n such that $P(|M_p(x)| \geq C_n) \leq \alpha_n$, which results in

$$C_n = Z_{\alpha_n/2} \sqrt{\frac{2\widehat{\sigma}^2 \sum_{j=0}^{2p} K_{j,2} \tilde{K}_{j,1}^*}{n h_n |\mathbf{K}|^2}}, \tag{2.6}$$

where α_n is a significance level and $\widehat{\sigma}^2$ is a consistent estimator of σ^2. By using C_n in (2.6), x would be detected as a jump point if $|M_p(x)| \geq C_n$, or equivalently,

$$|M_p(x)|/\widehat{\sigma} > Z_{\alpha_n/2} \sqrt{2 \frac{\sum_{j=0}^{2p} K_{j,2} \tilde{K}_{j,1}^*}{n h_n |\mathbf{K}|^2}},$$

where $|M_p(x)|/\widehat{\sigma}$ can be regarded as an estimated signal-to-noise ratio at x. If x is a true jump point with jump magnitude d_x, then the probability that it is detected as a jump point is asymptotically

$$P\left(\left|\sqrt{\frac{2\sigma^2\sum_{j=0}^{2p}K_{j,2}\tilde{K}_{j,1}^*}{nh_n|\mathbf{K}|^2}}Z+d_x\right|>C_n\right),$$

where Z denotes a random variable with a standard normal distribution. Note that, in the above expression, both $\sqrt{(2\sigma^2\sum_{j=0}^{2p}K_{j,2}\tilde{K}_{j,1}^*)/(nh_n|\mathbf{K}|^2)}$ and C_n would converge to 0 when n increases, under some regularity conditions given in Section 3. So, x would be detected with probability 1.

When the number of jumps J is unknown, in finite sample cases, certain false jump points would be detected around true jump points if a threshold value, such as C_n in (2.6), is used in jump detection. To delete those false jump points, Qiu (1994) proposed a modification procedure, briefly described below. Let $\{x_i^*, i=1,2,\ldots,m\}$ be the set of detected jump points satisfying

$$|M_p(x_i^*)|\geq C_n, \qquad \text{for } i=1,2,\ldots,m.$$

If there are $r_1<r_2$ such that the distance between any two consecutive points in $\{x_{r_1}^*, x_{r_1+1}^*, \ldots, x_{r_2}^*\}$ is smaller than or equal to h_n, $x_{r_1}^*-x_{r_1-1}^*>h_n$, and $x_{r_2+1}^*-x_{r_2}^*>h_n$, then we say that $\{x_{r_1}^*, x_{r_1+1}^*, \ldots, x_{r_2}^*\}$ forms a tie in $\{x_i^*, i=1,2,\ldots,m\}$ and the entire tie set is replaced by its central point $(x_{r_1}^*+x_{r_2}^*)/2$ for estimating the jump positions. After this modification, the detected jump points and the corresponding jump magnitudes are denoted as

$$\widehat{s}_j, \text{ and } \widehat{d}_j=M_p(\widehat{s}_j), \qquad \text{for } j=1,2,\ldots,\widehat{J}. \tag{2.7}$$

3 Statistical Properties of Jump Detection by $M_p(x)$

This section discusses certain properties of the jump detection criterion $M_p(x)$ and its detected jumps (cf., expressions (2.5) and (2.7)). Theorem 1 below shows that $M_p(x)$ would not be affected by the first p derivatives of the true regression function f.

Theorem 1. *In regression model (2.1), assume that f has right and left $(p+1)$-st-order derivatives at the jump points $\{s_j\}$, and that f is $(p+1)$-st-order differentiable at any other points in $[0,1]$. The bandwidth h_n satisfies the conditions that $h_n=o(1)$ and $1/(nh_n)=o(1)$. The two one-sided kernel functions satisfy the condition that $K_1(x)=$*

$K_2(-x)$, for $x \in (0,1]$, and they are both Lipschitz-1 continuous in their supports. Let $K_{t_1,t_2} = \int_0^1 u^{t_1} K_1^{t_2}(u)du$, for $t_1, t_2 = 0, 1, \ldots,$ and

$$\mathbf{K} = \begin{pmatrix} K_{0,1} & K_{1,1} & \cdots & K_{p,1} \\ K_{1,1} & K_{2,1} & \cdots & K_{p+1,1} \\ \vdots & \vdots & \ddots & \vdots \\ K_{p,1} & K_{p+1,1} & \cdots & K_{2p,1} \end{pmatrix}.$$

Then, if x is a true jump point, we have

$$\begin{aligned} & E(M_p(x)) \\ = \quad & (f_+(x) - f_-(x)) + \left(f_+^{(p+1)}(x) - f_-^{(p+1)}(x) \right) \frac{\sum_{j=0}^{p} K_{j,1}^{*} K_{p+j+1,1}}{(p+1)!|\mathbf{K}|} h_n^{p+1} + o(h_n^{p+1}), \end{aligned}$$

and

$$Var(M_p(x)) = \frac{2\sigma^2(x) \sum_{j=0}^{2p} K_{j,2} \tilde{K}_{j,1}^{*}}{nh_n |\mathbf{K}|^2} + o\left(\frac{1}{nh_n} \right),$$

where

$$K_{j,\,1}^{*} = (-1)^{(j+1)} \begin{vmatrix} K_{1,\,1} & \ldots & K_{j,\,1} & K_{j+2,\,1} & \ldots & K_{p+1,\,1} \\ K_{2,\,1} & \ldots & K_{j+1,\,1} & K_{j+3,\,1} & \ldots & K_{p+2,\,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ K_{p,\,1} & \ddots & K_{j+p-1,\,1} & K_{j+p+1,\,1} & \ldots & K_{2p,\,1} \end{vmatrix}, \text{ for } j = 0, 1, \ldots, p,$$

and

$$\tilde{K}_{j,1}^{*} = \sum_{\substack{j_1 + j_2 = j \\ j_1, j_2 = 0, 1, 2, \cdots, p}} K_{j_1,1}^{*} K_{j_2,1}^{*}.$$

Corollary 3.1. Under the assumptions of Theorem 1, if $x \in (0,1)$ is a continuity point of f, then

(i) $\mathrm{E}(M_p(x)) = O(h_n^{p+1})$, and $\mathrm{Var}(M_p(x)) = \frac{2\sigma^2(x) \sum_{j=0}^{2p} K_{j,2} \tilde{K}_{j,1}^{*}}{nh_n |\mathbf{K}|^2} + o\left(\frac{1}{nh_n} \right)$;

(ii) when $h_n \sim n^{-\frac{1}{2p+3}}$, the mean squared error (MSE) of $M_p(x)$ converges to 0 with the optimal rate $O(n^{-\frac{2p+2}{2p+3}})$.

From result (i) of Corollary 3.1, it seems that it is better to use higher-order local polynomial smoothing in jump detection, because $\mathrm{E}(M_p(x))$ is much smaller when p is larger. However, this is a large-sample result and it is based on the assumption that a same bandwidth is used in $M_p(x)$ when p varies. Under this assumption, when the sample size is finite, $\mathrm{Var}(M_p(x))$ can actually change a lot. For instance, when K_1 is chosen to be the Epanechnikov kernel function $K_1(x) = 1.5(1-x^2)$, for $x \in (0,1]$, and when $p = 0,1,2,3$, we have

$$\mathrm{Var}(M_0(x)) = \frac{2.40\sigma^2(x)}{nh_n} + o\left(\frac{1}{nh_n}\right), \quad \mathrm{Var}(M_1(x)) = \frac{4.49\sigma^2(x)}{nh_n} + o\left(\frac{1}{nh_n}\right),$$

$$\mathrm{Var}(M_2(x)) = \frac{8.89\sigma^2(x)}{nh_n} + o\left(\frac{1}{nh_n}\right), \quad \mathrm{Var}(M_3(x)) = \frac{22.86\sigma^2(x)}{nh_n} + o\left(\frac{1}{nh_n}\right).$$

Thus, in practice, we still need to choose p carefully for different data sets. To this end, some practical guidelines will be provided in the next section, based on a large simulation study.

The next two theorems establish strong consistency of $\widehat{f}^{+}_{p,\,h_n}(x)$, $\widehat{f}^{-}_{p,\,h_n}(x)$, and the detected jumps (cf., expressions (2.4) and (2.7)). First, we state some assumptions.

Assumption A: *Let* ν *be a positive number and* $\{\beta_n\}_{n=1}^{\infty}$ *be a series of positive numbers such that* $1/\beta_n = o(1)$.

Assumption B: *The bandwidth* h_n *is chosen such that* $h_n = o(1)$, *and* $1/(nh_n) = o(1)$.

Assumption C: ν, β_n, *and* h_n *satisfy the conditions that* $\frac{n^{2\nu}}{n\beta_n h_n} = O(1)$, *and* $\frac{n^{\nu-1/2}}{h_n^2 \beta_n \log n} = o(1)$.

Assumption D: *The design points* $\{x_i, i = 1,2,\ldots,n\}$ *satisfy the condition that* $\max_{1\le i\le n+1}|x_i - x_{i-1}| = O(1/n)$, *where* $x_0 = 0$ *and* $x_{n+1} = 1$.

Assumption E: *For* $1 \le i \le n$, $E(\sigma^2(X_i)\varepsilon_i^2) < M$, *where* M *is a positive number.*

Assumption F: $\frac{Z_{\alpha_n/2}}{(nh_n)^{1/2}} = o(1)$ *and* $\frac{h_n^{1/2}\beta_n \log(n)}{Z_{\alpha_n/2} n^{\nu-1/2}} = o(1)$.

Theorem 2. *Under Assumptions A – E, if* f *is* $(p+1)$*-th-order differentiable in* $[0,1]$ *and the other conditions in Theorem 1 hold, then*

$$\frac{n^{\nu}}{\beta_n \log(n)} \sup_{x\in[h_n/2,1]} \left|\widehat{f}^{+}_{p,\,h_n}(x) - f(x)\right| = o(1) \qquad a.s.\,, \tag{3.1}$$

$$\frac{n^{\nu}}{\beta_n \log(n)} \sup_{x\in[0,1-h_n/2]} \left|\widehat{f}^{-}_{p,\,h_n}(x) - f(x)\right| = o(1) \qquad a.s.\,. \tag{3.2}$$

Theorem 3. *Under Assumptions A – F and all conditions in Theorem 1, we have*

$$\lim_{n\to\infty} \widehat{J} = J \qquad a.s.\ , \tag{3.3}$$

$$\lim_{n\to\infty} \widehat{s}_j = s_j \qquad a.s.\ , \quad \textit{for } j = 1,2,\ldots,J\ , \tag{3.4}$$

$$\lim_{n\to\infty} M_p(\widehat{s}_j) = d_j \qquad a.s.\ , \quad \textit{for } j = 1,2,\cdots,J\ . \tag{3.5}$$

4 Bandwidth Selection

In the jump detection procedure (2.3)–(2.7), there are two parameters h_n and α_n to choose. In this section, we propose a bootstrap procedure for that purpose. For a given observed dataset $\mathcal{D} = \{(x_1,Y_1),(x_2,Y_2),\ldots,(x_n,Y_n)\}$ and given parameters h_n and α_n in (2.3)–(2.7), assume that the estimated jumps are $\widehat{S} = \{\widehat{s}_j, j = 1,2,\ldots,\widehat{J}\}$ and the estimated jump magnitudes are $\{\widehat{d}_j, j = 1,2,\ldots,\widehat{J}\}$. The Hausdorff distance between $\widehat{S}$ and the set of true jumps S is

$$d_H(S,\widehat{S};h_n,\alpha_n) = \max\left\{\sup_{s_1\in S}\inf_{s_2\in\widehat{S}}|s_1-s_2|, \sup_{s_1\in\widehat{S}}\inf_{s_2\in S}|s_1-s_2|\right\}.$$

<u>Bandwidth Selection Procedure</u>

- **Step 1:** Define new observations

  $$\tilde{Y}_i = Y_i - \sum_{j=1}^{\widehat{J}} \widehat{d}_j I(x_i > \widehat{s}_j),\ \text{for } i = 1,2,\ldots,n.$$

  Estimate g by local linear kernel smoothing with bandwidth h_{est} from data $\{(x_i,\tilde{Y}_i), i = 1,2,\ldots,n\}$, and the estimator is denoted as $\widehat{g}$. Then, define residuals

  $$\widehat{\varepsilon}_i = Y_i - \widehat{g}(x_i) - \sum_{j=1}^{\widehat{J}} \widehat{d}_j I(x_i > \widehat{s}_j),\ \text{for } i = 1,2,\ldots,n.$$

- **Step 2:** Obtain B batches of resampled residuals from $\{\widehat{\varepsilon}_i, i = 1,2,\ldots,n\}$, by random selection with replacement; each batch has n values. For the b-th batch of resampled residuals, denoted as $\{\widehat{\varepsilon}_i^*, i = 1,2,\ldots,n\}$, define pseudo-data as follows.

  $$Y_i^* = \widehat{g}(x_i) + \sum_{j=1}^{\widehat{J}} \widehat{d}_j I(x_i > \widehat{s}_j) + \widehat{\varepsilon}_i^*,\ \text{for } i = 1,2,\ldots,n.$$

- **Step 3:** Apply the jump detection procedure (2.3)–(2.7) with parameters h_n and α_n to the b-th pseudo-data, and the set of detected jumps is denoted as $\widehat{S}_b$. Then, the Hausdorff distance $d_H(S, \widehat{S}; h_n, \alpha_n)$ is estimated by

$$\widehat{d}_H(S, \widehat{S}; h_n, \alpha_n, h_{est}) = \frac{1}{B} \sum_{b=1}^{B} d_H(\widehat{S}, \widehat{S}_b; h_n, \alpha_n, h_{est}),$$

  where $d_H(\widehat{S}, \widehat{S}_b; h_n, \alpha_n, h_{est})$ denotes the Hausdorff distance between $\widehat{S}$ and $\widehat{S}_b$, which depends on parameters h_n, α_n, and h_{est}.

- **Step 4:** Parameters h_n and α_n are approximated by the solution of

$$\min_{h_n > 0} \min_{\alpha_n \in [0,1]} \left[\min_{h_{est} > 0} \widehat{d}_H(S, \widehat{S}; h_n, \alpha_n, h_{est}) \right].$$

In the literature, Gijbels and Goderniaux (2004) has discussed bootstrap bandwidth selection for jump detection. Our proposed procedure modifies theirs in several aspects. First, Gijbels and Goderniaux use the same bandwidth for jump detection and for estimating g. In our procedure, two different bandwidths h_n and h_{est} are used for the two purposes. Based on our numerical experience, the bandwidth for jump detection should be chosen smaller than the bandwidth for curve estimation. Therefore, this modification should improve jump detection performance. Second, Gijbels and Goderniaux discuss bandwidth selection in cases when the number of jumps is fixed, and they propose an estimator of the number of jumps using cross-validation. In our proposed procedure, bandwidth selection is discussed in the general case when the number of jumps is unknown, which simplifies its execution.

It should be pointed out that, in the above bandwidth selection procedure, we resample residuals when constructing the bootstrap estimator of the Hausdorff distance $d_H(S, \widehat{S}; h_n, \alpha_n)$, which has been shown effective in the literature (e.g., Gijbels and Goderniaux 2004) in cases when the error terms in model (2.1) are homogeneous. In cases when the error terms are heterogeneous, it might be more reasonable to resample original observation pairs $\{(x_i, Y_i), i = 1, 2, \ldots, n\}$.

5 A Simulation Study

In this section, we present some simulation results regarding jump detection by $M_p(x)$ discussed in the previous sections. In model (2.1), for simplicity, we set $\sigma(X_i) = \sigma$. The regression function f takes one of the following three forms:

$$f_1(x) = \begin{cases} 2/3 - 2x & 0 < x < 1/3 \\ 1 & 1/3 \le x < 2/3 \\ -2\,(x-2/3)\,(x-2) & 2/3 \le x \le 1 \end{cases}$$

$$f_2(x) = \begin{cases} 10 - 30x & 0 < x < 1/3 \\ -360\,(x-1/2)^2 + 11 & 1/3 \le x < 2/3 \\ \exp\left[15\,(x-2/3)\,/2\right] - 1 & 2/3 \le x \le 1 \end{cases}$$

$$f_3(x) = \begin{cases} 72\,(x-1/3)^2 & 0 < x < 1/3 \\ 8\sin(15\pi x) + 1 & 1/3 \le x < 2/3 \\ 25\left[\log(x+1/6) - \log(5/6)\right] & 2/3 \le x \le 1. \end{cases}$$

Three σ values 0.1, 0.25, and 0.5, and four n values 100, 200, 500, and 1000 are considered. The two one-sided kernel functions used in (2.3) are chosen to be

$$K_1(z) = 1.5(1-z^2)I(-1 < z < 0), \qquad K_2(z) = 1.5(1-z^2)I(0 < z < 1).$$

When $n = 100$ and $\sigma = 0.5$, one realization of observations from model (2.1) with the three regression functions is shown in Figure 1. From the figure, we can see that each of the three regression functions has two jumps at $x = 1/3$ and $x = 2/3$, and they have quite different curvature, with f_1 being linear in continuity regions, f_2 being curved in certain regions, and f_3 having large curvature in most part of the design interval.

We then apply the jump detection procedure (2.3)–(2.7) to this example. The procedure parameters h_n and α_n are chosen by the bootstrap procedure discussed in Section 4. Tables 1–3 present averaged Hausdorff distances between the set of detected jumps $\widehat{S}$ and the set of true jumps S, averaged from 100 replications, and the corresponding selected bandwidths (in parentheses), by the jump detection criterion $M_p(x)$ with $p = 0, 1, 2,$ and 3 in various cases. We also considered cases when $p > 3$. But results in such cases were found not to be the best in any of the scenarios considered. Therefore, they are not presented here.

From Table 1, we can see that $M_0(x)$ performs the best in all cases when $f = f_1$. Therefore, it seems reasonable to conclude that, when f is quite straight in continuity regions, jump detection using local constant kernel smoothing would be a good choice. Table 2 shows that, in the case when $f = f_2$, $M_2(x)$ performs the best when σ is relatively small or when n is relatively large (i.e., when $\sigma = 0.1$, or when $\sigma = 0.25$ and $n \ge 200$, or when

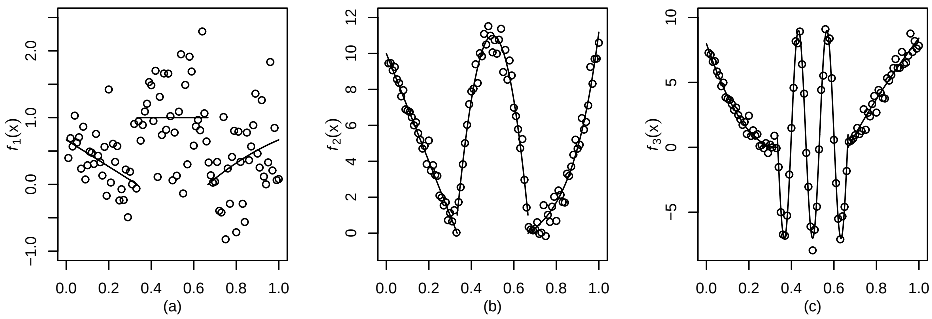

Figure 1: One realization of observations from model (2.1) with the three regression functions when $n = 100$ and $\sigma = 0.5$. (a) f_1, (b) f_2, and (c) f_3.

$\sigma = 0.5$ and $n \geq 500$). Otherwise, $M_1(x)$ performs well. These results imply that, when f has quite large curvature in certain small regions and is straight in majority part of the continuity regions, we can consider using $M_1(x)$ or $M_2(x)$, depending on the sample size and the noise level. If the noise level is low or the sample size is large, then $M_2(x)$ often performs better than $M_1(x)$. Otherwise, $M_1(x)$ would be a good choice. From Table 3, it can be seen that, when f has large curvature in majority part of the continuity regions, $M_2(x)$ or $M_3(x)$ would be good for jump detection. In such cases, if σ is small or n is large, then $M_3(x)$ could be used. Otherwise, $M_2(x)$ would be a good choice.

6 Applications

In this section, we apply the jump detection procedure (2.3)–(2.7) to two real data sets. The first data consist of thickness measures of two US pennies for each year from 1945 to 1989, which are shown in Figure 2(a). The second dataset, shown in Figure 2(b), includes weekly values of the Dow Jones Index Open Price from September 2000 to August 2002. From the plots, it can be seen that penny thickness measure has two jumps around years 1959 and 1975, which was confirmed by Gijbels and Goderniaux (2004), and the Dow Jones Index Open Price has a dramatic jump around the 56-th week, which is the week immediately after September 11, 2001 when the airplane suicide attacks by Al-Qaeda greatly hurt the US and world markets and the Dow Jones Industrial Average index fell 684 points, or 7.1%, on September 17, 2001.

In the jump detection procedure (2.3)–(2.7), parameters h_n and α_n are chosen by the

Table 1: Averaged Hausdorff distances between the set of detected jumps $\widehat{S}$ by jump detection criteria $M_p(x)$, for $p = 0, 1, 2$, and 3, and the set of true jumps S in the case when $f = f_1$. Numbers in parentheses denote selected bandwidths.

		n=100	n=200	n=500	n=1000
σ=0.1	$M_0(x)$	0.0033 (0.04)	0.0015 (0.02)	0.0016 (0.01)	0.0009 (0.002)
	$M_1(x)$	0.0271 (0.14)	0.0168 (0.14)	0.0131 (0.14)	0.0144 (0.14)
	$M_2(x)$	0.0279 (0.14)	0.0192 (0.10)	0.0173 (0.09)	0.0129 (0.14)
	$M_3(x)$	0.0354 (0.16)	0.0177 (0.11)	0.0103 (0.06)	0.0089 (0.07)
σ=0.25	$M_0(x)$	0.0135 (0.01)	0.0044 (0.04)	0.0091 (0.04)	0.0026 (0.01)
	$M_1(x)$	0.0489 (0.18)	0.0559 (0.17)	0.0369 (0.15)	0.0223 (0.15)
	$M_2(x)$	0.0996 (0.47)	0.0349 (0.23)	0.0206 (0.10)	0.0208 (0.14)
	$M_3(x)$	0.2406 (0.41)	0.0734 (0.36)	0.0189 (0.21)	0.0142 (0.10)
σ=0.5	$M_0(x)$	0.0839 (0.32)	0.0341 (0.23)	0.0111 (0.06)	0.0076 (0.03)
	$M_1(x)$	0.1376 (0.36)	0.0705 (0.27)	0.0500 (0.16)	0.0581 (0.16)
	$M_2(x)$	0.3499 (0.57)	0.1624 (0.56)	0.0703 (0.43)	0.0308 (0.16)
	$M_3(x)$	0.3643 (0.85)	0.2694 (0.82)	0.1712 (0.43)	0.0522 (0.34)

bootstrap procedure described in Section 4. The detected jumps by $M_p(x)$ when $p = 0, 1$, and 2 are listed in Table 4. From the table, we can see that, for the penny thickness data which are quite straight in continuity regions, $M_0(x)$ performs well and both $M_1(x)$ and $M_2(x)$ miss one jump. These results are consistent with those from Table 1. The Dow Jones Open Price data look curved in the entire 2-year (i.e., 104-week) range. For this data, jump detection by $M_2(x)$ identifies one jump at the 56-th week, while jump detection by $M_0(x)$ and $M_1(x)$ identifies two and four jumps, respectively, at certain other places. After checking the data carefully, we think that jump detection by $M_2(x)$ might be more reliable in this case.

7 Concluding Remarks

Jump detection in regression curves is important for certain applications. In the literature, a number of jump detectors have been proposed, among which jump detectors based on

Table 2: Averaged Hausdorff distances between the set of detected jumps $\widehat{S}$ by jump detection criteria $M_p(x)$, for $p = 0, 1, 2$, and 3, and the set of true jumps S in the case when $f = f_2$. Numbers in parentheses denote selected bandwidths.

		n=100	n=200	n=500	n=1000
σ=0.1	$M_0(x)$	0.2902 (0.04)	0.2705 (0.03)	0.2637 (0.02)	0.0397 (0.002)
	$M_1(x)$	0.0213 (0.16)	0.0256 (0.15)	0.0329 (0.14)	0.0390 (0.13)
	$M_2(x)$	0.0153 (0.15)	0.0116 (0.12)	0.0134 (0.09)	0.0165 (0.05)
	$M_3(x)$	0.0218 (0.16)	0.0152 (0.14)	0.0098 (0.07)	0.0678 (0.05)
σ=0.25	$M_0(x)$	0.1946 (0.05)	0.2896 (0.05)	0.2962 (0.04)	0.0721 (0.006)
	$M_1(x)$	0.0230 (0.17)	0.0260 (0.16)	0.0245 (0.16)	0.0289 (0.16)
	$M_2(x)$	0.0249 (0.16)	0.0155 (0.16)	0.0103 (0.15)	0.0110 (0.11)
	$M_3(x)$	0.0304 (0.24)	0.0304 (0.19)	0.0167 (0.17)	0.0123 (0.13)
σ=0.5	$M_0(x)$	0.0625 (0.03)	0.2555 (0.06)	0.2735 (0.05)	0.0818 (0.01)
	$M_1(x)$	0.0194 (0.17)	0.0220 (0.16)	0.0299 (0.16)	0.0296 (0.16)
	$M_2(x)$	0.0871 (0.19)	0.0505 (0.18)	0.0242 (0.17)	0.0118 (0.16)
	$M_3(x)$	0.2082 (0.28)	0.1399 (0.27)	0.0351 (0.25)	0.0254 (0.24)

one-sided local constant kernel smoothing and one-sided local linear kernel smoothing are the major ones. In this paper, we discuss jump detection based on the general framework of one-sided local polynomial kernel smoothing. Based on both theoretical and numerical arguments, different jump detectors under this framework are compared, and some practical guidelines are provided. We conclude that (i) when the underlying regression function f is quite straight in continuity regions, jump detection using $M_0(x)$ is recommended, (ii) when f has quite large curvature in certain small regions and is straight in majority part of the continuity regions, jump detection using $M_1(x)$ or $M_2(x)$ is recommended, depending on the sample size and the noise level (if the noise level is relatively low or the sample size is large, then $M_2(x)$ often performs better than $M_1(x)$; otherwise, $M_1(x)$ would be a good choice), and (iii) when f has large curvature in majority part of the continuity regions, $M_2(x)$ or $M_3(x)$ would be good for jump detection.

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

Table 3: Averaged Hausdorff distances between the set of detected jumps $\widehat{S}$ by jump detection criteria $M_p(x)$, for $p = 0, 1, 2$, and 3, and the set of true jumps S in the case when $f = f_3$. Numbers in parentheses denote selected bandwidths.

		n=100	n=200	n=500	n=1000
σ=0.1	$M_0(x)$	0.1213 (0.03)	0.2102 (0.03)	0.1402 (0.02)	0.1116 (0.015)
	$M_1(x)$	0.1755 (0.04)	0.1710 (0.04)	0.1475 (0.03)	0.1715(0.03)
	$M_2(x)$	0.0913 (0.09)	0.0992 (0.09)	0.0876 (0.09)	0.0638 (0.02)
	$M_3(x)$	0.1762 (0.08)	0.0955 (0.05)	0.0172 (0.04)	0.0075 (0.04)
σ=0.25	$M_0(x)$	0.2061 (0.05)	0.1950 (0.04)	0.1668 (0.03)	0.1594 (0.027)
	$M_1(x)$	0.1679 (0.05)	0.1427 (0.04)	0.1710 (0.05)	0.1714 (0.04)
	$M_2(x)$	0.0458 (0.09)	0.0294 (0.09)	0.0345 (0.09)	0.0688 (0.03)
	$M_3(x)$	0.0650 (0.16)	0.0713 (0.06)	0.0300 (0.06)	0.0159 (0.05)
σ=0.5	$M_0(x)$	0.2085 (0.07)	0.2118 (0.06)	0.1603 (0.02)	0.1778 (0.035)
	$M_1(x)$	0.1752 (0.06)	0.1609 (0.05)	0.1714 (0.05)	0.1461 (0.04)
	$M_2(x)$	0.0388 (0.09)	0.0233 (0.09)	0.0252 (0.09)	0.0266 (0.09)
	$M_3(x)$	0.0359 (0.17)	0.0275 (0.16)	0.0254 (0.16)	0.0861 (0.16)

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 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: We thank two referees for many helpful comments which greatly improved the quality of the paper. This research is supported in part by an NSF grant of USA.

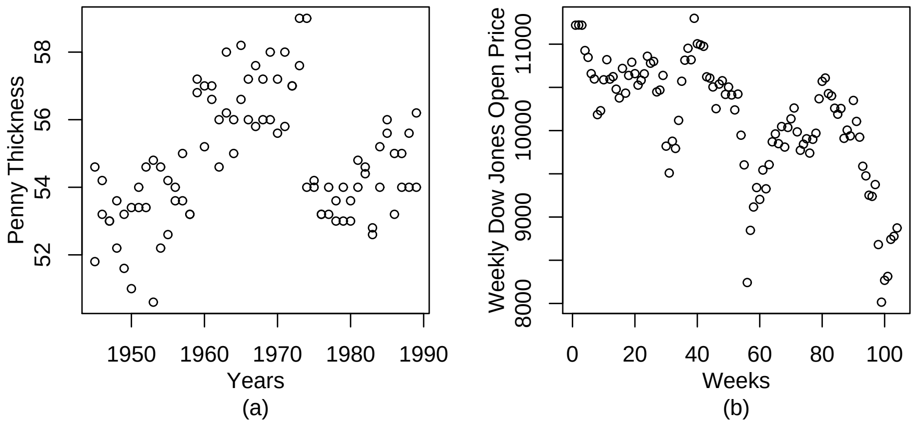

Figure 2: (a) The penny thickness data. (b) The weekly Dow Jones Open Price data.

Table 4: Detected jumps by $M_p(x)$ when $p = 0, 1$, and 2 for the penny thickness data and the weekly Dow Jones Open Price data.

	$M_0(x)$	$M_1(x)$	$M_2(x)$
Penny thickness data	1959, 1975	1975	1974
Dow Jones Open Price data	50, 80	34, 60, 78, 89	56

APPENDIX

A Proof of Theorem 1

In Section 3, we define $K_{t_1,t_2} = \int_0^1 u^{t_1} K_1^{t_2}(u)du = \int_{-1}^0 u^{t_1} K_2^{t_2}(u)du$, for $t_1, t_2 = 0, 1, \ldots$. By integration approximation with summations, for $t_1, t_2 = 0, 1, 2, \ldots, 2p$, we have

$$\frac{1}{nh_n}\sum_{i=0}^{n}\left(\frac{X_i - x}{h_n}\right)^{t_1} K_1^{t_2}\left(\frac{X_i - x}{h_n}\right) = \frac{1}{nh_n}\sum_{i=0}^{n}\left(\frac{X_i - x}{h_n}\right)^{t_1} K_2^{t_2}\left(\frac{X_i - x}{h_n}\right) = K_{t_1,t_2} + o(1).$$

For $E(\widehat{f}^{+}_{p,h_n}(x))$, by some routine algebraic manipulations and the above results, we have

$$\begin{aligned}
E(\widehat{f}^{+}_{p,h_n}(x)) &= \sum_{i=1}^{n} \frac{\sum_{j=0}^{p} (w^{(1)}_{j,h_n})^{*} (X_i - x)^j}{|\mathbf{W}^{(1)}|} K_1\left(\frac{X_i - x}{h_n}\right) f(X_i) \\
&= |\mathbf{K}|^{-1} \sum_{i=1}^{n} \sum_{j=0}^{p} (K_{j,1})^{*} \left(\frac{1}{nh_n}\right) \left(\frac{X_i - x}{h_n}\right)^j K_1\left(\frac{X_i - x}{h_n}\right) f(X_i) \\
&= |\mathbf{K}|^{-1} \sum_{j=0}^{p} (K_{j,1})^{*} \left[\sum_{i=1}^{n} \left(\frac{1}{nh_n}\right) \left(\frac{X_i - x}{h_n}\right)^j K_1\left(\frac{X_i - x}{h_n}\right) f(X_i)\right] \\
&= |\mathbf{K}|^{-1} \sum_{j=0}^{p} (K_{j,1})^{*} \left[\sum_{i=1}^{n} \left(\frac{1}{nh_n}\right) \left(\frac{X_i - x}{h_n}\right)^j K_1\left(\frac{X_i - x}{h_n}\right)\right. \\
&\qquad \left. \times \left(\sum_{s=0}^{p+1} \frac{f^{(s)}_{+}(x)}{s!} (X_i - x)^s + o((X_i - x)^{p+1})\right)\right] \\
&= |\mathbf{K}|^{-1} \sum_{j=0}^{p} \left[(K_{j,1})^{*} \sum_{s=0}^{p+1} \frac{f^{(s)}_{+}(x)}{s!} (h_n)^s \sum_{i=1}^{n} \frac{1}{nh_n} \left(\frac{X_i - x}{h_n}\right)^{(j+s)} K_1\left(\frac{X_i - x}{h_n}\right)\right] \\
&\quad + o(h_n^{p+1}) \\
&= |\mathbf{K}|^{-1} \sum_{j=0}^{p} (K_{j,1})^{*} \left[\sum_{s=0}^{p+1} \frac{f^{(s)}_{+}(x)}{s!} (h_n)^s K_{j+s,1}\right] + o(h_n^{p+1}) \\
&= |\mathbf{K}|^{-1} \sum_{s=0}^{p+1} \frac{f^{(s)}_{+}(x)(h_n)^s}{s!} \left[\sum_{j=0}^{p} (K_{j,1})^{*} K_{j+s,1}\right] + o(h_n^{p+1}) \\
&= f_{+}(x) + f^{(p+1)}_{+}(x) \frac{\sum_{j=0}^{p} K^{*}_{j,1} K_{p+j+1,1}}{(p+1)!|\mathbf{K}|} h_n^{p+1} + o(h_n^{p+1}).
\end{aligned} \tag{A.1}$$

Notice that the last equation holds because it is not difficult to check that

$$\sum_{j=0}^{p} (K_{j,1})^{*} K_{j+s,1} = \begin{cases} 0, & \text{if } 1 \leq s \leq p \\ |\mathbf{K}|, & \text{if } s = 0. \end{cases}$$

Similarly, we have

$$E(\widehat{f}^{-}_{p,h_n}(x)) = f_{-}(x) + f^{(p+1)}_{-}(x) \frac{\sum_{j=0}^{p} K^{*}_{j,1} K_{p+j+1,1}}{(p+1)!|\mathbf{K}|} h_n^{p+1} + o(h_n^{p+1}). \tag{A.2}$$

Therefore, by (A.1) and (A.2),

$$E(M_p(x)) = (f_{+}(x) - f_{-}(x)) + (f^{(p+1)}_{+}(x) - f^{(p+1)}_{-}(x)) \frac{\sum_{j=0}^{p} K^{*}_{j,1} K_{p+j+1,1}}{(p+1)!|\mathbf{K}|} h_n^{p+1} + o(h_n^{p+1}).$$

To prove the result about $Var(M_p(x))$, we rewrite $M_p(x) = \sum_{i=1}^n C_i Y_i$, where

$$C_i = \frac{\sum_{j=0}^p (w_{j,h_n}^{(1)})^* (X_i - x)^j}{|\mathbf{W}^{(1)}|} K_1\left(\frac{X_i - x}{h_n}\right) - \frac{\sum_{j=0}^p (w_{j,h_n}^{(2)})^* (X_i - x)^j}{|\mathbf{W}^{(2)}|} K_2\left(\frac{X_i - x}{h_n}\right).$$

Then,

$$[E(M_p(x))]^2 = \sum_{i=1}^n C_i^2 f^2(X_i) + \sum_{i,j=1}^n C_i C_j f(X_i) f(X_j)\ ,$$

$$E\left[(M_p(x))^2\right] = \sum_{i=1}^n C_i^2 f^2(X_i) + \sum_{i=1}^n C_i^2 \sigma^2(X_i) + \sum_{i,j=1}^n C_i C_j f(X_i) f(X_j)\ .$$

Thus,

$$\begin{aligned}
Var(M_p(x)) &= [E(M_p(x))]^2 - \left[(M_p(x))^2\right] = \textstyle\sum_{i=1}^n \sigma^2(X_i) C_i^2 \\
&= \sum_{i=1}^n \sigma^2(X_i) \left[\frac{\left(\sum_{j=0}^p (w_{j,h_n}^{(1)})^* (X_i - x)^j\right)^2}{|\mathbf{W}^{(1)}|^2} K_1^2\left(\frac{X_i - x}{h_n}\right) \right. \\
&\qquad + \frac{\left(\sum_{j=0}^p (w_{j,h_n}^{(2)})^* (X_i - x)^j\right)^2}{|\mathbf{W}^{(2)}|^2} K_2^2\left(\frac{X_i - x}{h_n}\right) \\
&\quad \left. -2\, \frac{\sum_{j=0}^p (w_{j,h_n}^{(1)})^* (X_i - x)^j}{|\mathbf{W}^{(1)}|} K_1\left(\frac{X_i - x}{h_n}\right) \frac{\sum_{j=0}^p (w_{j,h_n}^{(2)})^* (X_i - x)^j}{|\mathbf{W}^{(2)}|} K_2\left(\frac{X_i - x}{h_n}\right) \right] \\
&= 2\sum_{i=1}^n \sigma^2(X_i) \left[\frac{\left(\sum_{j=0}^p (w_{j,h_n}^{(1)})^* (X_i - x)^j\right)^2}{|\mathbf{W}^{(1)}|^2} K_1^2\left(\frac{X_i - x}{h_n}\right) \right] \\
&= 2\sum_{i=1}^n \sigma^2(X_i) \frac{\sum_{j=0}^{2p} \left[K_1^2(\frac{X_i - x}{h_n})(X_i - x)^j \left(\sum_{\substack{j_1 + j_2 = j \\ j_1, j_2 = 0,1,2,\cdots,p}} (w_{j_1,h_n}^{(1)})^* (w_{j_2,h_n}^{(1)})^* \right) \right]}{|\mathbf{W}^{(1)}|^2} \\
&= 2\sum_{i=1}^n [\sigma^2(x) + o(X_i - x)] \cdot \\
&\qquad \frac{\sum_{j=0}^{2p} \left[K_1^2(\frac{X_i - x}{h_n})(X_i - x)^j \left(\sum_{\substack{j_1 + j_2 = j \\ j_1, j_2 = 0,1,2,\cdots,p}} (w_{j_1,h_n}^{(1)})^* (w_{j_2,h_n}^{(1)})^* \right) \right]}{|\mathbf{W}^{(1)}|^2} \\
&= \frac{2\sigma^2(x)}{nh_n} \cdot \frac{\sum_{j=0}^{2p} K_{j,2} \left(\sum_{\substack{j_1 + j_2 = j \\ j_1, j_2 = 0,1,2,\cdots,p}} K_{j_1,1}^* K_{j_2,1}^* \right)}{|\mathbf{K}|^2} + o\left(\frac{1}{nh_n}\right).
\end{aligned}$$

B Proof of Theorem 2

Here, we only prove equation (3.1). Equation (3.2) can be proved in a similar way. From equation (2.4), by the Taylor's expansion of $f(X_i)$ at x, we have

$$\begin{aligned}
E[\widehat{f}^{+}_{p,h_n}] &= \sum_{i=1}^{n} \frac{\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*}(X_i-x)^j}{|\mathbf{W}^{(1)}|} f(X_i) K_1\left(\frac{X_i-x}{h_n}\right) \\
&= \sum_{i=1}^{n} \frac{\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*}(X_i-x)^j}{|\mathbf{W}^{(1)}|} K_1\left(\frac{X_i-x}{h_n}\right) \left(\sum_{s=0}^{p+1} \frac{f^{(s)}_{+}(x)}{s!}(X_i-x)^s + o((X_i-x)^{p+1})\right) \\
&= f(x) + f^{(p+1)}(x) \sum_{i=1}^{n} \frac{\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*} w^{(1)}_{p+j-1,\,h_n}}{(p+1)!|\mathbf{W}^{(1)}|} h_n^{(p+1)} + o(h_n^{(p+1)}).
\end{aligned}$$

The second equation follows from the facts that $\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*} w^{(1)}_{t+j-1,\,h_n} = 0$, for $1 \le t \le p$, and the definition

$$w^{(l)}_{j,\,h_n} = \sum_{i=0}^{n}(X_i-x)^j K_l\left(\frac{X_i-x}{h_n}\right), \qquad j = 0,1,2,\ldots,2p, \;\; l = 1,2.$$

So,

$$E[\widehat{f}^{+}_{p,h_n}] - f(x) = f^{(p+1)}(x) \sum_{i=1}^{n} \frac{\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*} w^{(1)}_{p+j-1,\,h_n}}{(p+1)!|\mathbf{W}^{(1)}|} h_n^{(p+1)} + o(h_n^{(p+1)}) \;. \tag{B.1}$$

Define

$$v^{(1)}_j = \int_{-1}^{0} x^j K_1(x)dx, \qquad \text{for} \qquad j = 0,1,2,\cdots,p+1.$$

Then, it is easy to check that

$$\frac{w^{(1)}_{j,\,h_n}}{nh_n^{j+1}} = v^{(1)}_j + o(1) \;. \tag{B.2}$$

Now, define

$$\tilde{\varepsilon}_i = \sigma(X_i)\varepsilon_i I(i^{1/2} - |\sigma(X_i)\varepsilon_i|), \qquad i = 1,2,\cdots,n,$$

$$g_n(x) = \sum_{i=1}^{n} K_1\left(\frac{X_i-x}{h_n}\right) \frac{\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*}(X_i-x)^j}{|\mathbf{W}^{(1)}|} \sigma(X_i)\varepsilon_i,$$

$$\tilde{g}_n(x) = \sum_{i=1}^{n} K_1\left(\frac{X_i-x}{h_n}\right) \frac{\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*}(X_i-x)^j}{|\mathbf{W}^{(1)}|} \tilde{\varepsilon}_i,$$

$$\tilde{g}_n(i) = K_1\left(\frac{X_i-x}{h_n}\right) \frac{\sum_{j=0}^{p}(w^{(1)}_{j,\,h_n})^{*}(X_i-x)^j}{|\mathbf{W}^{(1)}|} \tilde{\varepsilon}_i.$$

Then, by the exponential form of Chebyshev inequality, for any ε, we have

$$
\begin{aligned}
& P\left(\frac{n^{\nu}}{\beta_n \log n}\,[\tilde{g}_n(x)-E(\tilde{g}_n(x))] > \varepsilon\right) \\
\leq \quad & \exp\left(\,\log n^{-\varepsilon\beta_n^{1/2}}\right) E\left(\prod_{i=1}^{n} \exp\left(\frac{n^{\nu}}{\beta_n^{1/2}}\,[\tilde{g}_n(i)-E(\tilde{g}_n(i))]\right)\right) \\
\leq \quad & n^{-\varepsilon\beta_n^{1/2}} \exp\left(\frac{n^{2\nu}}{\beta_n}\sum_{i=1}^{n} \mathrm{Var}(\tilde{g}_n(i))\right). \qquad \text{(B.3)}
\end{aligned}
$$

By results (B.2) and Assumption E, it is easy to check that

$$
\sum_{i=1}^{n} \mathrm{Var}(\tilde{g}_n(i)) \;\leq\; M \sum_{i=1}^{n} K_1^2\left(\frac{X_i-x}{h_n}\right)\left[\frac{\sum_{j=0}^{p}(w_{j,\,h_n}^{(1)})^{*}(X_i-x)^j}{|w^{(1)}|}\right]^2 \;=\; \frac{M}{nh_n}\,C_1(K_1),
$$

where M and $C_1(K_1)$ are constants. By assumption C, we have

$$
P\left(\frac{n^{\nu}}{\beta_n \log n}\,[\tilde{g}_n(x)-\mathrm{E}(\tilde{g}_n(x))] > \varepsilon\right) \;=\; O(n^{-\varepsilon\beta_n^{1/2}}), \qquad \text{(B.4)}
$$

which is uniformly true for all $x \in [h_n/2,\, 1]$. We now define $D_n = \{x : |x| \leq n^{1/\delta}+1, x \in \mathrm{R}\}$ for some $\delta > 0$. Let E_n be the smallest subset of $\{i/n^2 : i = 1,2,\ldots,n^2\}$ such that, for any $x \in D_n$, there exists some $Z(x) \in E_n$ satisfying $|x - Z(x)| \leq n^{-2}$. Then, E_n has at most $N_n = [2n^2(n^{1/\delta}+1)]+1$ elements, where $[x]$ denotes the integer part of x. Clearly, we can write

$$
\frac{n^{\nu}}{\beta_n \log n}||\tilde{g}_n - E(\tilde{g}_n)||_{[h_n/2,\,1]\bigcap D_n} \leq\; S_{1n}+S_{2n}+S_{3n}, \qquad \text{(B.5)}
$$

where

$$
\begin{aligned}
S_{1n} &= \frac{n^{\nu}}{\beta_n \log n} \sup_{x\in[h_n/2,\,1]\bigcap D_n} |\tilde{g}_n(x)-\tilde{g}_n(Z(x))| \\
S_{2n} &= \frac{n^{\nu}}{\beta_n \log n} \sup_{x\in[h_n/2,\,1]\bigcap D_n} |\tilde{g}_n(Z(x))-E(\tilde{g}_n(Z(x)))| \qquad \text{(B.6)} \\
S_{3n} &= \frac{n^{\nu}}{\beta_n \log n} \sup_{x\in[h_n/2,\,1]\bigcap D_n} |E(\tilde{g}_n(Z(x)))-E(\tilde{g}_n(x))|\,.
\end{aligned}
$$

From (B.4) and (B.6), we have

$$
P(S_{2n} > \varepsilon) = \; o\left(N_n n^{-\varepsilon\beta_n^{1/2}}\right).
$$

By the Borel-Cantelli Lemma,

$$
\lim_{n\to\infty} S_{2n} = \; 0, \qquad \text{a.s.} \qquad \text{(B.7)}
$$

Now,

$$
\begin{aligned}
S_{1n} &= \frac{n^{\nu}}{\beta_n \log n} \sup_{x\in[h_n/2,\,1]\bigcap D_n} \left| \sum_{i=1}^{n} \left[K_1\left(\frac{X_i - x}{h_n}\right) \frac{\sum_{j=0}^{p} (w^{(1)}_{j,\,h_n})^{*}(X_i - x)^j}{|\mathbf{W}^{(1)}|} \right.\right. \\
&\quad \left.\left. - K_1\left(\frac{X_i - Z(x)}{h_n}\right) \frac{\sum_{j=0}^{p} (w^{(1)}_{j,\,h_n})^{*}(X_i - Z(x))^j}{|\mathbf{W}^{(1)}|} \right] \right| \\
&\leq \frac{n^{\nu+1/2}}{\beta_n \log n} \sup_{x\in[h_n/2,\,1]\bigcap D_n} \left| \frac{1}{h_n} \sum_{i=1}^{n} \left[K_1\left(\frac{X_i - x}{h_n}\right) \mathcal{H}(x) - K_1\left(\frac{X_i - Z(x)}{h_n}\right) \mathcal{H}(Z(x)) \right| \right. \\
&\leq \frac{n^{\nu+1/2}}{\beta_n \log n} \frac{C_2(K_1)}{n^2 h_n}
\end{aligned}
$$

where $C_2(K_1)$ is a constant, and $\mathcal{H}(x) = \sum_{j=0}^{p} (w^{(1)}_{j,\,h_n})^{*}(X_i - x)^j / |\mathbf{W}^{(1)}|$. Therefore

$$\lim_{n\to\infty} S_{1n} = 0 \qquad \text{a.s.} \tag{B.8}$$

Similarly,

$$\lim_{n\to\infty} S_{3n} = 0 \qquad \text{a.s.} \tag{B.9}$$

By (B.7) – (B.9), we have

$$\frac{n^{\nu}}{\beta_n \log n} ||\tilde{g}_n - E(\tilde{g}_n)||_{[h_n/2,\,1]\bigcap D_n} = o(1), \text{ a.s.} \tag{B.10}$$

Now,

$$||g_n - E(g_n)||_{[h_n/2,\,1]} \leq ||g_n - \tilde{g}_n||_{[h_n/2,\,1]} + ||\tilde{g}_n - E(\tilde{g}_n)||_{[h_n/2,\,1]} + ||E(\tilde{g}_n) - E(g_n)||_{[h_n/2,\,1]}.$$

By Assumption E, there exists a full set Ω_0 such that for each $\omega \in \Omega_0$ there exists a finite positive integer N_ω such that, for $i \geq N_\omega$, $\varepsilon_i(\omega) = \sigma(X_i)\tilde{\varepsilon}_i(\omega)$. So, for all $n \geq N_\omega$,

$$|\, g_n(x) - \tilde{g}_n(x)| \leq \frac{1}{nh_n} \sum_{i=1}^{N_\omega} K_1\left(\frac{X_i - x}{h_n}\right) \left| \mathcal{L}(x)(\sigma(X_i)\varepsilon_i - \tilde{\varepsilon}_i) \right| \leq \frac{C(N_\omega)}{nh_n}, \tag{B.11}$$

where $\mathcal{L}(x)$ is a continuous function. Therefore,

$$\frac{n^{\nu}}{\beta_n \log n} ||\tilde{g}_n - \tilde{g}_n||_{[h_n/2,\,1]} = o(1), \qquad \text{a.s.} \tag{B.12}$$

Similarly,

$$\frac{n^{\nu}}{\beta_n \log n} ||E(\tilde{g}_n) - E(g_n)||_{[h_n/2,\,1]} = o(1), \qquad \text{a.s.} \tag{B.13}$$

By (B.10), (B.12) and (B.13), we have

$$\frac{n^{\nu}}{\beta_n \log n} ||g_n - E(g_n)||_{[h_n/2,\,1]} = o(1), \text{ a.s.} \tag{B.14}$$

Thus, by (B.1) and (B.14), we have

$$\frac{n^{\nu}}{\beta_n \log n} ||\widehat{f}^{+}_{p,h_n} - f||_{[h_n/2,\,1]} = o(1), \text{ a.s.}$$

C Proof of Theorem 3

First, by Theorem 2, we have

$$\frac{n^{\nu}}{\beta_n \log(n)} ||M_p||_{[h_n, 1-h_n] \setminus \bigcup_{j=1}^{p} [s_j - h_n, s_j + h_n]} = 0, \qquad \text{a.s.}$$

This result and assumption F imply that, when n is large enough, none of the detected jumps (cf., Section 2) would fall into continuity regions $[h_n, 1-h_n] \setminus \bigcup_{j=1}^{p} [s_j - h_n, s_j + h_n]$.

On the other hand, it is easy to check, based on Theorems 1 and 2, that

$$\lim_{n\to\infty} M_p(s_j) = d_j, \quad \text{for } j = 1, 2, \cdots, p, \qquad \text{a.s.}$$

Therefore, for a given true jump s_j, the design point that is closest to s_j among all design points would be detected as a jump point. In the interval $[s_j - h_n, s_j + h_n]$, there might be multiple detected jumps; but, they would form a tie (cf., discussion at the end of Section 2) and be replaced by the central point of the tie. After the modification procedure, there would be one and only one detected jump in $[s_j - h_n, s_j + h_n]$. After combining the above results, conclusions (3.3)–(3.5) in Theorem 3.3 can be obtained.

REFERENCES

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.

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.

Joo, J., and Qiu, P. (2009), "Jump detection in a regression curve and its derivative," *Technometrics*, in press.

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

Müller, Ch.H. (2002), "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. (1991), "Estimation of a kind of jump regression functions," *Systems Science and Mathematical Sciences*, **4**, 1–13.

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., 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.

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

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.