A Jump-Preserving Curve Fitting Procedure Based On Local Piecewise-Linear Kernel Estimation

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

Abstract

It is known that the fitted regression function based on conventional local smoothing procedures is not statistically consistent at jump positions of the true regression function. In this article, a curve-fitting procedure based on local piecewise-linear kernel estimation is suggested. In a neighborhood of a given point, a piecewise-linear function with a possible jump at the given point is fitted by the weighted least squares procedure with the weights determined by a kernel function. The fitted value of the regression function at this point is then defined by one of the two estimators provided by the two fitted lines (the left and right lines) with the smaller value of the weighted residual sum of squares. It is proved that the fitted curve by this procedure is consistent in the entire design space. In other words, this procedure is jump-preserving. Several numerical examples are presented to evaluate its performance in small-to-moderate sample size cases.

Key Words: Jump-preserving curve fitting; Local piecewise-linear kernel estimation; Local smoothing; Nonparametric regression; Strong consistency.

1 Introduction

Regression analysis provides a tool to build functional relationships between dependent and independent variables. In some applications, regression models with jumps in the regression functions appear to be more appropriate to describe the data. For example, it was confirmed by several statisticians that the annual volume of the Nile river had a jump around year 1899 (Cobb, 1978).

The December sea-level pressure in Bombay India was found to have a jump discontinuity around year 1960 (Shea *at al.* 1994). Some physiological parameters can likewise jump after physical or chemical shocks. As an example, the percentage of time a rat in rapid-eye-movement state in each five-minute interval will most probably have an abrupt change after the lighting condition is suddenly changed (Qiu *et al.* 1999). The objective of this article is to provide a methodology to fit regression curves with jumps preserved.

Suppose that the regression model concerned is

$$y_i = f(x_i) + \epsilon_i, \text{ for } i = 1, 2, \cdots, n, \tag{1.1}$$

where $0 < x_1 < x_2 < \cdots < x_n < 1$ are design points, ϵ_i are i.i.d. random errors with mean 0 and variance σ^2. The regression function $f(\cdot)$ is continuous in $[0, 1]$ except at positions $0 < s_1 < s_2 < \cdots < s_m < 1$ where $f(\cdot)$ has jumps with magnitudes $d_j \neq 0$ for $j = 1, 2, \cdots, m$. Figure 1.1 below presents a case when $m = 2$.

It is known that the fitted curve by the conventional local smoothing procedures is not statistically consistent at positions where $f(\cdot)$ has jumps. For example, the local linear kernel smoother is based on the following minimization procedure (cf. Fan and Gijbels 1996):

$$\min_{a_0^*, a_1^*} \sum_{i=1}^{n} \{y_i - [a_0^* + a_1^*(x_i - x)]\}^2 K(\frac{x_i - x}{h_n}), \tag{1.2}$$

where $K(\cdot)$ is a kernel function with support $[-1/2, 1/2]$ and h_n is a bandwidth parameter. Then the solution of (1.2) for a_0^* is defined as the local linear kernel estimator of $f(x)$. In Figure 1.1, the solid curve denotes the true regression function. It has two jumps at $x = 0.3$ and $x = 0.7$. The dashed curve denotes the conventional fit by the local linear kernel smoothing procedure. It can be seen that "blurring" is present in the curve fitting around the two jumps. As a comparison, the fitted curve by the procedure suggested in this paper is represented by the dotted curve. The two jumps are preserved well by our procedure. More explanation of this plot is given in Section 4.

A major reason for the local linear kernel smoothing procedure (1.2) not to preserve jumps is that it uses a local "continuous" function (a linear function) to approximate the true regression function in a neighborhood of a given point x even if there is a jump at x. A natural idea to overcome this limitation is to fit a local piecewise-linear function at x as follows:

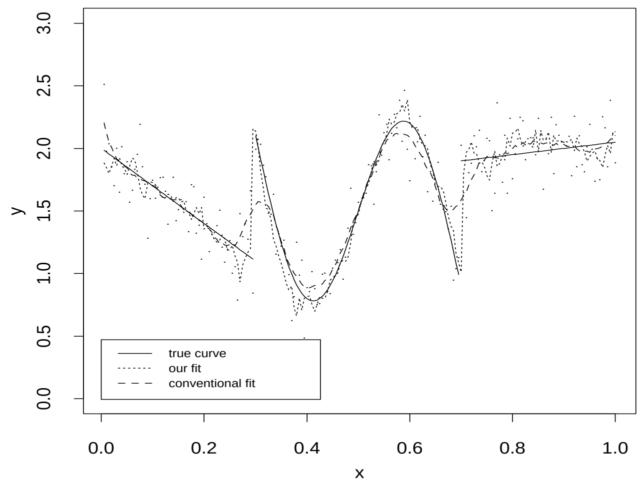

Figure 1.1: Small dots denote noisy data. The solid curve represents the true regression model. The dashed and dotted curves denote the conventional fit by the local linear kernel smoothing procedure and the fit by the procedure suggested in this paper.

$$\min_{a_{l,0},a_{l,1};a_{r,0},a_{r,1}} \sum_{i=1}^{n}\{y_i - [a_{l,0} + a_{l,1}(x_i - x)] - [(a_{r,0} - a_{l,0})I(x_i - x) + (a_{r,1} - a_{l,1})(x_i - x)I(x_i - x)]\}^2 K(\frac{x_i - x}{h_n}), \quad (1.3)$$

where $I(\cdot)$ is an indicator function defined by $I(a) = 1$ if $a \geq 0$ and $= 0$ otherwise. The minimization procedure (1.3) fits a piecewise-linear function $a_{l,0} + a_{l,1}(u - x) + (a_{r,0} - a_{l,0})I(u - x) + (a_{r,1} - a_{l,1})(u - x)I(u - x)$ in $u \in [x - h_n/2, x + h_n/2]$ with a possible jump at x. This is equivalent to fitting two different lines $a_{l,0} + a_{l,1}(u - x)$ and $a_{r,0} + a_{r,1}(u - x)$ in $[x - h_n/2, x)$ and $[x, x + h_n/2]$, respectively. Let $\{\widehat{a}_{l,j}(x), \widehat{a}_{r,j}(x), j = 0, 1\}$ denote the solution of (1.3). Then $\widehat{a}_{l,0}(x)$ and $\widehat{a}_{r,0}(x)$ are estimated from observations in $[x - h_n/2, x)$ and $[x, x + h_n/2]$, respectively. Thus they are good estimators of $f_-(x)$ and $f_+(x)$, the left and right limits of $f(\cdot)$ at x, in the case when x is a jump point. When there is no jump in $[x - h_n/2, x + h_n/2]$, both of them estimate $f(x)$ well. In the case when x itself is not a jump point but a jump point exists in its neighborhood $[x - h_n/2, x + h_n/2]$, only one of $\widehat{a}_{l,0}(x)$ and $\widehat{a}_{r,0}(x)$ provides a good estimator of $f(x)$. Therefore we need to choose one of them as an estimator of $f(x)$ in such case. By combining all these considerations, we define

$$\widehat{f}(x) = \widehat{a}_{l,0}(x)I^*(RSS_r(x) - RSS_l(x)) + \widehat{a}_{r,0}(x)I^*(RSS_l(x) - RSS_r(x)) \quad (1.4)$$

as an estimator of $f(x)$ for $x \in [h_n/2, 1 - h_n/2]$, where $I^*(a)$ is defined by $I^*(a) = 1$ if $a > 0$, 1/2 if $a = 0$ and 0 if $a < 0$; $RSS_l(x)$ and $RSS_r(x)$ are the weighted residual sums of squares (RSS) with respect to observations in $[x - h_n/2, x)$ and $[x, x + h_n/2]$, respectively. That is,

$$RSS_l(x) = \sum_{x_i < x} \{y_i - \widehat{a}_{l,0}(x) - \widehat{a}_{l,1}(x)(x_i - x)\}^2 K(\frac{x_i - x}{h_n});$$

$$RSS_r(x) = \sum_{x_i \geq x} \{y_i - \widehat{a}_{r,0}(x) - \widehat{a}_{r,1}(x)(x_i - x)\}^2 K(\frac{x_i - x}{h_n}).$$

Basically $\widehat{f}(x)$ is defined by one of $\widehat{a}_{l,0}(x)$ and $\widehat{a}_{r,0}(x)$ with the smaller RSS value.

In the literature, there are several existing procedures to fit regression curves with jumps preserved. McDonald and Owen (1996) proposed an algorithm based on three local ordinary least squares estimates of the regression function, corresponding to the observations on the right, left and both sides of a given point, respectively. They then constructed their "split linear fit" as a weighted average of these three estimates, with weights determined by the goodness-of-fit values of the estimates. Hall and Titterington (1992) suggested an alternative but simpler method by establishing some relations among three local linear smoothers and using them to detect the jumps. The regression curve was then fitted as usual in regions separated by the detected jumps. Our procedure is different from these two procedures in that we put the problem to fit regression curves with jumps preserved in the same framework as that of local linear kernel estimation except that a local piecewise-linear function is fitted at a given point in our procedure, making the curve estimator (1.4) easier to use.

Most other jump-preserving curve fitting procedures in the literature consist of two steps: (i) detecting possible jumps under the assumption that the number of jumps is known (it is often assumed to be 1) and (ii) fitting the regression curve as usual in design subintervals separated by the detected jump points. Various jump detectors are based on one-sided constant kernel smoothing (Müller 1992, Qiu *et al.* 1991, Wu and Chu 1993), one-sided linear kernel smoothing (Loader 1996), local least squares estimation (Qiu and Yandell 1998), wavelet transformation (Wang 1995), semiparametric modeling (Eubank and Speckman 1994) and smoothing spline modeling (Koo 1997, Shiau *et al.* 1986). The case when the number of jumps is unknown is considered by several authors including Qiu (1994) and Wu and Chu (1993). They first estimated the number of jumps and jump positions by performing a series of hypothesis tests and then fitted the regression curve in subintervals separated by the detected jump points. Comparing with the above mentioned methods, the method presented in this paper automatically accommodates the jumps in fitting the regression curve without knowing the number of jumps and without performing any hypothesis tests.

This paper is organized as follows. In next section, we discuss the jump-preserving curve fitting procedure (1.4) in some detail. Properties of the fitted curve are discussed in Section 3. In Section 4, we present some numerical examples concerning the goodness-of-fit and bandwidth selection. The procedure is applied to a real-life dataset in Section 5. Section 6 contains some concluding remarks.

2 The Jump-Preserving Curve Fitting Procedure

First we notice that the minimization procedure (1.3) is equivalent to the combination of:

$$\min_{a_{l,0},a_{l,1}} \sum_{i=1}^{n} \{y_i - a_{l,0} - a_{l,1}(x_i - x)\}^2 K_l(\frac{x_i - x}{h_n}) \tag{2.1}$$

and

$$\min_{a_{r,0},a_{r,1}} \sum_{i=1}^{n} \{y_i - a_{r,0}I(x_i - x) - a_{r,1}(x_i - x)I(x_i - x)\}^2 K_r(\frac{x_i - x}{h_n}), \tag{2.2}$$

where $K_l(\cdot)$ is defined by $K_l(x) = K(x)$ if $x \in [-1/2, 0)$ and 0 otherwise and $K_r(\cdot)$ is defined by $K_r(x) = K(x)$ if $x \in [0, 1/2]$ and 0 otherwise. Clearly, (2.1) is equivalent to the local linear kernel smoothing procedure to fit $f_-(x)$ by the observations in $[x - h_n/2, x)$, the left half of $[x - h_n/2, x + h_n/2]$, and (2.2) is equivalent to the local linear kernel smoothing procedure to fit $f_+(x)$ by the observations in $[x, x + h_n/2]$, the right half of $[x - h_n/2, x + h_n/2]$. The subscripts "l" and "r" in notations $\{a_{l,j}, a_{r,j}, j = 0, 1\}$, $K_l(\cdot)$ and $K_r(\cdot)$ represent "left" and "right", respectively, which are also used in other notation defined below.

Solutions of (2.1) and (2.2) can be written as:

$$\widehat{a}_{l,0}(x) = \sum_{i=1}^{n} y_i K_l(\frac{x_i - x}{h_n}) \frac{w_{l,2} - w_{l,1}(x_i - x)}{w_{l,0}w_{l,2} - w_{l,1}^2}$$

$$\widehat{a}_{l,1}(x) = \sum_{i=1}^{n} y_i K_l(\frac{x_i - x}{h_n}) \frac{w_{l,0}(x_i - x) - w_{l,1}}{w_{l,0}w_{l,2} - w_{l,1}^2}$$

$$\widehat{a}_{r,0}(x) = \sum_{i=1}^{n} y_i K_r(\frac{x_i - x}{h_n}) \frac{w_{r,2} - w_{r,1}(x_i - x)}{w_{r,0}w_{r,2} - w_{r,1}^2}$$

$$\widehat{a}_{r,1}(x) = \sum_{i=1}^{n} y_i K_r(\frac{x_i - x}{h_n}) \frac{w_{r,0}(x_i - x) - w_{r,1}}{w_{r,0}w_{r,2} - w_{r,1}^2}$$

where $w_{l,j} = \sum_{i=1}^{n} K_l(\frac{x_i - x}{h_n})(x_i - x)^j$ and $w_{r,j} = \sum_{i=1}^{n} K_r(\frac{x_i - x}{h_n})(x_i - x)^j$ for $j = 0, 1, 2$.

Figure 2.1 presents $\widehat{a}_{l,0}(\cdot), \widehat{a}_{r,0}(\cdot)$ and $\widehat{f}(\cdot)$ by the dotted, dashed and solid curves in the case of Figure 1.1 except that the noise in the data has been ignored by setting $\sigma = 0$. It can be seen that blurring occurs in $[x_0, x_0 + h_n/2]$ if $\widehat{a}_{l,0}(\cdot)$ is used to fit $f(\cdot)$ and point x_0 is a jump point. Similarly, blurring occurs in $[x_0 - h_n/2, x_0)$ if $\widehat{a}_{r,0}(\cdot)$ is used to fit $f(\cdot)$ and point x_0 is a jump point. Our estimator $\widehat{f}(\cdot)$, however, can preserve the jumps well because $\widehat{f}(\cdot)$ is defined as $\widehat{a}_{l,0}(\cdot)$ in $[x_0 - h_n/2, x_0)$ and as $\widehat{a}_{r,0}(\cdot)$ in $[x_0, x_0 + h_n/2]$ when x_0 is a jump point.

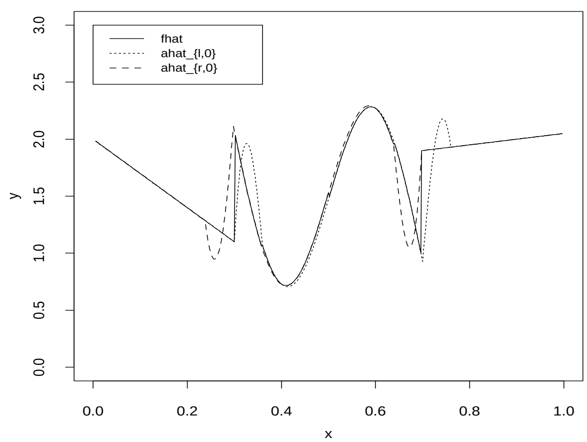

Figure 2.1: The dotted, dashed and solid curves denote $\widehat{a}_{l,0}(\cdot)$, $\widehat{a}_{r,0}(\cdot)$ and $\widehat{f}(\cdot)$ in the case of Figure 1.1 except that the noise in data has been ignored by setting $\sigma = 0$.

When x is in boundary regions $[0, h_n/2)$ and $(1 - h_n/2, 1]$, estimator of $f(x)$ is not defined by (1.4). In such case there are several possible approaches to estimate $f(x)$ if no jumps exist in $[0, h_n)$ and $(1-h_n, 1]$. For example, $\widehat{f}(x)$ could be defined by the conventional local linear kernel estimator constructed from observations in $[0, x + h_n/2]$ or $[x - h_n/2, 1]$ depending on whether $x \in [0, h_n/2)$ or $x \in (1 - h_n/2, 1]$. In the following sections, we define $\widehat{f}(x) = \widehat{a}_{r,0}(x)$ when $x \in [0, h_n/2)$ and $\widehat{f}(x) = \widehat{a}_{l,0}(x)$ when $x \in (1-h_n/2, 1]$ for simplicity. If there are jump points in $[0, h_n)$ (or $(1-h_n, 1]$), however, estimation of $f(x)$ in boundary region $[0, h_n/2)$ (or $(1 - h_n/2, 1]$) is still an open problem.

In the literature, there are several existing data-driven bandwidth selection procedures such as the plug-in procedures, the cross-validation procedure, the Mellow's C_p criterion and the Akaike's information criterion (cf. e.g., Chu and Marron 1991; Loader 1999). Since the exact expressions for the mean and variance of the jump-preserving estimator $\widehat{f}(\cdot)$ in (1.4) are not available at this moment, the plug-in procedures are not considered here. In the numerical examples presented in Sections 4 and 5, we determine the bandwidth h_n by the cross-validation procedure. That is, the optimal h_n is chosen by minimizing the following cross-validation criterion:

$$CV(h_n) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{f}_{-i}(x_i)\right)^2, \tag{2.3}$$

where $\widehat{f}_{-i}(x)$ is the "leave-1-out" estimator of $f(x)$ with bandwidth h_n. Namely, the observation (x_i, y_i) is left out in constructing $\widehat{f}_{-i}(x)$, for $i = 1, 2, \cdots, n$. A numerical example in Section 4 shows that the chosen bandwidth based upon (2.3) performs well.

3 Strong Consistency

The conventional local smoothing estimators of $f(\cdot)$ such as the one from (1.2) are not statistically consistent at jump positions. In this section we establish the almost sure consistency of the jump-preserving estimator $\widehat{f}(\cdot)$ which says that $\widehat{f}(\cdot)$ converges almost surely to the true regression function in the entire design space $[0,1]$ under some regularity conditions. That is, $\widehat{f}(\cdot)$ is jump-preserving. First we have the following result for $\widehat{a}_{l,0}(\cdot)$ and $\widehat{a}_{r,0}(\cdot)$.

Theorem 3.1 Suppose that $f(\cdot)$ has a continuous second-order derivative in $[0,1]$; $\max_{1\le i\le n+1}(x_i - x_{i-1}) = O(1/n)$ where $x_0 = 0$ and $x_{n+1} = 1$; the kernel function $K(\cdot)$ is Lipschitz (1) continuous; the bandwidth $h_n = O(n^{-1/5})$. Then

$$\frac{n^{2/5}}{\log n \log\log n}\|\widehat{a}_{l,0} - f\|_{[h_n/2,1]} = o(1), \quad a.s. \tag{3.1}$$

$$\frac{n^{2/5}}{\log n \log\log n}\|\widehat{a}_{r,0} - f\|_{[0,1-h_n/2]} = o(1), \quad a.s. \tag{3.2}$$

where $\|g\|_{[a,b]}$ denotes $\max_{a\le x\le b}|g(x)|$.

Theorem 3.1 establishes the almost sure uniform consistency of $\widehat{a}_{l,0}(\cdot)$ and $\widehat{a}_{r,0}(\cdot)$ when $f(\cdot)$ is continuous in the design space $[0,1]$. Its proof is given in Appendix A. When $f(\cdot)$ has jumps in $[0,1]$ as specified by model (1.1), Theorem 3.1 also gives almost sure consistency of $\widehat{f}(\cdot)$ in continuous regions $D_1 := [0,1]\backslash\bigcup_{j=1}^{m}(s_j - h_n/2, s_j + h_n/2)$ since $\|\widehat{f} - f\|_{D_1} \le \max(\|\widehat{a}_{l,0} - f\|_{D_1}, \|\widehat{a}_{r,0} - f\|_{D_1})$ by (1.4). In the neighborhood of jump points $D_2 := \bigcup_{j=1}^{m}(s_j - h_n/2, s_j + h_n/2)$, we have the following result.

Theorem 3.2 Suppose that x is a given point in $(0,1)$; $\max_{1\le i\le n+1}(x_i - x_{i-1}) = O(1/n)$ where $x_0 = 0$ and $x_{n+1} = 1$; the kernel function $K(\cdot)$ is Lipschitz (1) continuous; $\lim_{n\to\infty} h_n = 0$ and

$\lim_{n\to\infty} nh_n = \infty$. If $f(\cdot)$ has a continuous first-order derivative in $[x, x+h_n/2]$, then

$$RSS_r(x) = v_{r,0}\sigma^2 nh_n + o(nh_n), \quad a.s. \tag{3.3}$$

If $f(\cdot)$ has a jump in $[x, x+h_n/2]$ at $x_\tau := x + \tau h_n$ with magnitude d_τ where $0 \le \tau \le 1/2$ and $f(\cdot)$ has a continuous first-order derivative in $[x, x+h_n/2]$ except at x_τ at which $f(\cdot)$ has a right (when $\tau = 0$) or left (when $\tau = 1/2$) or both (when $0 < \tau < 1/2$) first-order derivatives $f'_+(x_\tau)$ and $f'_-(x_\tau)$, then

$$RSS_r(x) = (v_{r,0}\sigma^2 + d_\tau^2 C_\tau^2) nh_n + o(nh_n), \quad a.s., \tag{3.4}$$

where

$$\begin{aligned} C_\tau^2 \;=\; & \frac{1}{(v_{r,0}v_{r,2} - v_{r,1}^2)^2} \int_0^\tau \left[\int_\tau^{1/2} (v_{r,2} - v_{r,1}x) K_r(x)dx + u \int_\tau^{1/2} (v_{r,0}x - v_{r,1}) K_r(x)dx \right]^2 K_r(u)du + \\ & \frac{1}{(v_{r,0}v_{r,2} - v_{r,1}^2)^2} \int_\tau^{1/2} \left[\int_0^\tau (v_{r,2} - v_{r,1}x) K_r(x)dx - u \int_\tau^{1/2} (v_{r,0}x - v_{r,1}) K_r(x)dx \right]^2 K_r(u)du \end{aligned}$$

and $v_{r,j} = \int_0^{1/2} x^j K_r(x)dx$ for $j = 0, 1, 2$.

Similar results could be derived for $RSS_l(x)$. It can be checked that C_τ^2 is positive when $\tau \in (0, 1/2)$ and 0 when $\tau = 0$ or $1/2$. If the kernel function $K(\cdot)$ is chosen to be the Epanechnikov function defined by $K(x) = 1.5(1 - 4x^2)$ when $x \in [-1/2, 1/2]$ and 0 otherwise (cf. Section 3.2.6, Fan and Gijbels 1996), then C_τ^2 as a function of τ is displayed in Figure 3.1.

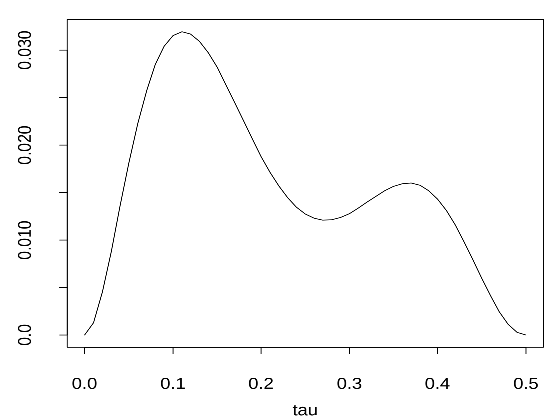

Figure 3.1: C_τ^2 as a function of τ when $K(\cdot)$ is chosen to be the Epanechnikov function.

By (3.3) and (3.4), if there is a jump in $[x - h_n/2, x + h_n/2]$, a neighborhood of a given point x, and this jump point is located on the right side of x, then $RSS_l(x) < RSS_r(x)$, a.s., when n is large enough. Consequently, $\widehat{f}(x) = \widehat{a}_{l,0}(x)$, a.s., when n is large enough. On the other hand, if the jump point is located on the left side of x, then $RSS_l(x) > RSS_r(x)$, a.s., and $\widehat{f}(x) = \widehat{a}_{r,0}(x)$, a.s., when n is large enough. By combining this fact and (3.1)-(3.2) in Theorem 3.1, we have the following results.

Theorem 3.3 Suppose that $f(\cdot)$ has a continuous second-order derivative in $[0, 1]$ except at the jump positions $\{s_j, j = 1, 2, \cdots, m\}$ where $f(\cdot)$ has left and right second-order derivatives; $\max_{1\leq i\leq n+1}(x_i - x_{i-1}) = O(1/n)$ where $x_0 = 0$ and $x_{n+1} = 1$; the kernel function $K(\cdot)$ is Lipschitz (1) continuous; and the bandwidth $h_n = O(n^{-1/5})$. Then

(i)

$$\frac{n^{2/5}}{\log n \log\log n}\|\widehat{f} - f\|_{D_1} = o(1), \quad a.s.;$$

(ii) for each $x \in D_2$,

$$\frac{n^{2/5}}{\log n \log\log n}(\widehat{f}(x) - f(x)) = o(1), \quad a.s.;$$

(iii) for any small number $0 < \delta < 1/4$,

$$\frac{n^{2/5}}{\log n \log\log n}\|\widehat{f} - f\|_{D_{2,\delta}} = o(1), \quad a.s.,$$

where $D_{2,\delta} = \bigcup_{j=1}^{m}\{[s_j - (1/2 - \delta)h_n, s_j - \delta h_n] \bigcup [s_j + \delta h_n, s_j + (1/2 - \delta)h_n]\}$.

Theorem 3.3 says that $\widehat{f}(\cdot)$ is uniformly consistent in continuous regions D_1 with rate $o(n^{-2/5} \log n \log\log n)$. In the neighborhood of jump points, it is consistent pointwise with the same rate. Because C_τ^2 has a positive lower bound when $\tau \in [\delta, 1/2 - \delta]$ for any given number $0 < \delta < 1/4$, $\widehat{f}(\cdot)$ is also uniformly consistent with rate $o(n^{-2/5} \log n \log\log n)$ in $D_{2,\delta}$ which equals to $D_2 \backslash D_\delta$ where $D_\delta = \bigcup_{j=1}^{m}[(s_j - h_n/2, s_j - (1/2 - \delta)h_n) \bigcup (s_j - \delta h_n, s_j + \delta h_n) \bigcup (s_j + (1/2 - \delta)h_n, s_j + h_n/2)]$.

4 Simulation Study

We present some simulation results regarding bandwidth selection and the numerical performance of the jump-preserving curve fitting procedure (1.4) in this section. Let us revisit the example of Figure 1.1 first. The true regression function in this example is $f(x) = -3x + 2$ when $x \in [0, 0.3)$; $f(x) = -3x + 3 - \sin((x - 0.3)\pi/0.2)$ when $x \in [0.3, 0.7)$; and $f(x) = 0.5x + 1.55$ when $x \in [0.7, 1]$. It has two jump points at $x = 0.3$ and $x = 0.7$. Both jump magnitudes are equal to 1. Observations are generated from model (1.1) with $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, \cdots, n$. The bandwidth used in procedure (1.4) is assumed to have the form $h_n = k/n$, where k is an odd integer, for convenience. Without confusion, k is sometimes called the bandwidth in this section.

Figure 4.1 presents the MSE values of the fitted curve by the jump-preserving procedure (1.4) with several k values when $n = 200$ and $\sigma = 0.2$. To remove some randomness in the results, all MSE values presented in this section are actually averages of 1000 replications. It can be seen from the plot that the MSE value first decreases and then increases when k increases. The bandwidth k works as a tuning parameter to balance "underfit" and "overfit" as in the conventional local smoothing procedures. The best bandwidth in this case is $k = 29$ which makes the MSE reach the minimum. The dotted curve in Figure 1.1 shows one realization of the fitted curve with the best bandwidth $k = 29$. The dashed curve shows the conventional local linear kernel estimator with the same bandwidth.

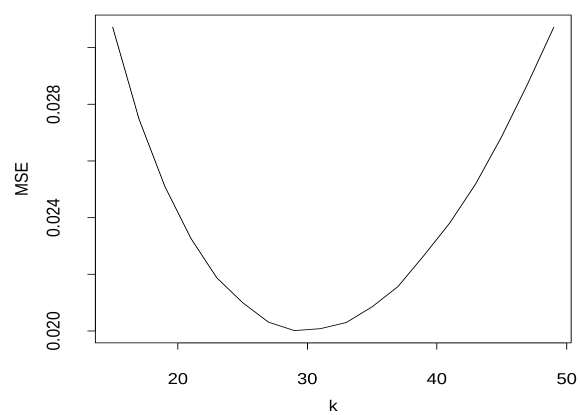

Figure 4.1: MSE values of the fitted curve by the jump-preserving procedure (1.4) with several k values when $n = 200$ and $\sigma = 0.2$.

We then perform simulations with several different n and σ values. The optimal bandwidths and the corresponding MSE values are presented in Figures 4.2(a) and 4.2(c), respectively. From the plots, it can be seen that (1) the optimal k increases when sample size n increases or σ increases and (2) the corresponding MSE value decreases when n increases or σ decreases. The first finding suggests that the bandwidth should be chosen larger when the sample size is larger or the data is noisier, which is intuitively reasonable. The second finding might reflect the consistency of the fitted curve.

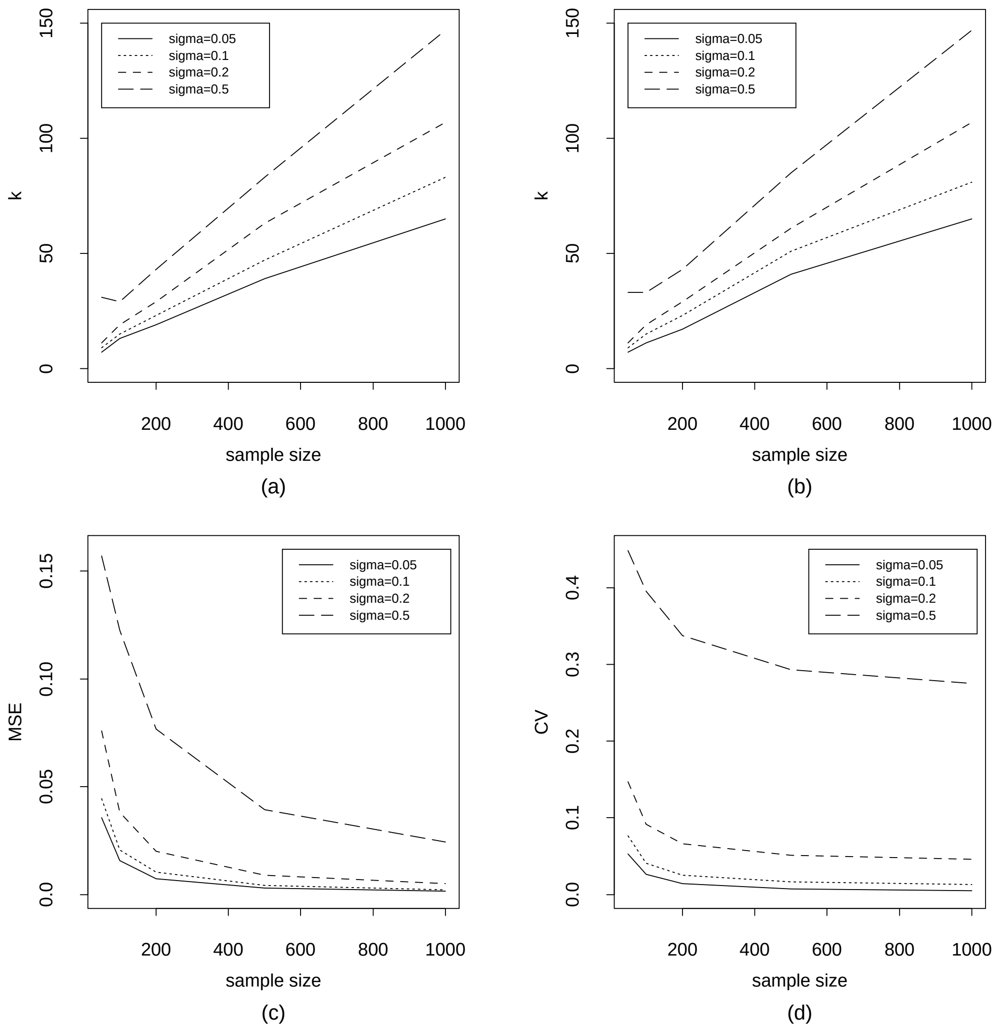

Figure 4.2: (a) The optimal bandwidths by the MSE criterion; (b) the optimal bandwidths by the CV criterion; (c) the corresponding MSE values when the bandwidths in plot (a) are used; (d) the corresponding CV values when the bandwidths in plot (b) are used.

As a comparison, the optimal bandwidths by the cross-validation procedure are presented in Figure 4.2(b). The corresponding CV values (defined by equation (2.3)) are shown in Figure 4.2(d). By comparing Figures 4.2(a) and 4.2(b), it can be seen that bandwidths selected by the cross-validation procedure are close to the optimal bandwidths based on the MSE criterion.

From Figure 1.1, it can be seen that bluring occurs around the jump points if $f(\cdot)$ is estimated by the conventional local linear kernel estimator. The jump-preserving estimator (1.4) preserves the jumps quite well, which is further confirmed by Figure 4.3. In Figure 4.3(a), the solid curve denotes the true regression model, the dotted curve denotes the averaged estimator by the jump-preserving procedure which is calculated from 1000 replications. The lower and upper dashed curves represent the 2.5 and 97.5 percentiles of these 1000 replications. We can see that the two sharp jumps are preserved well by the procedure (1.4). As a comparison, the averaged estimator and the corresponding percentiles by the conventional local linear kernel smoothing procedure with the same bandwidth are presented in Figure 4.3(b). It can be seen that the two jumps are blurred.

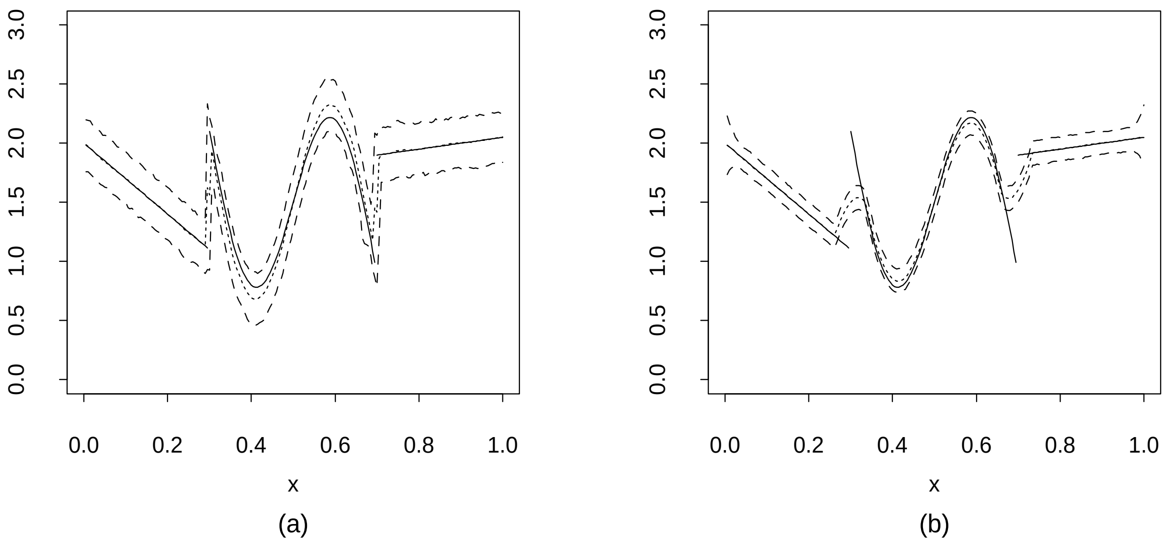

Figure 4.3: The solid curve denotes the true regression model, the dotted curve denotes the averaged estimator which is calculated from 1000 replications. The lower and upper dashed curves represent the 2.5 and 97.5 percentiles of these 1000 replications. (a) Results from the jump-preserving procedure (1.4); (b) results from the conventional local linear kernel smoothing procedure.

5 An Application

In this section, we apply the jump-preserving curve fitting procedure (1.4) to a sea-level pressure dataset. In Figure 5.1, small dots denote the December sea-level pressures during 1921-1992 observed by the Bombay weather station in India. Meteorologists (cf. Shea *et al.* 1994) noticed a jump around year 1960 in this dataset and the existence of this jump was confirmed by Qiu and Yandell (1998) with their local polynomial jump detection algorithm.

In Figure 5.1, the solid curve denotes the fitted regression curve by our jump-preserving curve fitting procedure (1.4). In the procedure, the bandwidth is chosen to be $k = 25$ which is determined by the cross-validation procedure (2.3). As indicated by the plot, the jump around year 1960 is preserved well by our procedure.

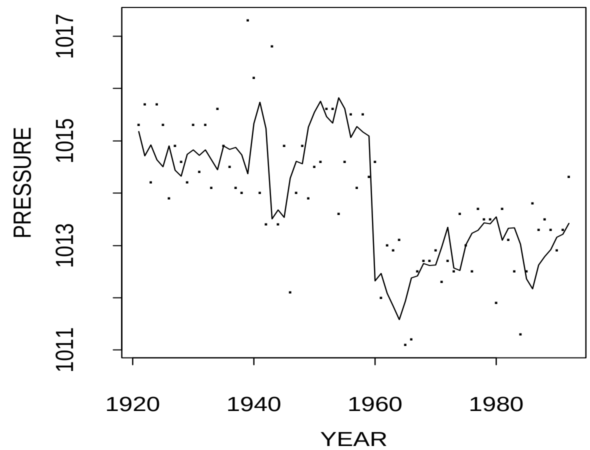

Figure 5.1: Small dots denote the December sea-level pressures during 1921-1992 observed by the Bombay weather station in India. The solid curve is the jump-preserving estimator by the procedure (1.4).

6 Concluding Remarks

We have presented a jump-preserving curve fitting procedure which automatically accommodates possible jumps of the regression curve without knowing the number of jumps. The fitted curve is proved to be statistically consistent in the entire design space. Numerical examples show that it works reasonably well in applications. The following issues related to this topic need further investigation. First, the procedure (1.4) works well in boundary regions $[0, h_n/2)$ and $(1-h_n/2, 1]$ only under the condition that there are no jumps in $[0, h_n)$ and $(1-h_n, 1]$. This condition can always be satisfied when the sample size is large. When the sample size is small, however, this condition may not be true in some cases and it is still an open problem to fit $f(\cdot)$ when jumps exist in the boundary regions. Second, the plog-in procedures to choose bandwidth of a local smoother are often based on the bias-variance trade-off of the fitted regression model. Exact expressions for the mean and variance of the jump-preserving procedure (1.4) are not available yet, which needs further research.

Acknowledgement: The author would like to thank Mr. Alexandre Lambert of the Institut de Statistique at Universite catholique de Louvain in Belgium for pointing out a mistake in the expression of C_τ^2 appeared in (3.4).

Appendix

A Proof of Theorem 3.1

We only prove equation (3.1) here. Equation (3.2) can be proved similarly. First of all,

$$E(\widehat{a}_{l,0}(x)) = \sum_{i=1}^{n} f(x_i) K_l(\frac{x_i - x}{h_n}) \frac{w_{l,2} - w_{l,1}(x_i - x)}{w_{l,0} w_{l,2} - w_{l,1}^2}. \tag{A.1}$$

We notice that the summation on the right hand side of (A.1) is only for those x_i in $[x - h_n/2, x)$. By Taylor's expansion,

$$f(x_i) = f(x) + f'(x)(x_i - x) + \frac{1}{2} f''(x)(x_i - x)^2 + o(h_n^2), \tag{A.2}$$

where $x_i \in [x - h_n/2, x)$. By combining (A.1) and (A.2), we have

$$E(\widehat{a}_{l,0}(x)) = f(x) + f''(x) \frac{w_{l,2}^2 - w_{l,1} w_{l,3}}{2(w_{l,0} w_{l,2} - w_{l,1}^2)} + o(h_n^2), \tag{A.3}$$

where $w_{l,3} = \sum_{i=1}^{n} K_l(\frac{x_i - x}{h_n})(x_i - x)^3$. Furthermore it can be checked that

$$\frac{w_{l,0}}{n h_n} = v_{l,0} + o(1),\ \frac{w_{l,1}}{n h_n^2} = v_{l,1} + o(1),\ \frac{w_{l,2}}{n h_n^3} = v_{l,2} + o(1),\ \frac{w_{l,3}}{n h_n^4} = v_{l,3} + o(1), \tag{A.4}$$

where $v_{l,j} = \int_{-1/2}^{0} x^j K_l(x) dx$ for $j = 0, 1, 2, 3$. By combining (A.3) and (A.4), we have

$$E(\widehat{a}_{l,0}(x)) = f(x) + f''(x) \frac{v_{l,2}^2 - v_{l,1} v_{l,3}}{2(v_{l,0} v_{l,2} - v_{l,1}^2)} h_n^2 + o(h_n^2).$$

Therefore

$$E(\widehat{a}_{l,0}(x)) - f(x) = f''(x) \frac{v_{l,2}^2 - v_{l,1} v_{l,3}}{2(v_{l,0} v_{l,2} - v_{l,1}^2)} h_n^2 + o(h_n^2). \tag{A.5}$$

Now let

$$\begin{aligned}
\tilde{\epsilon}_i &= \epsilon_i I(i^{1/2} - |\epsilon_i|),\ i = 1, 2, \cdots, n \\
g_n(x) &= \sum_{i=1}^{n} K_l(\frac{x_i - x}{h_n}) \frac{w_{l,2} - w_{l,1}(x_i - x)}{w_{l,0} w_{l,2} - w_{l,1}^2} \epsilon_i \\
\tilde{g}_n(x) &= \sum_{i=1}^{n} K_l(\frac{x_i - x}{h_n}) \frac{w_{l,2} - w_{l,1}(x_i - x)}{w_{l,0} w_{l,2} - w_{l,1}^2} \tilde{\epsilon}_i \\
&=: \sum_{i=1}^{n} \tilde{g}_n(i)
\end{aligned}$$

For any $\varepsilon > 0$,

$$
\begin{aligned}
& P(\frac{n^{2/5}}{\log n \log\log n}[\tilde{g}_n(x) - E(\tilde{g}_n(x))] > \varepsilon) \\
\leq \; & \exp(\log n^{-\varepsilon(\log\log n)^{1/2}}) E(\Pi_{i=1}^n \exp(\frac{n^{2/5}}{(\log\log n)^{1/2}}[\tilde{g}_n(i) - E(\tilde{g}_n(i))])) \\
\leq \; & n^{-\varepsilon(\log\log n)^{1/2}} \exp(\frac{n^{4/5}}{\log\log n}\sum_{i=1}^n Var(\tilde{g}_n(i)))
\end{aligned}
$$

by an application of the Chebyshev's inequality of the exponential form. Now

$$
\begin{aligned}
\sum_{i=1}^n Var(\tilde{g}_n(i)) &\leq \sigma^2 \sum_{i=1}^n K_l^2(\frac{x_i - x}{h_n})[\frac{w_{l,2} - w_{l,1}(x_i - x)}{w_{l,0}w_{l,2} - w_{l,1}^2}]^2 \\
&= \frac{\sigma^2}{nh_n}\int_{-1/2}^0 K_l^2(x)(\frac{v_{l,2}^2 - v_{l,1}x}{v_{l,0}v_{l,2} - v_{l,1}^2})^2 dx \\
&= \frac{\sigma^2}{nh_n} C_{l,1}(K),
\end{aligned}
$$

where $C_{l,1}(K)$ is a constant. So

$$
P(\frac{n^{2/5}}{\log n \log\log n}[\tilde{g}_n(x) - E(\tilde{g}_n(x))] > \varepsilon) = O(n^{-\varepsilon(\log\log n)^{1/2}}) \tag{A.6}
$$

for all $x \in [h_n/2, 1]$.

We now define $D_n = \{x : |x| \leq n^{1/\delta} + 1, x \in R\}$, for some $\delta > 0$. Let E_n be a set such that, for any $x \in D_n$, there exists some $Z(x) \in E_n$ such that $|x - Z(x)| < n^{-2}$, and E_n has at most $N_n = [2n^2(n^{1/\delta} + 1)] + 1$ elements, where $[x]$ denotes the integral part of x. Then

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

where

$$
\begin{aligned}
S_{1n} &= \frac{n^{2/5}}{\log n \log\log n} \sup_{x\in[h_n/2,1]\bigcap D_n} |\tilde{g}_n(x) - \tilde{g}_n(Z(x))| \\
S_{2n} &= \frac{n^{2/5}}{\log n \log\log n} \sup_{x\in[h_n/2,1]\bigcap D_n} |\tilde{g}_n(Z(x)) - E(\tilde{g}_n(Z(x)))| \\
S_{3n} &= \frac{n^{2/5}}{\log n \log\log n} \sup_{x\in[h_n/2,1]\bigcap D_n} |E(\tilde{g}_n(Z(x))) - E(\tilde{g}_n(x))|
\end{aligned}
$$

From (A.6), $P(S_{2n} > \varepsilon) = O(N_n n^{-\varepsilon(\log\log n)^{1/2}})$. By the Borel-Cantelli Lemma,

$$
\lim_{n\to\infty} S_{2n} = 0, \; a.s. \tag{A.7}
$$

Now

$$\begin{aligned}
S_{1n} &= \frac{n^{2/5}}{\log n \log\log n} \sup_{x \in [h_n/2,1] \bigcap D_n} |\sum_{i=1}^{n} [K_l(\frac{x_i - x}{h_n}) \frac{w_{l,2} - w_{l,1}(x_i - x)}{w_{l,0}w_{l,2} - w_{l,1}^2} - \\
& \quad K_l(\frac{x_i - Z(x)}{h_n}) \frac{w_{l,2} - w_{l,1}(x_i - Z(x))}{w_{l,0}w_{l,2} - w_{l,1}^2}]\tilde{\epsilon}_i| \\
&\leq \frac{n^{2/5} n^{1/2}}{\log n \log\log n} \sup_{x \in [h_n/2,1] \bigcap D_n} |\frac{1}{nh_n} \sum_{i=1}^{n} [K_l(\frac{x_i - x}{h_n}) \frac{v_{l,2} - v_{l,1}(x_i - x)/h_n}{v_{l,0}v_{l,2} - v_{l,1}^2} - \\
& \quad K_l(\frac{x_i - Z(x)}{h_n}) \frac{v_{l,2} - v_{l,1}(x_i - Z(x))/h_n}{v_{l,0}v_{l,2} - v_{l,1}^2}]| \\
&\leq \frac{n^{2/5+1/2}}{\log n \log\log n} \frac{C_{l,2}(K)}{n^2 h_n},
\end{aligned}$$

where $C_{l,2}(K)$ is a constant. In the last inequality above, we have used the Lipschitz (1) property of $K_l(\cdot)$. Therefore

$$\lim_{n \to \infty} S_{1n} = 0, \ a.s. \tag{A.8}$$

Similarly,

$$\lim_{n \to \infty} S_{3n} = 0. \tag{A.9}$$

By combining (A.7)-(A.9), we have

$$\frac{n^{2/5}}{\log n \log\log n} \|\tilde{g}_n - E(\tilde{g}_n)\|_{[h_n/2,1] \bigcap D_n} = o(1), \ a.s. \tag{A.10}$$

Now,

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

Since $E(\epsilon_1^2) < \infty$, there exists a full set Ω_0 such that for each $\omega \in \Omega_0$ there exists a finite positive integer N_ω and for $n \geq N_\omega$,

$$\epsilon_n(\omega) = \tilde{\epsilon}_n(\omega).$$

So for all $n \geq N_\omega$,

$$\begin{aligned}
|g_n(x) - \tilde{g}_n(x)| &\leq \frac{1}{nh_n} \sum_{i=1}^{N_\omega} K_l(\frac{x_i - x}{h_n})|\frac{v_{l,2} - v_{l,1}(x_i - x)/h_n}{v_{l,0}v_{l,2} - v_{l,1}^2}||\epsilon_i - \tilde{\epsilon}_i| \\
&\leq \frac{C(N_\omega)}{nh_n},
\end{aligned}$$

where $C(N_\omega)$ is a constant. Therefore,

$$\frac{n^{2/5}}{\log n \log\log n} \|g_n - \tilde{g}_n\|_{[h_n/2,1]} = o(1), \ a.s. \tag{A.11}$$

Similarly,

$$\frac{n^{2/5}}{\log n \log\log n}\|E(\tilde{g}_n) - E(g_n)\|_{[h_n/2,1]} = o(1). \tag{A.12}$$

By (A.10)-(A.12), we have

$$\frac{n^{2/5}}{\log n \log\log n}\|g_n - E(g_n)\|_{[h_n/2,1]} = o(1),\ a.s. \tag{A.13}$$

By (A.5) and (A.13), we get equation (3.1).

B Proof of Theorem 3.2

By the definition of $RSS_r(x)$,

$$\begin{aligned}
RSS_r(x) &= \sum_{i=1}^n [y_i - \widehat{a}_{r,0}(x) - \widehat{a}_{r,1}(x)(x_i - x)]^2 K_r(\frac{x_i - x}{h_n}) \\
&= \sum_{i=1}^n [\epsilon_i + f(x_i) - \widehat{a}_{r,0}(x) - \widehat{a}_{r,1}(x)(x_i - x)]^2 K_r(\frac{x_i - x}{h_n}) \\
&= \sum_{i=1}^n \epsilon_i^2 K_r(\frac{x_i - x}{h_n}) + 2\sum_{i=1}^n \epsilon_i [f(x_i) - \widehat{a}_{r,0}(x) - \widehat{a}_{r,1}(x)(x_i - x)] K_r(\frac{x_i - x}{h_n}) + \\
&\quad \sum_{i=1}^n [f(x_i) - \widehat{a}_{r,0}(x) - \widehat{a}_{r,1}(x)(x_i - x)]^2 K_r(\frac{x_i - x}{h_n}) \\
&=: I_1 + I_2 + I_3
\end{aligned}$$

Let us first prove equation (3.3) under the condition that $f(\cdot)$ has continuous first-order derivative in $[x, x + h_n/2]$. By similar auguments to those in Appendix A,

$$I_1 = v_{r,0}\sigma^2 n h_n + o(nh_n),\ a.s. \tag{B.1}$$

Now

$$\begin{aligned}
I_2 &= 2\sum_{i=1}^n \epsilon_i [f(x) + f'(x)(x_i - x) - \widehat{a}_{r,0}(x) - \widehat{a}_{r,1}(x)(x_i - x) + o(h_n)] K_r(\frac{x_i - x}{h_n}) \\
&= 2(f(x) - \widehat{a}_{r,0}(x)) \sum_{i=1}^n \epsilon_i K_r(\frac{x_i - x}{h_n}) + 2(f'(x) - \widehat{a}_{r,1}(x)) \sum_{i=1}^n \epsilon_i K_r(\frac{x_i - x}{h_n})(x_i - x) + o(nh_n) \\
&= o(nh_n) + o(h_n^{-1}) \times O(nh_n) \times O(h_n) + o(nh_n) \\
&= o(nh_n)
\end{aligned} \tag{B.2}$$

In the third equation above, we have used the results that $f(x) - \widehat{a}_{r,0}(x) = o(1),\ a.s.$, and $f'(x) - \widehat{a}_{r,1}(x) = o(1/h_n),\ a.s.$, where the first result is from Theorem 3.1 and the second result can be derived by similar arguments to those in Appendix A. It can be similarly checked that

$$I_3 = o(nh_n),\ a.s. \tag{B.3}$$

By combining (B.1)-(B.3), we get equation (3.3).

Next we prove equation (3.4) under the condition that $f(\cdot)$ has a jump in $[x, x + h_n/2]$ at $x_\tau = x + \tau h_n$ where $0 \le \tau \le 1/2$ is a constant. First,

$$\begin{aligned}
\widehat{a}_{r,0}(x) &= \sum_{i=1}^{n} y_i K_r(\frac{x_i - x}{h_n}) \frac{w_{r,2} - w_{r,1}(x_i - x)}{w_{r,0}w_{r,2} - w_{r,1}^2} \\
&= \sum_{x_i < x_\tau} f(x_i) K_r(\frac{x_i - x}{h_n}) \frac{w_{r,2} - w_{r,1}(x_i - x)}{w_{r,0}w_{r,2} - w_{r,1}^2} + \sum_{x_i \ge x_\tau} f(x_i) K_r(\frac{x_i - x}{h_n}) \frac{w_{r,2} - w_{r,1}(x_i - x)}{w_{r,0}w_{r,2} - w_{r,1}^2} \\
&\quad + \sum_{i=1}^{n} \epsilon_i K_r(\frac{x_i - x}{h_n}) \frac{w_{r,2} - w_{r,1}(x_i - x)}{w_{r,0}w_{r,2} - w_{r,1}^2} \\
&= \sum_{x_i < x_\tau} (f_-(x_\tau) + o(1)) K_r(\frac{x_i - x}{h_n}) \frac{w_{r,2} - w_{r,1}(x_i - x)}{w_{r,0}w_{r,2} - w_{r,1}^2} \\
&\quad + \sum_{x_i \ge x_\tau} (f_-(x_\tau) + d_\tau + o(1)) K_r(\frac{x_i - x}{h_n}) \frac{w_{r,2} - w_{r,1}(x_i - x)}{w_{r,0}w_{r,2} - w_{r,1}^2} + o(1),\ a.s. \\
&= f_-(x_\tau) + \frac{d_\tau \int_\tau^{1/2} K_r(x)(v_{r,2} - v_{r,1}x)dx}{v_{r,0}v_{r,2} - v_{r,1}^2} + o(1),\ a.s.
\end{aligned} \tag{B.4}$$

In the last equation above, we have used (A.4). Similarly we can check that

$$\widehat{a}_{r,1}(x) = \frac{d_\tau \int_\tau^{1/2} K_r(x)(v_{r,0}x - v_{r,1})dx}{h_n(v_{r,0}v_{r,2} - v_{r,1}^2)} + o(1/h_n),\ a.s. \tag{B.5}$$

Then

$$\begin{aligned}
I_2 &= 2\sum_{x_i < x_\tau} \epsilon_i [f(x_i) - f_-(x_\tau) - \frac{d_\tau \int_\tau^{1/2} K_r(x)(v_{r,2} - v_{r,1}x)dx}{v_{r,0}v_{r,2} - v_{r,1}^2}] K_r(\frac{x_i - x}{h_n}) + \\
&\quad 2\sum_{x_i \ge x_\tau} \epsilon_i [f(x_i) - f_-(x_\tau) - \frac{d_\tau \int_\tau^{1/2} K_r(x)(v_{r,2} - v_{r,1}x)dx}{v_{r,0}v_{r,2} - v_{r,1}^2}] K_r(\frac{x_i - x}{h_n}) - \\
&\quad 2\sum_{i=1}^{n} \epsilon_i (x_i - x) K_r(\frac{x_i - x}{h_n}) \times \frac{d_\tau \int_\tau^{1/2} K_r(x)(v_{r,0}x - v_{r,1})dx}{h_n(v_{r,0}v_{r,2} - v_{r,1}^2)} + o(nh_n) \\
&= -2\frac{d_\tau \int_\tau^{1/2} K_r(x)(v_{r,2} - v_{r,1}x)dx}{v_{r,0}v_{r,2} - v_{r,1}^2} \sum_{x_i < x_\tau} \epsilon_i K_r(\frac{x_i - x}{h_n}) + \\
&\quad 2d_\tau (1 - \frac{\int_\tau^{1/2} K_r(x)(v_{r,2} - v_{r,1}x)dx}{v_{r,0}v_{r,2} - v_{r,1}^2}) \sum_{x_i \ge x_\tau} \epsilon_i K_r(\frac{x_i - x}{h_n}) + o(nh_n),\ a.s. \\
&= o(nh_n),\ a.s.
\end{aligned} \tag{B.6}$$

and

$$
\begin{aligned}
I_3 &= \sum_{x_i<x_\tau}[f(x_i)-f_-(x_\tau)-\frac{d_\tau\int_\tau^{1/2}K_r(x)(v_{r,2}-v_{r,1}x)dx}{v_{r,0}v_{r,2}-v_{r,1}^2}- \\
&\quad \frac{d_\tau\int_\tau^{1/2}K_r(x)(v_{r,0}x-v_{r,1})dx}{h_n(v_{r,0}v_{r,2}-v_{r,1}^2)}(x_i-x)]^2K_r(\frac{x_i-x}{h_n})+ \\
&\quad \sum_{x_i\geq x_\tau}[f(x_i)-f_-(x_\tau)-\frac{d_\tau\int_\tau^{1/2}K_r(x)(v_{r,2}-v_{r,1}x)dx}{v_{r,0}v_{r,2}-v_{r,1}^2}- \\
&\quad \frac{d_\tau\int_\tau^{1/2}K_r(x)(v_{r,0}x-v_{r,1})dx}{h_n(v_{r,0}v_{r,2}-v_{r,1}^2)}(x_i-x)]^2K_r(\frac{x_i-x}{h_n})+o(nh_n),\ a.s. \\
&= nh_n\int_0^\tau[\frac{d_\tau\int_\tau^{1/2}K_r(x)(v_{r,2}-v_{r,1}x)dx}{v_{r,0}v_{r,2}-v_{r,1}^2}+\frac{d_\tau\int_\tau^{1/2}K_r(x)(v_{r,0}x-v_{r,1})dx}{v_{r,0}v_{r,2}-v_{r,1}^2}u]^2K_r(u)du+ \\
&\quad nh_n\int_\tau^{1/2}[\frac{d_\tau\int_0^\tau K_r(x)(v_{r,2}-v_{r,1}x)dx}{v_{r,0}v_{r,2}-v_{r,1}^2}-\frac{d_\tau\int_\tau^{1/2}K_r(x)(v_{r,0}x-v_{r,1})dx}{v_{r,0}v_{r,2}-v_{r,1}^2}u]^2K_r(u)du+ \\
&\quad o(nh_n),\ a.s. \\
&= d_\tau^2C_\tau^2nh_n+o(nh_n),\ a.s. \qquad \text{(B.7)}
\end{aligned}
$$

By combining (B.1), (B.6) and (B.7), we get equation (3.4).

References

Chu, C.K., and Marron, J.S. (1991), "Choosing a kernel regression estimator," *Statistical Science 6,* 404-436.

Cobb, G.W.(1978), "The problem of the Nile: conditional solution to a changepoint problem," *Biometrika 65*, 243-251.

Eubank, R.L., and Speckman, P.L.(1994),"Nonparametric estimation of functions with jump discontinuities," IMS Lecture Notes, vol. 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,* Chapman & Hall: London.

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

Hastie, T., and Loader, C. (1993), "Local regression: automatic kernel carpentry," *Statistical Science 8*, 120-143.

Koo, J.Y. (1997), "Spline estimation of discontinuous regression functions," Journal of Computational and Graphical Statistics 6, 266-284.

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

Loader, C.R. (1999), "Bandwidth selection: classical or plug-in?", *The Annals of Statistics 27,* 415-438.

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

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

Qiu, P., Chappell, R., Obermeyer, W., and Benca, R. (1999), "Modelling daily and subdaily cycles in rat sleep data," *Biometrics 55*, 930-935.

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

Shea, D.J., Worley, S.J., Stern, I.R., and Hoar, T.J. (1994),"An introduction to atmospheric and oceanographic data," NCAR/TN-404+IA, Climate and Global Dynamics Division, National Center For Atmospheric Research, Boulder, Colorado.

Shiau, J.H., Wahba, G., and Johnson, D.R. (1986), "Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two- and three-dimensional objective analysis," *Journal of Atmospheric and Oceanic Technology 3*, 714-725.

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.