

# Jump-preserving regression and smoothing using local linear fitting: a compromise\*

Irène Gijbels<sup>1</sup>, Alexandre Lambert<sup>2</sup> and Peihua Qiu<sup>3</sup>

August 4, 2005

## Abstract

This paper deals with nonparametric estimation of a regression curve, where the estimation method should preserve possible jumps in the curve. At each point  $x$  at which one wants to estimate the regression function, the method chooses in an adaptive way among three estimates: a local linear estimate using only datapoints to the left of  $x$ , a local linear estimate based on only datapoints to the right of  $x$ , and finally a local linear estimate using data in a two-sided neighbourhood around  $x$ . The choice among these three estimates is made by looking at differences of the weighted residual mean squares of the three fits. The resulting estimate preserves the jumps well and in addition gives smooth estimates of the continuity parts of the curve. This property of compromise between local smoothing and jump-preserving is what distinguishes our method from most previously proposed methods, that mainly focused on local smoothing and consequently blurred possible jumps, or mainly focused on jump-preserving and hence led to rather noisy estimates in continuity regions of the underlying regression curve. Strong consistency of the estimator is established and its performance is tested via a simulation study. This study also compares the current method with some existing methods. The current method is illustrated in analyzing a real dataset.

KEY WORDS: Consistency, Jump-preserving estimation, Local linear fit, Nonparametric regression, Smoothing, Weighted Residual Mean Square.

## 1 Introduction

When one wants to estimate a regression function that possibly shows a discontinuous behaviour at certain places, two smoothing approaches have been adopted in the literature. The first approach, which we will call the *indirect* approach, estimates the discontinuity locations first and then considers different segments of the design interval, on which the underlying function is assumed to be continuous. The final estimate of the regression function is obtained by using a

---

<sup>1</sup> University Center of Statistics, University of Leuven, Belgium; <sup>2</sup> Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium; <sup>3</sup> School of Statistics, University of Minnesota, USA. \*This research was supported by ‘Projet d’Actions de Recherche Concertées’, No. 98/03–217, and by the IAP research network nr. P5/24 of the Belgian government. For the third author it was also supported by a grant from National Security Agency of USA.

conventional estimator (e.g., a local linear estimator) on each of the segments. Such an approach can provide estimates of jump sizes simultaneously. By this approach, good estimation of the regression function depends on accurate estimates of the jump locations. It also requires users to implement each of its two steps properly. As long as the jump locations are accurately estimated, its estimates in continuity regions often show the desired smoothness. The second approach, called the *direct* approach, estimates the regression curve directly, without first estimating the number and locations of discontinuity points. By this approach one starts with the idea that each point in the design interval is a potential discontinuity point and thus the curve estimation method should adapt at each point to a possible discontinuity. Therefore it is convenient to use and also preserves potential jumps well. A consequence of this built-in flexibility is that the resulting estimates often show a quite ‘unsmooth’ behaviour in regions where the underlying regression function is actually continuous. For both approaches one can use methods based on kernel smoothing, splines, wavelets, etc.

The two approaches mentioned above are quite different in nature and a thorough comparison of them is lacking. It is even questionable whether such a comparison makes much sense because their major objectives are quite different. By the first approach, the major goal is to obtain good estimates of jump locations and jump sizes. Estimation of the regression function in its continuity regions is often secondary. On the other hand, the major goal of the second approach is to obtain an *overall* good estimation of the regression function, and it should be noted that a good overall estimate of the regression function may not necessarily reveal jump locations and sizes well. Due to this difference in major goals, it is even difficult to choose an appropriate criterion for comparisons.

This paper suggests a “greedy” jump-preserving curve estimation method which has the properties that: (i) it is still a direct method and therefore jumps are not detected explicitly before curve estimation, (ii) its curve estimate behaves ‘smoother’ in the continuity regions than those of most existing direct methods, and (iii) it still preserves possible jumps well. Therefore the proposed new methodology combines the major benefits of existing direct and indirect jump-curve estimation methods. It might be greedy to achieve these goals simultaneously.

For estimating continuous regression functions, there exist many smoothing methods in the literature. Among the kernel smoothing methods there are the Nadaraya-Watson estimator, the Gasser-Müller estimator, the local linear kernel estimator, and several others. In this paper we focus on local linear kernel smoothing because of its special merits when estimating regression functions at boundary points. The discontinuity points of a discontinuous regression function are similar in nature to boundary points of a continuous regression function because the discontinuous regression function is continuous within the interval formed by two consecutive discontinuity points. Therefore it would be advantageous to use local linear kernel smoothing in the context of discontinuous curve estimation. See Fan and Gijbels (1996) for a more complete discussion about local linear kernel smoothing techniques.

It is known that the conventional local linear kernel estimator using observations in a two-sided neighbourhood of a given point has a good smoothing property. But it assumes that the true regression function is smooth, and hence would not preserve jumps. Around a jump point a better way for constructing the local linear kernel estimate is to use data points located on a

single (left or right) side of that point, avoiding the dependence on the smoothness assumption of the underlying regression function at the point. The basic idea of our method is to consider three possible estimates at each point  $x$ : the conventional local linear estimate, the local linear estimate using data only to the left of  $x$ , and the local linear estimate using data only to the right of  $x$ . Ideally, one should use the conventional local linear estimate at continuity points of the regression function, and one of the two one-sided estimates near or at discontinuity points. Since the jump locations are often unknown, we suggest a data-driven criterion for choosing among the three estimates as well for choosing the involved bandwidth parameter.

Methods based on similar ideas have been proposed in the literature. McDonald and Owen (1986) suggested to obtain at each point three local linear fits, via least squares, and to construct a ‘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 Titterton (1992) proposed an alternative to this split linear fitting method. They considered at each point three nearest-neighbour type estimates, and propose various diagnostics to decide whether the regression function was continuous at that point or not. The method of Hall and Titterton (1992) is in fact an indirect method, since it searched for discontinuity points first and then constructed the curve estimate in each continuity region. Recently Qiu (2003) proposed a jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. For each point  $x$  two one-sided local linear estimates were considered, and based on a comparison of the residual sums of squares of the two one-sided fits the curve estimate at  $x$  was defined by one of the two estimates or their average. This estimate preserves jumps quite well, but shows a quite ‘rough’ behaviour in continuity regions of the underlying regression curve, as mentioned at the beginning of this section regarding direct jump curve estimation methods. The proposed method in this paper represents an improvement in this regard, by compromising between local smoothing and jump-preserving.

The literature on indirect estimation methods includes Müller (1992), Wu and Chu (1993), Eubank and Speckman (1994), Müller and Song (1997), Qiu and Yandell (1998), Gijbels, Hall and Kneip (1999), Kang, Koo and Park (2000), Gijbels and Goderniaux (2004), among others. Spline estimation of discontinuous regression functions has been discussed in Koo (1997).

A completely different adaptive estimation procedure that simultaneously adapts to the smoothness of the regression curve and to its discontinuities as well has been proposed by Spokoiny (1998). Our method distinguishes from this method by its computational simplicity and by the fact that it is a non-iterative procedure. Other relevant references are the papers on Sigma filtering by Lee (1983) and on M-smoothers by Chu, Glad, Godtliebsen and Marron (1998) and Rue, Chu, Godtliebsen and Marron (2002).

The paper is organized as follows. In Section 2 we first briefly introduce the estimation method proposed by Qiu (2003) and then discuss possible improvements based on theoretical considerations. We focus on one particular improved estimation method and establish its strong consistency in Section 3. Proofs of the theoretical results are deferred to the Appendix. The estimation method involves selection of a threshold and a smoothing parameter, which is discussed in Section 4. A simulation study investigating the performances of the proposed methods is provided in Section 5. An application of the proposed method to real data is demonstrated in Section 6.

## 2 Estimation procedures

In this section, we first discuss some preliminary results about the three local linear kernel estimates (i.e., the conventional estimate and the two one-sided estimates), and then introduce several jump-preserving curve estimation procedures, all based on the three local linear kernel estimates.

### 2.1 Model and preliminary results

Consider the regression model

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

where the  $x_i = i/n$  are  $n$  equally-spaced design points in the design interval  $[0, 1]$ ,  $\varepsilon_i$  are  $n$  *iid* random errors with zero mean and finite variance  $\sigma^2$ , and  $f$  is a nonparametric regression function. In this paper, we consider the case when  $f$  has jumps at points  $s_q$  in  $[0, 1]$  with jump magnitudes  $d_q$ , for  $q = 1, \dots, m$ . Without loss of generality,  $f$  is assumed to be right-continuous at all jump locations. The number of jumps  $m$ , the jump locations  $s_q$ 's and the jump magnitudes  $d_q$ 's are all unknown. In (2.1) the design points are assumed equally-spaced for convenience. All methodologies developed in this paper can actually be applied to cases when the design points are unequally-spaced or even random, under some regularity conditions. Moreover, the methodologies also apply when the conditional variance is a function of  $x$ , denoted as  $\sigma^2(x)$ , i.e. in the heteroscedastic case. The theoretical results have been proved for the random design and heteroscedastic case, but for ease of presentation, we present here only the proofs for the fixed design and homoscedastic case.

Let  $K$  be a bounded symmetric density kernel function supported on  $[-1/2, 1/2]$ . Two one-sided kernel functions are defined by:

$$K_\ell(x) = \begin{cases} K(x), & \text{when } x \in [-1/2, 0) \\ 0, & \text{otherwise} \end{cases}; \quad K_r(x) = \begin{cases} K(x), & \text{when } x \in [0, 1/2] \\ 0, & \text{otherwise} \end{cases}.$$

Namely,  $K_\ell$  is defined as the left-half of  $K$  on its support and  $K_r$  is the right-half. Then three sets of local linear estimators of  $f$  and its first-order derivative  $f'$  at  $x$  are considered. The *left* local linear estimator is defined by:

$$(\hat{a}_{\ell,0}(x), \hat{a}_{\ell,1}(x)) = \arg \min_{a,b} \sum_{i=1}^n [Y_i - a - b(x_i - x)]^2 K_\ell\left(\frac{x_i - x}{h_n}\right), \quad (2.2)$$

where  $h_n > 0$  is a bandwidth parameter. The *right* and *conventional* local linear estimators, denoted by respectively  $(\hat{a}_{r,0}(x), \hat{a}_{r,1}(x))$  and  $(\hat{a}_{c,0}(x), \hat{a}_{c,1}(x))$  are obtained by using respectively  $K_r$  and  $K$  instead of  $K_\ell$  in (2.2). Obviously the estimators  $(\hat{a}_{c,0}(x), \hat{a}_{c,1}(x))$  are the usual local linear estimators of  $f(x)$  and  $f'(x)$ , based on data in the neighbourhood  $[x - \frac{h_n}{2}, x + \frac{h_n}{2}]$  of  $x$ . The first set of estimators are based on data in the left-sided interval  $[x - \frac{h_n}{2}, x)$ , and the second estimator relies only on data in the right-sided interval  $[x, x + \frac{h_n}{2}]$ .

The quality of the three local linear fits can be evaluated via their Weighted Residual Mean Squares (WRMSs), defined as

$$\text{WRMS}_\ell(x) = \frac{\sum_i [Y_i - \hat{a}_{\ell,0} - \hat{a}_{\ell,1}(x_i - x)]^2 K_\ell\left(\frac{x_i - x}{h_n}\right)}{\sum_i K_\ell\left(\frac{x_i - x}{h_n}\right)}, \quad (2.3)$$

for the left local linear estimate. The residual quantities  $\text{WRMS}_r(x)$  and  $\text{WRMS}_c(x)$  are defined similarly by replacing  $(\hat{a}_{\ell,0}, \hat{a}_{\ell,1}, K_\ell)$  by respectively  $(\hat{a}_{r,0}, \hat{a}_{r,1}, K_r)$  and  $(\hat{a}_{c,0}, \hat{a}_{c,1}, K)$  in (2.3). If  $f$  is continuous around a given point  $x$ , then all three estimators  $\hat{a}_{\ell,0}(x)$ ,  $\hat{a}_{r,0}(x)$  and  $\hat{a}_{c,0}(x)$  are consistent estimators of  $f(x)$  (cf. Section 2.2 below). The following proposition tells us that all three weighted residual mean square quantities are consistent estimators of the error variance  $\sigma^2$  in such cases.

**Proposition 2.1** *Assume that  $f$  has continuous second-order derivatives in  $[0, 1] \setminus \bigcup_{q=1}^m [s_q - \frac{h_n}{2}, s_q + \frac{h_n}{2}]$ ; the kernel function  $K$  is uniformly Lipschitz continuous;  $h_n \rightarrow 0$  and  $nh_n \rightarrow \infty$ . Then at any  $x \in [\frac{h_n}{2}, 1 - \frac{h_n}{2}] \setminus \bigcup_{q=1}^m [s_q - \frac{h_n}{2}, s_q + \frac{h_n}{2}]$ , we have*

$$\text{WRMS}_j(x) = \sigma^2 + R_{n,j,1}(x), \quad j = \ell, r, c, \quad (2.4)$$

where  $R_{n,\ell,1}(x)$ ,  $R_{n,r,1}(x)$  and  $R_{n,c,1}(x)$  are random variables tending to 0 almost surely and uniformly in  $x \in [\frac{h_n}{2}, 1 - \frac{h_n}{2}] \setminus \bigcup_{q=1}^m [s_q - \frac{h_n}{2}, s_q + \frac{h_n}{2}]$ .

The proof of this result is along the same lines as the proof of Theorem 3.2 in Qiu (2003) and is omitted here.

The behaviours of the weighted residual mean squares are quite different near jump points. Next we discuss this behaviour for an arbitrary jump point  $s$  with associated jump magnitude  $d$ . Any point  $x$  in the neighbourhood  $[s - \frac{h_n}{2}, s + \frac{h_n}{2}]$  of  $s$  can be denoted as

$$x = s + \tau h_n, \quad \text{with } \tau \in [-1/2, 1/2]. \quad (2.5)$$

In the left-half of the neighbourhood, i.e., for  $x = s + \tau h_n$  with  $\tau \in [-\frac{1}{2}, 0)$ , intuition tells us that only  $\text{WRMS}_\ell(x)$  provides a consistent estimator for  $\sigma^2$  and the other two weighted residual mean squares would be affected by the jump at  $s$ . Similarly, in the right-half of the neighbourhood, i.e. for  $x = s + \tau h_n$  with  $\tau \in [0, \frac{1}{2}]$ , only  $\text{WRMS}_r(x)$  provides a consistent estimator for  $\sigma^2$ . The following proposition formally states these results. For points  $x = s + \tau h_n$ ,  $\tau \in [-1/2, 1/2]$ , in the neighbourhood of a jump point  $s$ , we introduce the following notations.

$$\begin{aligned} C_{\tau,j}^2 &= q_j^{-2} \int_{-\tau}^{1/2} \left[ \int_{-1/2}^{-\tau} s_{1,j}(x) K_j(x) dx - u \int_{-\tau}^{1/2} s_{0,j}(x) K_j(x) dx \right]^2 K_j(u) du \\ &\quad + q_j^{-2} \int_{-1/2}^{-\tau} \left[ \int_{-\tau}^{1/2} s_{1,j}(x) K_j(x) dx + u \int_{-\tau}^{1/2} s_{0,j}(x) K_j(x) dx \right]^2 K_j(u) du \quad (2.6) \end{aligned}$$

where  $j$  denotes  $r, \ell, c$  respectively (with  $K_c = K$ ) and

$$q_j = v_{0,j}v_{1,j} - v_{1,j}^2, \quad s_{0,j}(x) = v_{0,j}x - v_{1,j}, \quad s_{1,j}(x) = v_{2,j} - v_{1,j}x \text{ and } v_{k,j} = \int_{-1/2}^{1/2} u^k K_j(u) du. \quad (2.7)$$

**Proposition 2.2** *Assume that  $f$  has continuous second-order derivatives in  $[s - \frac{h_n}{2}, s + \frac{h_n}{2}]$  except at a jump point  $s$  (with jump magnitude  $d$ ) at which  $f$  has a second-order left derivative (for (i) below) or a second-order right derivative (for (ii) below); the kernel function  $K$  is uniformly Lipschitz continuous;  $h_n \rightarrow 0$  and  $nh_n \rightarrow \infty$ . Then, we have*

(i). for any  $x = s + \tau h_n$  with  $\tau \in [-\frac{1}{2}, 0)$ :

$$\begin{aligned} WRMS_\ell(x) &= \sigma^2 + R_{n,\ell,2}(x), \\ WRMS_r(x) &= \sigma^2 + d^2 C_{\tau,r}^2 + R_{n,r,2}(x), \\ WRMS_c(x) &= \sigma^2 + d^2 C_{\tau,c}^2 + R_{n,c,2}(x), \end{aligned} \quad (2.8)$$

where  $R_{n,\ell,2}(x), R_{n,r,2}(x)$  and  $R_{n,c,2}(x)$  are random variables tending to 0 almost surely and uniformly in  $x \in [s - \frac{h_n}{2}, s)$ ;

(ii). for any  $x = s + \tau h_n$  with  $\tau \in [0, \frac{1}{2}]$ :

$$\begin{aligned} WRMS_\ell(x) &= \sigma^2 + d^2 C_{\tau,\ell}^2 + R_{n,\ell,3}(x), \\ WRMS_r(x) &= \sigma^2 + R_{n,r,3}(x), \\ WRMS_c(x) &= \sigma^2 + d^2 C_{\tau,c}^2 + R_{n,c,3}(x), \end{aligned} \quad (2.9)$$

where  $R_{n,\ell,3}(x), R_{n,r,3}(x)$  and  $R_{n,c,3}(x)$  are random variables tending to 0 almost surely and uniformly in  $x \in [s, s + \frac{h_n}{2}]$ .

The expressions for  $WRMS_\ell(x)$  and  $WRMS_r(x)$  in (2.8) and (2.9) were derived by Qiu (2003). The expressions for  $WRMS_c(x)$  can be derived in a similar way.

Since our curve estimation method depends heavily on the three weighted residual mean squares, which in turn depend on the quantities  $C_{\tau,\ell}^2$ ,  $C_{\tau,r}^2$  and  $C_{\tau,c}^2$ , we explore these quantities a bit further.

From their expressions, the quantities  $C_{\tau,\ell}^2$ ,  $C_{\tau,r}^2$  and  $C_{\tau,c}^2$  depend on  $\tau$  and the kernel function  $K$ . When  $K$  is one of the Uniform, Epanechnikov and Triangular kernels on  $[-1/2, 1/2]$ , the three quantities are shown in Figure 2.1 as functions of  $\tau$ . From Figure 2.1, we can discover some of their common properties: all three quantities seem to be bimodal functions of  $\tau$  and this bimodality is stronger for kernels with smaller values at 0. Furthermore, the quantities  $C_{\tau,\ell}^2$  and  $C_{\tau,r}^2$  are symmetric versions of one another and  $C_{\tau,c}^2$  is always symmetric about  $\tau = 0$ . Moreover, we have  $C_{\tau,r}^2 \leq C_{\tau,c}^2$  for all values of  $\tau$ . All these properties have been proved formally. Expressions for finding the local maxima of the three functions of  $\tau$  are also available.

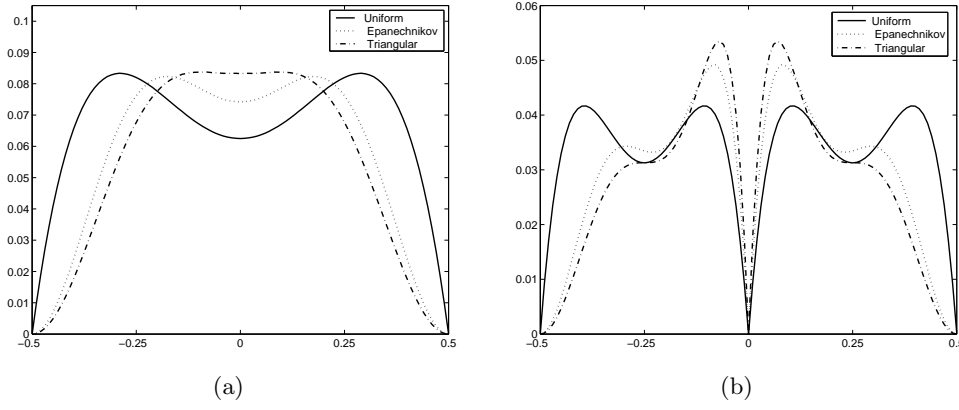


Figure 2.1: (a) Plots of  $C_{\tau,c}^2$ ; (b) Plots of  $C_{\tau,\ell}^2$  and  $C_{\tau,r}^2$ , for the three kernel functions (Uniform, Epanechnikov and Triangular).

## 2.2 Estimation procedures

Qiu (2003) recently proposed the following estimator of  $f(x)$ :

$$\hat{f}_1(x) = \begin{cases} \hat{a}_{\ell,0}(x) & \text{if } \text{WRMS}_{\ell}(x) < \text{WRMS}_r(x) \\ \hat{a}_{r,0}(x) & \text{if } \text{WRMS}_{\ell}(x) > \text{WRMS}_r(x) \\ (\hat{a}_{\ell,0}(x) + \hat{a}_{r,0}(x))/2 & \text{if } \text{WRMS}_{\ell}(x) = \text{WRMS}_r(x), \end{cases} \quad (2.10)$$

which is defined by one of the left and right estimates with smaller WRMS value. Qiu (2003) proved that  $\hat{f}_1(x)$  is a consistent estimator of  $f(x)$  in the entire design interval. In practice it appears that this estimator preserves jumps well, but is quite noisy in continuity regions of  $f$ , due to the fact that only one-sided (left- or right-sided) observations are used in its construction.

In order to get a better insight into the different behaviour of the three estimators  $\hat{a}_{\ell,0}(x)$ ,  $\hat{a}_{r,0}(x)$  and  $\hat{a}_{c,0}(x)$ , next we look at their asymptotic bias and variance expressions, in cases when  $x$  is far away from any jump points and when  $x$  is close to a jump point. The following two propositions formally state the asymptotic Mean Squared Error (MSE) expressions for the three estimates.

**Proposition 2.3** *Assume that  $f$  has continuous second-order derivatives in  $[0, 1] \setminus \bigcup_{q=1}^m [s_q - \frac{h_n}{2}, s_q + \frac{h_n}{2}]$ ; the kernel function  $K$  is uniformly Lipschitz continuous;  $h_n \rightarrow 0$  and  $nh_n \rightarrow \infty$ . Then at any  $x \in [\frac{h_n}{2}, 1 - \frac{h_n}{2}] \setminus \bigcup_{q=1}^m [s_q - \frac{h_n}{2}, s_q + \frac{h_n}{2}]$ , we have*

$$\text{MSE}(\hat{a}_{j,0}(x)) = \left[ \frac{h_n^2}{2} f''(x) B_{j,K} \right]^2 + \frac{\sigma^2}{nh_n} V_{j,K} + o\left(h_n^4 + \frac{1}{nh_n}\right), \quad j = \ell, r, c, \quad (2.11)$$

where

$$\begin{aligned}
B_{\ell,K} &= \frac{v_{2,\ell}^2 - v_{1,\ell}v_{3,\ell}}{v_{0,\ell}v_{2,\ell} - v_{1,\ell}^2}, & V_{\ell,K} &= \int_{-1/2}^0 K^2(u) \left[ \frac{v_{2,\ell} - v_{1,\ell}u}{v_{2,\ell}v_{0,\ell} - v_{1,\ell}^2} \right]^2 du, \\
B_{r,K} &= \frac{v_{2,r}^2 - v_{1,r}v_{3,r}}{v_{0,r}v_{2,r} - v_{1,r}^2}, & V_{r,K} &= \int_0^{1/2} K^2(u) \left[ \frac{v_{2,r} - v_{1,r}u}{v_{2,r}v_{0,r} - v_{1,r}^2} \right]^2 du, \\
B_{c,K} &= \int_{-1/2}^{1/2} u^2 K(u) du = v_{2,c}, & V_{c,K} &= \int_{-1/2}^{1/2} K^2(u) du,
\end{aligned} \tag{2.12}$$

with  $v_{k,\ell}$  and  $v_{k,r}$  ( $k = 1, 2, 3$ ) as defined in (2.7).

For a formal proof of these results readers are referred to Fan and Gijbels (1996). From Proposition 2.3 we can conclude that the three estimators are consistent in mean square sense and have the same rate of convergence in continuity regions of  $f$ . The only differences among them appear in the terms of (2.12). For the Epanechnikov kernel, as an example, we have  $B_{\ell,K} = B_{r,K} \approx -0.02$ ,  $B_{c,K} = 0.05$ ,  $V_{\ell,K} = V_{r,K} \approx 8.995$  and  $V_{c,K} = 1.2$ . So asymptotically the right or the left estimator with the same  $h_n$  increases variance by a factor of 7.5 compared to the conventional estimator, while decreases bias by a factor of 2.5.

Proposition 2.4 below provides expressions for asymptotic biases and variances of the three estimators  $\widehat{a}_{\ell,0}(x)$ ,  $\widehat{a}_{r,0}(x)$  and  $\widehat{a}_{c,0}(x)$  on a left-sided or right-sided interval around a jump point  $s$ .

**Proposition 2.4** *Assume that  $f$  has continuous second-order derivatives in  $[s - \frac{h_n}{2}, s + \frac{h_n}{2}]$  except at  $s$  at which  $f$  has a second-order left derivative (for (i) below) or a second-order right derivative (for (ii) below); the kernel function  $K$  is uniformly Lipschitz continuous;  $h_n \rightarrow 0$  and  $nh_n \rightarrow \infty$ . Then, we have*

(i). for any  $x = s + \tau h_n$  with  $\tau \in [-\frac{1}{2}, 0)$ :

$$\begin{aligned}
MSE(\widehat{a}_{\ell,0}(x)) &= \left[ \frac{h_n^2}{2} f''(s-) B_{\ell,K} \right]^2 + \frac{\sigma^2}{nh_n} V_{\ell,K} + o(h_n^4 + \frac{1}{nh_n}), \\
MSE(\widehat{a}_{r,0}(x)) &= \left[ d \int_{|\tau|}^{1/2} K_r(u) \frac{v_{2,r} - v_{1,r}u}{v_{0,r}v_{2,r} - v_{1,r}^2} du \right]^2 + \frac{\sigma^2}{nh_n} V_{r,K} + o(1), \\
MSE(\widehat{a}_{c,0}(x)) &= \left[ d \int_{|\tau|}^{1/2} K(u) du \right]^2 + \frac{\sigma^2}{nh_n} V_{c,K} + o(1);
\end{aligned} \tag{2.13}$$

(ii). for any  $x = s + \tau h_n$  with  $\tau \in [0, \frac{1}{2}]$ :

$$\begin{aligned}
MSE(\widehat{a}_{\ell,0}(x)) &= \left[ -d \int_{-1/2}^{-|\tau|} K_{\ell}(u) \frac{v_{2,\ell} - v_{1,\ell}u}{v_{0,\ell}v_{2,\ell} - v_{1,\ell}^2} du \right]^2 + \frac{\sigma^2}{nh_n} V_{\ell,K} + o(1), \\
MSE(\widehat{a}_{r,0}(x)) &= \left[ \frac{h_n^2}{2} f''(s+) B_{r,K} \right]^2 + \frac{\sigma^2}{nh_n} V_{r,K} + o(h_n^4 + \frac{1}{nh_n}), \\
MSE(\widehat{a}_{c,0}(x)) &= \left[ -d \int_{-1/2}^{-|\tau|} K(u) du \right]^2 + \frac{\sigma^2}{nh_n} V_{c,K} + o(1),
\end{aligned} \tag{2.14}$$

where  $f''(s-)$  (respectively  $f''(s+)$ ) denotes the left-hand (respectively right-hand) second-order derivative of  $f$  at  $s$  and  $d$  is the jump magnitude of  $f$  at the point  $s$ .



A formal proof for the result about  $\text{MSE}(\widehat{a}_{c,0}(x))$  can be found in Hamrouni (1999) and Grégoire and Hamrouni (2002). The results about  $\text{MSE}(\widehat{a}_{\ell,0}(x))$  and  $\text{MSE}(\widehat{a}_{r,0}(x))$  can be proved along the same lines. The asymptotic expressions in (2.13) reveal that  $\widehat{a}_{c,0}(x)$  and  $\widehat{a}_{r,0}(x)$  are not consistent at any point in the neighborhood  $[s - h_n/2, s]$  which is  $\tau h_n$  away from  $s$  with  $\tau \in [-\frac{1}{2}, 0)$ . A similar discussion can be given for points on the right-side interval of the jump point  $s$  by using expressions (2.14). Since these left and right neighbourhoods tend to the empty set when  $n$  tends to  $\infty$ , the Mean Integrated Squared Error of the three estimates all tends to zero as  $n$  tends to  $\infty$ . See Hamrouni (1999) for details about this result.

As an illustration of these asymptotic expressions, we plot in Figure 2.2 the asymptotic bias functions, as function of  $x$  on  $[0.1, 0.9]$ , for the three estimators and the regression function  $f = g_1$  considered in Section 5 (see (5.1) and the top left panel of Figure 5.1). From Figure 2.2 it can be seen that all three estimates have relatively low biases in continuity regions (equal to zero in linear regions and depending on the second derivative and the constants  $B_{j,K}$  as in (2.12) in the sinusoidal regions). Note also that one of the two one-sided estimates has a considerably smaller bias than the other one near a jump point.

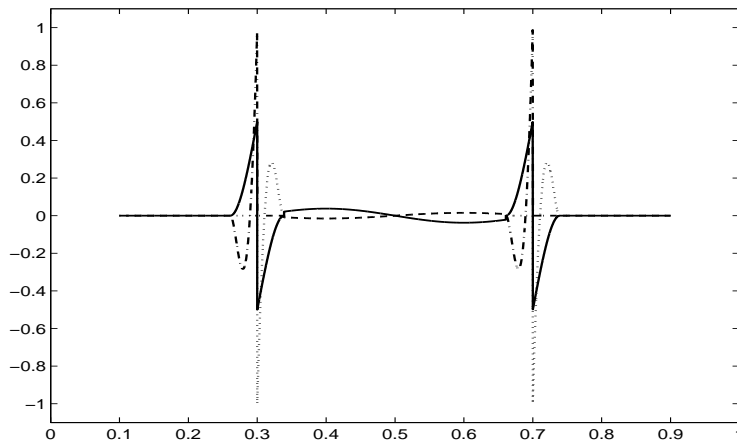


Figure 2.2: Asymptotic bias functions of  $\widehat{a}_{c,0}(x)$  (solid curve),  $\widehat{a}_{\ell,0}(x)$  (dotted curve) and  $\widehat{a}_{r,0}(x)$  (dashed-dotted curve) for the regression function  $f = g_1$  defined in (5.1), using  $h_n = 0.078$ .

Based on this detailed discussion about the MSEs of the three estimators and the discussion about their weighted residual mean squares, provided in Propositions 2.1 and 2.2, next we propose some curve estimation procedures which try to balance jump-preservation and local smoothing.

As mentioned above,  $\widehat{f}_1$  is quite noisy in continuity regions of  $f$ .

To overcome this problem, we propose two different modifications. By the first modification

we introduce the conventional estimator  $\widehat{a}_{c,0}$  in (2.10), and define

$$\widehat{f}_2(x) = \begin{cases} \widehat{a}_{c,0}(x) & \text{if } \frac{\text{WRMS}_c(x)}{C} \leq \min(\text{WRMS}_\ell(x), \text{WRMS}_r(x)) \\ \widehat{a}_{\ell,0}(x) & \text{if } \frac{\text{WRMS}_c(x)}{C} > \min(\text{WRMS}_\ell(x), \text{WRMS}_r(x)) = \text{WRMS}_\ell(x) \\ \widehat{a}_{r,0}(x) & \text{if } \frac{\text{WRMS}_c(x)}{C} > \min(\text{WRMS}_\ell(x), \text{WRMS}_r(x)) = \text{WRMS}_r(x) \\ (\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x))/2 & \text{if } \frac{\text{WRMS}_c(x)}{C} > \text{WRMS}_\ell(x) = \text{WRMS}_r(x). \end{cases} \quad (2.15)$$

The constant  $C$  should be chosen such that the conventional estimate is selected in continuity regions of  $f$  (i.e.,  $\frac{\text{WRMS}_c(x)}{C} \leq \min(\text{WRMS}_\ell(x), \text{WRMS}_r(x))$  in such cases), and one of the two one-sided estimates is selected near a jump point (i.e.,  $\frac{\text{WRMS}_c(x)}{C} > \min(\text{WRMS}_\ell(x), \text{WRMS}_r(x))$ ). The case  $C \leq 1$  should be avoided since it can be shown that often  $\text{WRMS}_c(x) > \min(\text{WRMS}_\ell(x), \text{WRMS}_r(x))$  under some regularity conditions (cf. Lemma A.1 in the Appendix), which implies that the conventional linear estimate will almost never be selected in the continuity regions of  $f$  by the above definition. By Propositions 2.1 and 2.2, the above two requirements are asymptotically equivalent to

$$1 < C < 1 + \frac{d^2}{\sigma^2} C_{\tau,c}^2, \quad (2.16)$$

which depends on the signal-to-noise ratio  $d/\sigma$ .

By the second modification, the three WRMS's are compared by their differences instead of ratios in (2.15), and the resulting estimate is defined by:

$$\widehat{f}_3(x) = \begin{cases} \widehat{a}_{c,0}(x) & \text{if } \text{diff}(x) \leq u \\ \widehat{a}_{\ell,0}(x) & \text{if } \text{diff}(x) > u \text{ and } \text{WRMS}_\ell(x) < \text{WRMS}_r(x) \\ \widehat{a}_{r,0}(x) & \text{if } \text{diff}(x) > u \text{ and } \text{WRMS}_\ell(x) > \text{WRMS}_r(x) \\ (\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x))/2 & \text{if } \text{diff}(x) > u \text{ and } \text{WRMS}_\ell(x) = \text{WRMS}_r(x), \end{cases} \quad (2.17)$$

where  $u$  is a threshold value and

$$\text{diff}(x) = \max(\text{WRMS}_c(x) - \text{WRMS}_\ell(x), \text{WRMS}_c(x) - \text{WRMS}_r(x)). \quad (2.18)$$

In continuity regions of  $f$ ,  $\text{WRMS}_\ell(x)$ ,  $\text{WRMS}_r(x)$  and  $\text{WRMS}_c(x)$  are all consistent estimates of  $\sigma^2$ . So  $\text{diff}(x)$  is close to zero in such cases. From Proposition 2.2 we know that  $\text{diff}(x) = d^2 C_{\tau,c}^2 + o(1)$ , *a.s.*, near a jump point. By combining these two properties of  $\text{diff}(x)$ , the threshold value  $u$  needs to be selected such that

$$0 < u < d^2 C_{\tau,c}^2. \quad (2.19)$$

In contrast to the (asymptotic) constraint on  $C$  (cf. (2.16)), we can see that the (asymptotic) constraint on  $u$  does not have the error variance  $\sigma^2$  involved, which leads to certain advantages for the second modification, as shown by the simulation study in Section 5.

It should be mentioned that in case of multiple jump points, the constraints (2.16) and (2.19) should be replaced by

$$1 < C < 1 + \frac{d_q^2}{\sigma^2} C_{\tau_q,c}^2 \text{ and } 0 < u < d_q^2 C_{\tau_q,c}^2,$$

for  $q = 1, 2, \dots, m$ , where  $\tau_q$  is defined by  $\tau_q = \frac{x-sq}{h_n}$ .

Practical choices of the parameters  $h_n$ ,  $C$ ,  $u$  involved in the modified estimates will be discussed in Section 4.

### 3 Consistency

From our extensive simulation study in Section 5, it appears that  $\widehat{f}_3$  has some preferable properties compared to the other estimates  $\widehat{f}_1$  and  $\widehat{f}_2$ . In this section, its strong consistency is proved. First, we establish the uniform strong consistency of the three estimates  $\widehat{a}_{c,0}(x)$ ,  $\widehat{a}_{\ell,0}(x)$  and  $\widehat{a}_{r,0}(x)$ , on which  $\widehat{f}_3$  is based.

**Theorem 3.1** *If  $f$  is second-order differentiable on  $[0, 1]$ ,  $f''$  is uniformly bounded on  $[0, 1]$ , the kernel function  $K$  is uniformly Lipschitz continuous, and  $h_n \rightarrow 0$ ,  $nh_n^3 \rightarrow \infty$  and  $\sqrt{\frac{nh_n^5}{\ln n}} \rightarrow 0$  as  $n$  goes to  $\infty$ , then for any  $0 < \rho < 1$  we have:*

$$\begin{aligned} \sup_{x \in [\rho, 1-\rho]} \sqrt{\frac{nh_n}{\ln n}} |\widehat{a}_{c,0}(x) - f(x)| &= O(1), a.s., \\ \sup_{x \in [\rho, 1]} \sqrt{\frac{nh_n}{\ln n}} |\widehat{a}_{\ell,0}(x) - f(x)| &= O(1), a.s., \\ \sup_{x \in [0, 1-\rho]} \sqrt{\frac{nh_n}{\ln n}} |\widehat{a}_{r,0}(x) - f(x)| &= O(1), a.s., \end{aligned}$$

for  $n$  sufficiently large.

Theorem 3.2 below states that  $\widehat{f}_3$  is uniformly strong consistent on the interval  $[0, 1]$  excluding all neighbourhoods of jump points. In the neighbourhoods of jump points, it is pointwise consistent. Finally, in the neighbourhoods of jump points excluding any small regions around the neighborhood endpoints and centers, it is still uniformly consistent. The three different regions mentioned above are denoted by:

$$\begin{aligned} D_1 &= [\rho, 1 - \rho] \setminus \bigcup_{q=1}^m [s_q - h_n/2, s_q + h_n/2] \\ D_2 &= \bigcup_{q=1}^m [s_q - h_n/2, s_q + h_n/2] \\ D_{2,\delta} &= \bigcup_{q=1}^m \left\{ [s_q - (1/2 - \delta)h_n, s_q - \delta h_n] \cup [s_q + \delta h_n, s_q + (1/2 - \delta)h_n] \right\}, \end{aligned} \quad (3.1)$$

where  $0 < \rho < 1$  and  $0 < \delta < 1/4$  are two constants.

**Theorem 3.2** *Suppose that  $f$  is second-order differentiable and  $f''$  is uniformly bounded on  $[0, 1]$  except the jump points  $\{s_q, q = 1, \dots, m\}$  at which  $f$  has left and right bounded second-order derivatives. It is further assumed that the kernel function  $K$  is uniformly Lipschitz continuous, the bandwidth  $h_n$  satisfies the conditions that  $h_n \rightarrow 0$ ,  $nh_n^3 \rightarrow \infty$  and  $\sqrt{\frac{nh_n^5}{\ln n}} \rightarrow 0$ , and the threshold  $u = u_n \rightarrow 0$  as  $n$  goes to  $\infty$ . Then we have:*

(i)

$$\sup_{x \in D_1} \sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_3(x) - f(x)| = O(1), a.s.;$$

(ii) For any  $x \in D_2$ ,

$$\sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_3(x) - f(x)| = O(1), a.s.;$$

(iii)

$$\sup_{x \in D_{2,\delta}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_3(x) - f(x)| = O(1), a.s.,$$

for  $n$  sufficiently large.

The proofs of Theorems 3.1 and 3.2 are given in the Appendix. It can be proved along the same lines that the other two estimates  $\widehat{f}_1$  and  $\widehat{f}_2$  are also strong consistent for estimating  $f$ . A way to compare these estimates theoretically is via studying their biases and variances. However, it is not an easy task to derive useful formulas for their biases and variances, because of their complicated definitions in which weighted residual mean squares are used in indicator functions. In this paper, we compared these estimates by studying their finite-sample biases and variances through an extensive simulation study in Section 5.

## 4 Error criteria and choice of procedure parameters

We discuss two important issues in this section. To compare different curve estimation procedures, an error criterion is needed for measuring their performance. This is discussed in Section 4.1. Then in Section 4.2, we discuss selection of the procedure parameters.

### 4.1 Error criteria for comparison

Recall that the main objective of the curve estimation procedures discussed in Section 2 is to give good overall reconstruction of the regression function  $f$ . Therefore a natural criterion for evaluating an estimate  $\widehat{f}$  is the Mean Integrated Squared Error,  $MISE = E \left[ \int_0^1 (\widehat{f}(x) - f(x))^2 f_X(x) dx \right]$ , where  $f_X(x)$  denotes the design density in the random design case, and equals one in the fixed design case. For estimating a jump regression curve, the curve estimates also need to be jump-preserving. To measure jump-preserving around a given jump point  $s$ , we propose to use the following local MISE:

$$MISE_s = E \left[ \int_{s-0.05}^{s+0.05} (\widehat{f}(x) - f(x))^2 f_X(x) dx \right], \quad (4.1)$$

which measures the MISE between  $\widehat{f}$  and  $f$  in the interval  $[s - 0.05, s + 0.05]$ . The half-width of this interval, 0.05, is subjectively selected. In applications a reasonable choice for this number is  $h_n/2$ , by which 0.05 is reasonable for most numerical examples in Section 5.

From Proposition 2.4,  $\widehat{a}_{c,0}(x)$  is not pointwise consistent when  $x$  is in the neighborhood  $[s - h_n/2, s + h_n/2]$ , i.e.,  $x = s + \tau h_n$  with  $\tau \in [-1/2, 1/2]$ . So by the local MISE criterion, this estimate would not perform well. However, due to the facts that the width of this neighborhood is small and it often performs better than the other estimates in continuity regions of  $f$ , its global performance measured by MISE would be good. As a matter of fact, it can be checked

that its MISE is of order  $O(h_n + \frac{1}{nh_n})$ , which implies that  $\widehat{a}_{c,0}$  is  $L_2$ -consistent on the entire design interval  $[0, 1]$ .

## 4.2 Choice of procedure parameters

### 4.2.1 Bandwidth parameter $h_n$

The choice of the bandwidth parameter  $h_n$  is crucial. There exist many bandwidth selection procedures in the literature. Some of them are difficult to use here (e.g., the plug-in procedures). Some others are complicated to compute (e.g., the bootstrap procedures). For simplicity we opt for the cross-validation procedure, i.e. we select  $h_n$  as

$$\widehat{h}_n = \arg \min_{h_n} \sum_{i=1}^n \left[ Y_i - \widehat{f}^{-i}(x_i) \right]^2, \quad (4.2)$$

where  $\widehat{f}^{-i}(x_i)$  is one of the proposed estimates of  $f(x_i)$  based on all data except the  $i$ -th observation  $(x_i, Y_i)$ .

### 4.2.2 Thresholds $C$ and $u$

The threshold parameters  $C = C_n$  and  $u = u_n$  used in  $\widehat{f}_2$  and  $\widehat{f}_3$  can be selected, together with the bandwidth  $h_n$ , by the cross-validation procedure:

$$(\widehat{h}_n, \widehat{C}_n) = \arg \min_{h_n, C_n} \sum_{i=1}^n \left[ Y_i - \widehat{f}^{-i}(x_i) \right]^2, \quad (\widehat{h}_n, \widehat{u}_n) = \arg \min_{h_n, u_n} \sum_{i=1}^n \left[ Y_i - \widehat{f}^{-i}(x_i) \right]^2, \quad (4.3)$$

where  $\widehat{f}^{-i}$  is one of  $\widehat{f}_2^{-i}$  and  $\widehat{f}_3^{-i}$ .

To solve the minimization problems in (4.3), we need to specify a grid for the  $C_n$ -values and a grid for the  $u_n$ -values. Recall that the (asymptotic) constraint (2.16) needs to be imposed on  $C_n$ , which has two unknown quantities  $|d|$  and  $\sigma^2$  involved. The quantity  $|d|$  can be estimated by  $\max_{x \in [\rho, 1-\rho]} |\widehat{a}_{l,0}(x) - \widehat{a}_{r,0}(x)|$  where  $\rho > 0$  is a small constant, and  $\sigma^2$  can be estimated by the residual mean square  $\frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{f}_1(x_i))^2$  of  $\widehat{f}_1$  (recall that  $\widehat{f}_1$  preserves the jumps well). Then the range for  $C_n$ -values can be specified by (2.16). We know that when  $C = 1$ ,  $\widehat{f}_2(x)$  is close to  $\widehat{f}_1(x)$  in such a case, which is good in preserving the jumps. If  $C$  is large, then  $\widehat{f}_2(x)$  is close to  $\widehat{a}_{c,0}(x)$ , which behaves best in the continuity regions. Therefore the cross-validation procedure considers many estimates that behave in between the two estimates  $\widehat{f}_1(x)$  and  $\widehat{a}_{c,0}(x)$ . By selecting  $C_n$ , we would automatically end up with a compromise between *local smoothing* and *jump preserving*. Similar remarks can be made regarding the choice of the grid for the  $u_n$ -values, although selection of  $u_n$  is generally simpler since the (asymptotic) constraint (2.19) does not have  $\sigma^2$  involved.

One might think that choosing both parameters  $h$  and  $C$  (or  $u$ ) by the CV procedure (4.3) would require a big computational demand, but this is not true. In view of (2.15) and (2.17), once we have the three estimators ( $\widehat{a}_{c,0}$ ,  $\widehat{a}_{l,0}$  and  $\widehat{a}_{r,0}$ ) (for fixed  $h$ ), we do not have to re-compute these estimators for different values of  $C$  (or  $u$ ), which saves a great amount of computation.

## 5 Simulation Study

In this section we investigate the performance of the curve estimation procedures discussed in Section 2.2 by a simulation study. We also provide some comparison with performances of other methods.

### 5.1 Simulation models and results

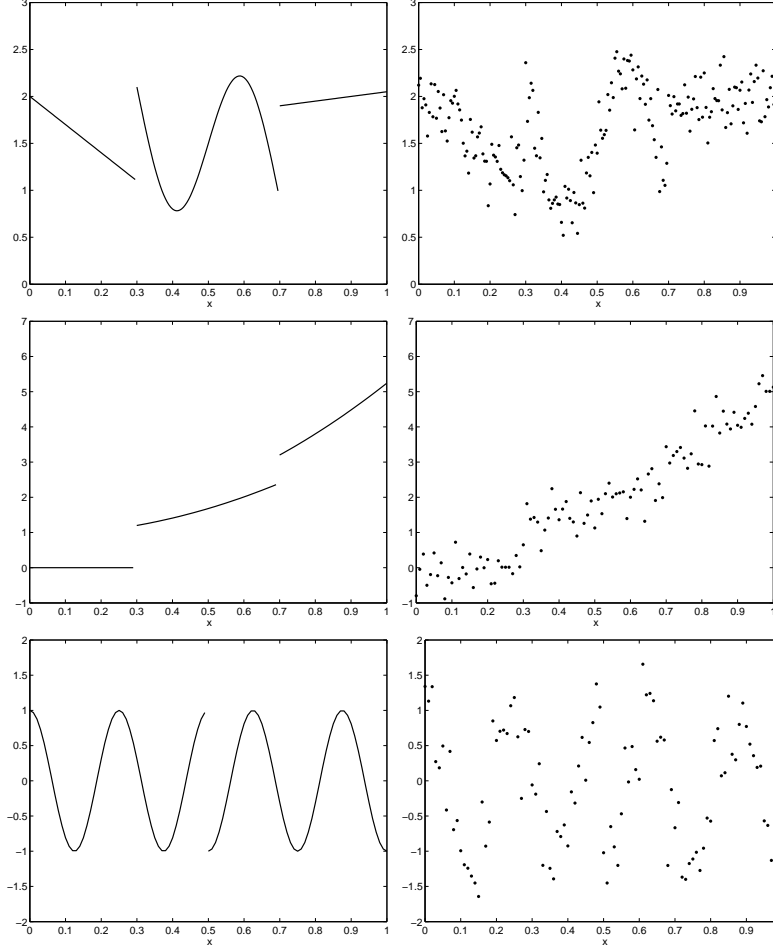


Figure 5.1: Left panel: graphs of the functions  $g_1, g_2$  and  $g_3$ . Right panel: simulated data from model (2.1) with  $f = g_1$  ( $n = 200, \varepsilon \sim N(0; 0.2^2)$ );  $f = g_2$  ( $n = 100, \varepsilon \sim N(0; 0.4^2)$ ), and  $f = g_3$  ( $n = 100, \varepsilon \sim N(0; 0.4^2)$ ).

In model (2.1), suppose that  $\varepsilon_1 \sim N(0; \sigma^2)$  and the regression function  $f$  is one of the following three functions:

$$\begin{aligned}
 g_1(x) &= \begin{cases} -3x + 2, & \text{on } [0, 0.3) \\ -3x + 3 - \sin((x - 0.3)\pi/0.2), & \text{on } [0.3, 0.7) \\ x/2 + 1.55, & \text{on } [0.7, 1] \end{cases} \\
 g_2(x) &= (3x^2 + 0.93)I[0.3 \leq x < 0.7] + (4x^2 + 1.24)I[0.7 \leq x \leq 1] \\
 g_3(x) &= \cos(8\pi(0.5 - x))I[0 \leq x < 0.5] - \cos(8\pi(0.5 - x))I[x \geq 0.5]. \quad (5.1)
 \end{aligned}$$

The left panels of Figure 5.1 depict the three regression functions. Note that  $g_1$  has two jump points at positions 0.3 and 0.7 with the same jump magnitude 1. The function  $g_2$  has two jumps at 0.3 and 0.7, with respective jump magnitudes 1.2 and 0.8. This function is chosen to investigate the effect of jumps with different sizes on the performance of the proposed estimates. The function  $g_3$  has one discontinuity at  $x = 0.5$  with jump size  $-2$ . This function might be most difficult to estimate among the three functions, because it is steep at several different locations and these locations could be confused with jump locations by the curve estimation procedures. The right panels of Figure 5.1 present noisy versions of the three regression functions with  $n$  and  $\sigma$  values as specified in the figure caption.

For a given regression function  $f$ ,  $N = 200$  samples are then generated from model (2.1) with  $\varepsilon_1 \sim N(0; \sigma^2)$ . We use the Epanechnikov kernel:  $K(x) = 1.5(1 - 4x^2)I[-1/2 \leq x \leq 1/2]$ . The MISE value of a curve estimate  $\hat{f}$  is estimated by

$$\widehat{\text{MISE}} = \frac{1}{N} \sum_{k=1}^N \widehat{\text{ISE}}_k, \quad \text{with} \quad \widehat{\text{ISE}}_k = \sum_{i=1}^{n-1} \frac{\text{SE}_k(x_i) + \text{SE}_k(x_{i+1})}{2n}, \quad (5.2)$$

where  $\text{SE}_k(x_i) = (f(x_i) - \hat{f}^k(x_i))^2$  with  $\hat{f}^k$  denoting the curve estimate  $\hat{f}$  constructed from the  $k$ -th simulated sample.

Around a jump point  $s$ , the local MISE value,  $\text{MISE}_s$ , of  $\hat{f}$  (cf. (4.1)) can be estimated by

$$\widehat{\text{MISE}}_s = \frac{1}{N} \sum_{k=1}^N \widehat{\text{ISE}}_{s,k}, \quad (5.3)$$

where

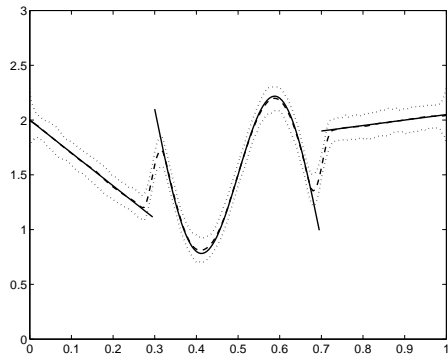
$$\widehat{\text{ISE}}_{s,k} = \sum_{i=1}^{n-1} \frac{\text{SE}_k(x_i) + \text{SE}_k(x_{i+1})}{2n} I[s - 0.05 < x_i < x_{i+1} < s + 0.05].$$

When  $f = g_1$ , 6 different values of  $\sigma^2$  are considered, with a fixed sample size  $n = 200$ . The performance of the curve estimation procedures is evaluated by their  $\widehat{\text{MISE}}$  and  $\widehat{\text{MISE}}_{s=0.3} + \widehat{\text{MISE}}_{s=0.7}$  values, which are presented in Table 5.1, in the first and second columns respectively. The procedure parameters are selected as described in Section 4.2. From the table, it can be seen that  $\hat{a}_{c,0}$  has decent  $\widehat{\text{MISE}}$  values, but its local MISE values around the jump points are relatively large, due to its bias caused by two-sided smoothing around jump points. The estimate  $\hat{f}_1$  has opposite behaviour, namely, it has relatively small  $\widehat{\text{MISE}}_{s=0.3} + \widehat{\text{MISE}}_{s=0.7}$  values (i.e., good jump-preserving), but relatively large  $\widehat{\text{MISE}}$  values (caused by its large variation in continuity regions). The other curve estimates behave in between, which implies that they indeed found the compromise between local smoothing and jump-preserving. Further, we can see that for all estimates  $\widehat{\text{MISE}}$  increases with  $\sigma^2$ . For a large value of  $\sigma^2$  all estimates (except  $\hat{f}_1$ ) show similar results. This means that for such a large value of  $\sigma^2$  smoothing might be more important than jump-preserving, and all these estimates behave like the conventional estimate in such cases. We also remark that  $\hat{f}_3$  behaves better than  $\hat{f}_2$  for median values of  $\sigma$ . This is no surprise since from their definitions provided in Section 2.2, it is already clear that  $\sigma^2$  is less involved in  $\hat{f}_3$  (also see the constraints (2.16) and (2.19)).

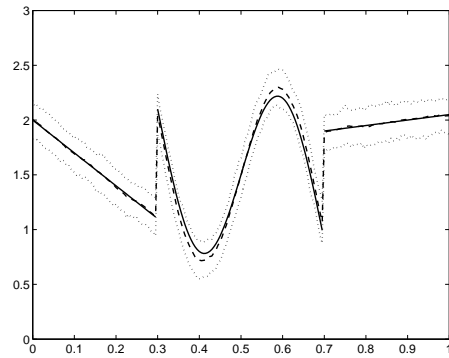
Figure 5.2 gives a graphical display of the performance of the four estimates  $\hat{a}_{c,0}$ ,  $\hat{f}_1$ ,  $\hat{f}_2$  and  $\hat{f}_3$  when  $f = g_1$ . In each plot, the true regression function  $g_1$  is represented by the solid

Table 5.1:  $\widehat{\text{MISE}}$  (first columns) and  $\widehat{\text{MISE}}_{s=0.3} + \widehat{\text{MISE}}_{s=0.7}$  values (second columns) when  $f = g_1$ ,  $N = 200$ ,  $n = 200$  and  $\varepsilon_1 \sim N(0; \sigma^2)$ .

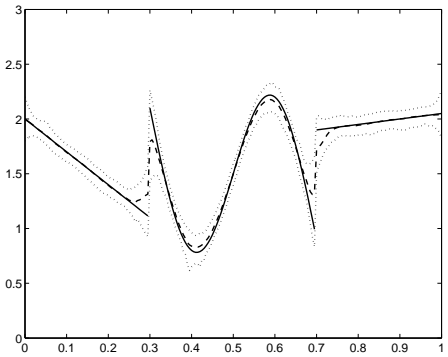
Method	$\sigma=0.1$		$\sigma=0.15$		$\sigma=0.2$		$\sigma=0.25$		$\sigma=0.3$		$\sigma=0.4$	
$\widehat{a}_{c,0}$	0.0056	0.0039	0.0084	0.0054	0.0113	0.0069	0.0144	0.0084	0.0179	0.0099	0.0247	0.0128
$\widehat{f}_1$	0.0039	0.0008	0.0075	0.0015	0.0125	0.0029	0.0190	0.0050	0.0267	0.0077	0.0444	0.0139
$\widehat{f}_2$	0.0025	0.0011	0.0065	0.0040	0.0108	0.0070	0.0146	0.0093	0.0180	0.0110	0.0243	0.0143
$\widehat{f}_3$	0.0027	0.0014	0.0055	0.0030	0.0088	0.0048	0.0127	0.0068	0.0168	0.0090	0.0248	0.0126



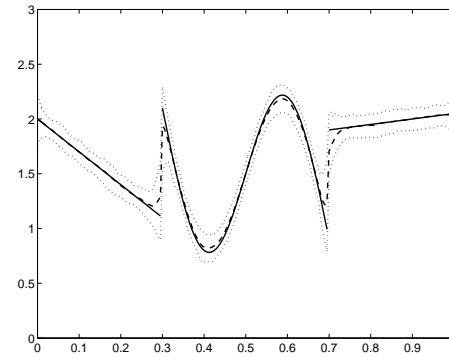
(a)



(b)



(c)



(d)

Figure 5.2: Plots of four curve estimates when  $f = g_1$ ,  $n = 200$ ,  $\sigma = 0.2$  and  $N = 200$ . In each plot, the true regression function is denoted by the solid curve. The average of  $N$  replicated fits is denoted by the dashed curve. The corresponding 5<sup>th</sup> percentile and the 95<sup>th</sup> percentile curves are denoted by dotted curves. Estimates: (a)  $\widehat{a}_{c,0}$ ; (b)  $\widehat{f}_1$ ; (c)  $\widehat{f}_2$ ; and (d)  $\widehat{f}_3$ .



curve. The dashed curve represents the average of  $N = 200$  replicated fits. The dotted curves represent the corresponding 5<sup>th</sup> percentile and the 95<sup>th</sup> percentile. The curve of the averaged fit can be seen as an estimator of  $\text{Bias}(\hat{f}(x))$ , and the difference between the upper and lower dotted curves can be regarded as a measure of  $\text{Var}(\hat{f}(x))$ . From the plots, it can be seen that the conventional estimator  $\hat{a}_{c,0}$  blurs the two jumps, but its variance is quite small. The estimate  $\hat{f}_1$  preserves the jumps very well, but its variance is quite large in continuity regions. For  $\hat{f}_2$  and  $\hat{f}_3$ , their variances are close to that of  $\hat{a}_{c,0}$ , and they preserve the jumps much better than the conventional estimate (but not as good as the estimate  $\hat{f}_1$ ). We also notice that their variances near the two jumps have been increased, which is mainly due to the fact that in these regions they take the value of one of the two one-sided estimates which have larger asymptotic variances than the conventional estimate if all of them use the same bandwidth.

Table 5.2:  $\widehat{\text{MISE}}$ ,  $\widehat{\text{MISE}}_{s=0.3}$  and  $\widehat{\text{MISE}}_{s=0.7}$  values and their integrated squared bias ( $\widehat{\text{ISB}}_s$ ) and integrated variance ( $\widehat{\text{IV}}_s$ ) decompositions when  $f = g_2$ ,  $N = 200$  and  $\varepsilon_1 \sim N(0; 0.16)$ .

sample size	Method	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_{s=0.3}$	$\widehat{\text{ISB}}_{s=0.3}$	$\widehat{\text{IV}}_{s=0.3}$	$\widehat{\text{MISE}}_{s=0.7}$	$\widehat{\text{ISB}}_{s=0.7}$	$\widehat{\text{IV}}_{s=0.7}$
$n = 200$	$\hat{a}_{c,0}$	0.0242	0.0092	0.0074	0.0018	0.0047	0.0033	0.0014
	$\hat{f}_1$	0.0311	0.0054	0.0006	0.0048	0.0052	0.0009	0.0043
	$\hat{f}_2$	0.0235	0.0090	0.0048	0.0042	0.0055	0.0035	0.0020
	$\hat{f}_3$	0.0235	0.0077	0.0025	0.0052	0.0054	0.0026	0.0028
$n = 500$	$\hat{a}_{c,0}$	0.0145	0.0057	0.0047	0.0010	0.0029	0.0021	0.0008
	$\hat{f}_1$	0.0153	0.0033	0.0005	0.0029	0.0034	0.0007	0.0027
	$\hat{f}_2$	0.0125	0.0039	0.0012	0.0027	0.0037	0.0016	0.0021
	$\hat{f}_3$	0.0121	0.0036	0.0008	0.0028	0.0036	0.0014	0.0022

When  $f = g_2$ , we consider two different sample sizes  $n = 200, 500$  and one error variance value  $\sigma^2 = 0.16$ . The global and local MISE values of various estimates are presented in Table 5.2. In addition we show here how the estimated  $\widehat{\text{MISE}}_{s=0.3}$  and  $\widehat{\text{MISE}}_{s=0.7}$  decompose into estimated Integrated Squared Bias ( $\widehat{\text{ISB}}$ ) and Integrated Variance ( $\widehat{\text{IV}}$ ), providing as such information on the separate contributions of bias and variance. We can see that for all estimates  $\widehat{\text{MISE}}$  decreases as  $n$  increases. Furthermore, the estimate  $\hat{f}_1$  has the largest  $\widehat{\text{MISE}}$ , and all other estimates have similar  $\widehat{\text{MISE}}$  values when  $n$  is small due to the fact that  $\sigma^2$  is quite large here. For the local measure  $\widehat{\text{MISE}}_{s=0.3}$  around the jump point  $s = 0.3$ , which has a large jump size, we remark that  $\hat{a}_{c,0}$  has the largest  $\widehat{\text{MISE}}_{s=0.3}$  values (due to jump blurring) and  $\hat{f}_1$  has the smallest  $\widehat{\text{MISE}}_{s=0.3}$  values (i.e., good jump preserving). The other estimates behave in between them. Around the second jump point, it can be seen that  $\hat{a}_{c,0}$  performs the best and all other estimates have similar performance. For both jumps points the integrated squared bias is largest for the conventional estimator (not jump preserving) and smallest for the estimator  $\hat{f}_1$  (privileged jump preserving). On the other hand, the estimator  $\hat{a}_{c,0}$  shows the smallest contribution from the integrated variance (privileged smoothing). Since the noise level is quite large in this example the smoothing operation results in a bigger impact on the MISE-measures, especially for the jump point with the smallest jump size (the point 0.7).

Figure 5.3 depicts the performance of the four estimates  $\hat{a}_{c,0}$ ,  $\hat{f}_1$ ,  $\hat{f}_2$  and  $\hat{f}_3$  when  $f = g_2$ . The

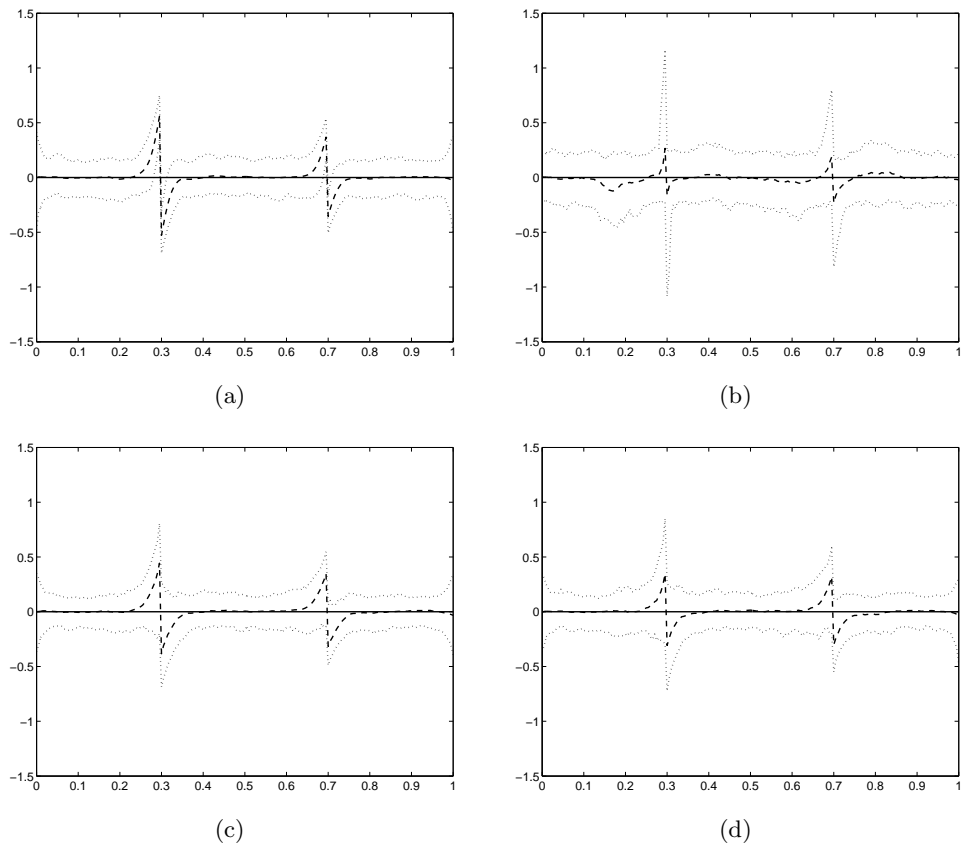


Figure 5.3: Plots of four curves  $\hat{f} - f$  when  $f = g_2$ ,  $n = 200$ ,  $\sigma = 0.4$  and  $N = 200$ . In each plot, the zero line is denoted by the solid curve. The average of  $\hat{f} - f$  using  $N$  replicated fits is denoted by a dashed curve. The corresponding  $5^{th}$  percentile and the  $95^{th}$  percentile are denoted by dotted curves. Corresponding estimates: (a)  $\hat{a}_{c,0}$ ; (b)  $\hat{f}_1$ ; (c)  $\hat{f}_2$ ; and (d)  $\hat{f}_3$ .

dashed curve represents the average of  $\widehat{f} - f$  using  $N = 200$  replicated fits. The dotted curves represent the corresponding 5<sup>th</sup> percentile and the 95<sup>th</sup> percentile. The solid line represents the zero line. Similar conclusions to those from Figure 5.2 can be drawn here, except that the first jump seems to be better preserved by  $\widehat{f}_1, \widehat{f}_2$  and  $\widehat{f}_3$ .

When  $f = g_3$ , we also use three different sample sizes  $n = 100, 200, 300$  and one error variance  $\sigma^2 = 0.16$ , as in the previous example. The simulation results are summarized in Table 5.3. We can see that for all estimates  $\widehat{\text{MISE}}$  decreases as  $n$  increases. Values of  $\widehat{\text{MISE}}$  are larger than those in the previous examples when  $f = g_1$  and  $g_2$ , because  $g_3$  is more difficult to estimate as explained before. Once again, it can be noticed that  $\widehat{f}_1$  has the largest  $\widehat{\text{MISE}}$  values but the smallest  $\widehat{\text{MISE}}_s$  values,  $\widehat{a}_{c,0}$  shows the opposite behaviour, and the other estimates behave in between them. We can also see that  $\widehat{f}_3$  performs considerably better than  $\widehat{f}_2$  in this case.

Table 5.3:  $\widehat{\text{MISE}}$  and  $\widehat{\text{MISE}}_{s=0.5}$  values when  $f = g_3$ ,  $N = 200$  and  $\varepsilon_1 \sim N(0; 0.16)$ .

Method	$n = 100$		$n = 200$		$n = 300$	
	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_{s=0.5}$	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_{s=0.5}$	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_{s=0.5}$
$\widehat{a}_{c,0}$	0.0531	0.0224	0.0346	0.0158	0.0275	0.0132
$\widehat{f}_1$	0.0909	0.0082	0.0572	0.0048	0.0433	0.0041
$\widehat{f}_2$	0.0743	0.0226	0.0355	0.0155	0.0252	0.0115
$\widehat{f}_3$	0.0484	0.0168	0.0277	0.0094	0.0210	0.0073

Figure 5.4 depicts the performance (using  $\widehat{f} - f$ ) of the four estimates  $\widehat{a}_{c,0}, \widehat{f}_1, \widehat{f}_2$  and  $\widehat{f}_3$  when  $f = g_3$ . Similar conclusions to those from Figures 5.2 and 5.3 can be drawn here. Note that the bias of the four estimates in the continuous regions has some sinusoidal behaviour. This can be explained by the sinusoidal behaviour of the second derivative of  $g_3$  which appears in the expression of the asymptotic bias of  $\widehat{a}_{c,0}, \widehat{a}_{r,0}$  and  $\widehat{a}_{\ell,0}$  (see Propositions 2.3 and 2.4).

Figure 5.5 presents the estimated bias (left panels) and variance (right panels) functions of  $\widehat{a}_{c,0}, \widehat{f}_1$  and  $\widehat{f}_3$  when  $f$  equals  $g_1$ . It can be seen from the plots that the conventional estimator  $\widehat{a}_{c,0}$  has the largest bias around jump points, as seen in Figure 2.2, and the bias of  $\widehat{f}_1$  is the smallest at such places. In continuity regions,  $\widehat{f}_1$  has relatively large bias, and the biases of  $\widehat{f}_3$  and  $\widehat{a}_{c,0}$  are similar and small, due to the fact that  $\widehat{f}_3$  most probably equals  $\widehat{a}_{c,0}$  in such cases. For the estimated variance functions, we can see that  $\widehat{f}_1$  has the largest variance in continuity regions, which is consistent with the asymptotic results obtained in Section 2.2. The variances  $\widehat{f}_2$  and  $\widehat{f}_3$  are bigger around the jump points because they have to choose among three estimates at such places, which increases their variances.

In Section 4 we proved the consistency of  $\widehat{f}_3$ , in which the threshold  $u_n$  is required to tend to zero as  $n$  tends to infinity. In the practical implementation of  $\widehat{f}_3$ ,  $u_n$  is selected by cross-validation, as explained in Section 4.2. To investigate the behaviour of the selected  $\widehat{u}_n$  values when  $n$  increases, we performed  $N = 20$  replicated simulations in each case when  $f = g_1$ ,  $\sigma^2 = 0.04$  and  $n = 100, 200, 400, 600, 1000$  and 5000. For each sample size  $n$ , the 20 selected  $\widehat{u}_n$  values from the 20 replicated simulations were retained and presented by a boxplot in Figure 5.6. From the plot, it can be seen that the cross-validation choice of  $u_n$  does tend to zero when

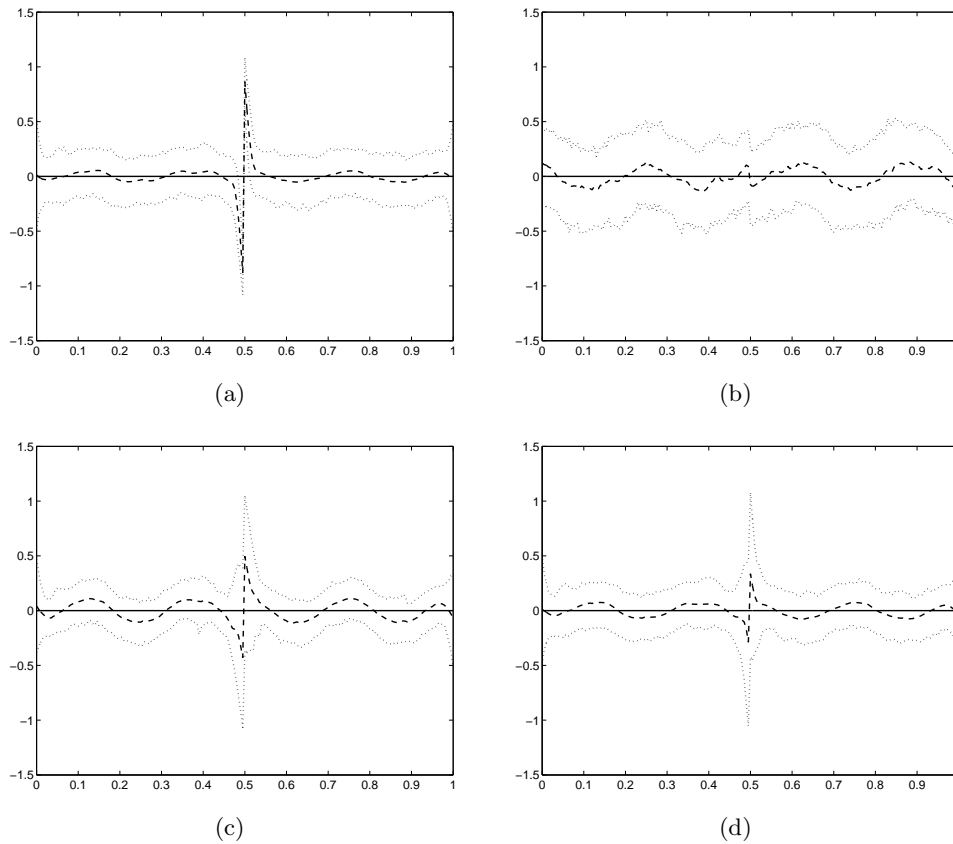


Figure 5.4: Plots of four curves  $\hat{f} - f$  when  $f = g_3$ ,  $n = 200$ ,  $\sigma = 0.4$  and  $N = 200$ . In each plot, the zero line is denoted by the solid curve. The average of  $\hat{f} - f$  using  $N$  replicated fits is denoted by a dashed curve. The corresponding 5<sup>th</sup> percentile and the 95<sup>th</sup> percentile are denoted by dotted curves. Corresponding estimates: (a)  $\hat{a}_{c,0}$ ; (b)  $\hat{f}_1$ ; (c)  $\hat{f}_2$ ; and (d)  $\hat{f}_3$ .

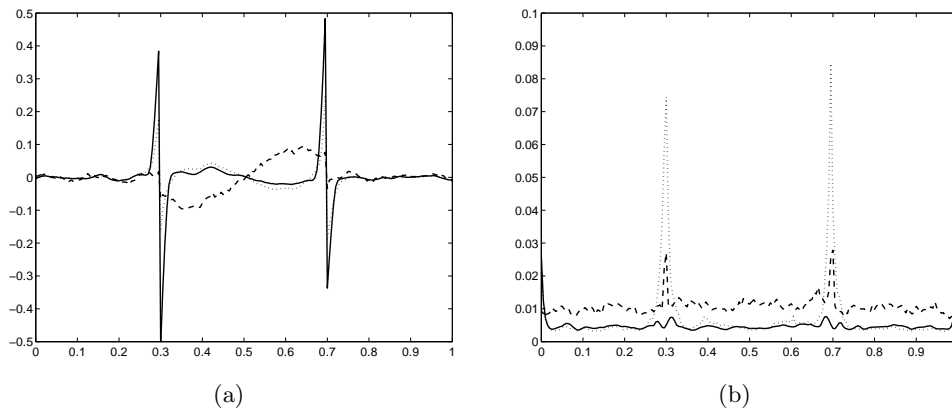


Figure 5.5: Plots of the estimated bias function (left panel) and variance function (right panel) of the estimates  $\hat{a}_{c,0}$  (solid curves),  $\hat{f}_1$  (dashed curves) and  $\hat{f}_3$  (dotted curves) for  $f = g_1$ ,  $n = 200$ ,  $\sigma = 0.2$  and  $N = 200$  simulated samples.

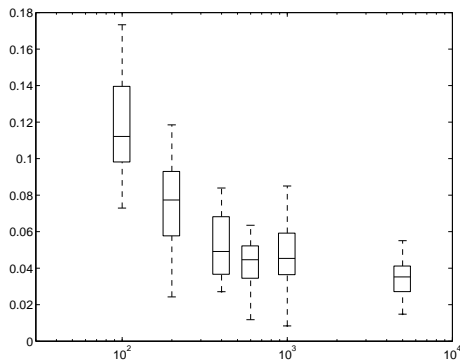


Figure 5.6: Boxplots of  $\hat{u}_n$  values obtained from 20 replicated simulations in cases when  $f = g_1$ ,  $\sigma^2 = 0.04$  and  $n = 100, 200, 400, 600, 1000$  and  $5000$ .

$n$  gets larger, which coincides with the theoretical condition imposed on  $u_n$  in Theorem 3.2.

Based on the extensive simulation study presented above, we recommend to use the estimate  $\hat{f}_3$  for estimating a jump curve because it seems to provide a good compromise between local smoothing and jump-preserving, and in most cases performs best among all estimates that were designed to search for such a compromise.

## 5.2 Comparison with some other existing methods

We now run some simulations to compare our estimator  $\hat{f}_3$  to some existing procedures in the literature which include the Sigma filter (Lee 1983), the  $M$ -constant smoother (Chu et al. 1998), and the adaptive weight smoothing (AWS) procedure (Polzehl and Spokoiny 2002, 2003). This last method is based on a similar idea to that in Spokoiny (1998). The three existing methods mentioned above are all direct methods, in the sense that they estimate jump regression functions without first detecting jump positions explicitly. The conventional local linear kernel estimator is used as a reference in this comparison.

Each related procedure has one or more parameters to select. Not all methods discuss data-driven choices for the involved parameters. Therefore, for each method, their optimum values are chosen such that they minimize the estimated MISE value based on  $N = 200$  simulations. More specifically, we choose  $h_n$  and  $u_n$  in  $\hat{f}_3$ , two bandwidths in the Sigma filter and the  $M$ -constant smoother, and  $h_{max}$  in the AWS procedure. In the AWS procedure, there are a number of other parameters which are taken to be their default values (see Polzehl and Spokoiny (2002, 2003) for more explanations). The AWS procedure is run for four different values of  $p$ , i.e.,  $p = 0, 1, 2, 3$ , where  $p$  is the degree of local polynomial used in local function approximation.

The MISE results are reported in Table 5.4. In the table,  $\hat{f}_3$  (local constant fit) corresponds to the same procedure as (2.17) except that the local constant estimator (i.e., the Nadaraya-Watson estimator) instead of the local linear estimator is used. In Section 1, we mentioned that the local linear estimator has preferable asymptotic properties near a boundary point. But, when the sample size is small to moderate, the Nadaraya-Watson estimator often leads to better MISE results because it has smaller variability. A referee also suggested to modify the proposed estimator such that the three estimates are all based on data from intervals of length  $h_n$ . The

Table 5.4:  $\widehat{MISE}$  values for several jump-preserving curve estimation procedures in the case when  $n = 200$  and  $N = 200$ .

Method	$g_1 (\sigma = 0.2)$	$g_2 (\sigma = 0.4)$	$g_3 (\sigma = 0.4)$
$\widehat{f}_3$	0.0063	0.0230	0.0232
$\widehat{f}_3$ (modified)	0.0055	0.0199	0.0214
$\widehat{f}_3$ (local constant fit)	0.0069	0.0170	0.0200
Sigma filter	0.0078	0.0186	0.0264
M-constant smoother	0.0083	0.0174	0.0238
AWS ( $p = 0$ )	0.0072	0.0159	0.0325
AWS ( $p = 1$ )	0.0093	0.0212	0.0282
AWS ( $p = 2$ )	0.0067	0.0168	0.0219
AWS ( $p = 3$ )	0.0077	0.0201	0.0211
$\widehat{a}_{c,0}$	0.0105	0.0230	0.0328

second line in Table 5.4 (entitled  $\widehat{f}_3$  (modified)) reports the results for this modified estimator. The results are slightly better. Note however that for such an estimator the boundary regions are enlarged, and that the theoretical behaviours of the Weighted Residual Mean Square quantities as well as of the difference function are more involved and need to be studied.

From the table it can be seen that all jump-preserving methods outperform the conventional local linear kernel estimator  $\widehat{a}_{c,0}$ , which is not a surprise due to the discontinuity feature of the three models considered. We can also see that  $\widehat{f}_3$  outperforms the Sigma filter and the  $M$ -constant smoother, except for model 2 when local linear estimation is used in  $\widehat{f}_3$ . The AWS procedure and  $\widehat{f}_3$  seem to be quite competitive depending on the models and the value of  $p$ . Overall, the proposed method gives reasonably good results and thus is a good competitor of the existing ones. It also has the advantage of simple computation.

## 6 Real data analysis

The data set consists of measurements in mils of the thickness of 90 US Lincoln pennies. There are two measurements each year, from 1945 through 1989. Penny thickness was reduced in World War II and restored to its original thickness sometime around 1960 and reduced again in the 70's. These data are given in Scott (1992) and are displayed in Figure 6.1. Speckman (1994) found that there were changes in thickness around the years 1958 and 1974. Similar findings were obtained by Gijbels and Goderniaux (2004) using their jump detection procedure. We depict in Figure 6.1 the three estimates  $\widehat{a}_{c,0}$ ,  $\widehat{f}_1$  and  $\widehat{f}_3$ . As expected, the conventional estimator (presented as a dotted curve) blurs the discontinuities. The estimate  $\widehat{f}_1$  (the dashed curve) preserves the discontinuities well but is quite noisy in continuity regions. Finally, the estimate  $\widehat{f}_3$  (the solid curve) is close to  $\widehat{f}_1$  in the discontinuity regions and close to  $\widehat{a}_{c,0}$  in the continuous regions. Hence  $\widehat{f}_3$  preserves well the discontinuities and removes the noise efficiently in continuity regions.

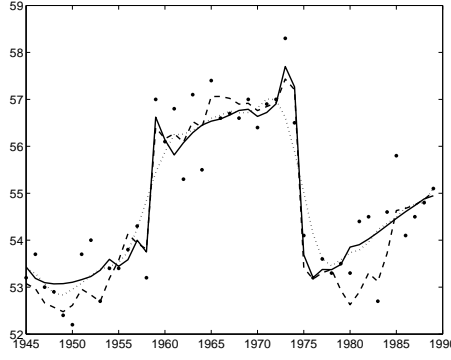


Figure 6.1: The data (small dots) and the estimates  $\widehat{a}_{c,0}(x)$  (dotted curve),  $\widehat{f}_1(x)$  (dashed curve) and  $\widehat{f}_3(x)$  (solid curve). The cross-validation scores of the three estimates are 0.8475, 1.0568 and 0.7380, respectively.

## Appendix

In this section, we provide proofs for Theorems 3.1 and 3.2. We first give two lemmas which will be used in the proofs of the theorems.

**Lemma A.1** *For any  $x \in [0, 1]$ , we have*

$$WRMS_c(x) \geq \min(WRMS_\ell(x), WRMS_r(x)).$$

**Proof:** Let  $g_\ell(x; a, b)$ ,  $g_r(x; a, b)$  and  $g_c(x; a, b)$  denote the objective functions in (2.2). Then it is easy to see that

$$g_c(x; a, b) = g_\ell(x; a, b) + g_r(x; a, b), \text{ for any } a, b, x. \quad (\text{A.1})$$

The three WRMS's in (2.3) can be written as

$$WRMS_\ell(x) = g_\ell(x; \widehat{a}_{\ell,0}(x), \widehat{a}_{\ell,1}(x)) / \sum_{x_i < x} k_i(x), \quad (\text{A.2})$$

where  $k_i(x) = K((x_i - x)/h_n) \geq 0$  for all  $i$ . For  $WRMS_r(x)$  (respectively  $WRMS_c(x)$ ) replace the index  $\ell$  by  $r$  (respectively  $c$ ) and  $\sum_{x_i < x}$  by  $\sum_{x_i \geq x}$  (respectively  $\sum_{x_i}$ ) in (A.2). By (A.1) and (A.2), we find

$$\begin{aligned} WRMS_c(x) &= \{g_\ell(x; \widehat{a}_{c,0}(x), \widehat{a}_{c,1}(x)) + g_r(x; \widehat{a}_{c,0}(x), \widehat{a}_{c,1}(x))\} / \sum_{x_i} k_i(x) \\ &\geq \{g_\ell(x; \widehat{a}_{\ell,0}(x), \widehat{a}_{\ell,1}(x)) + g_r(x; \widehat{a}_{r,0}(x), \widehat{a}_{r,1}(x))\} / \sum_{x_i} k_i(x) \\ &= \frac{\sum_{x_i < x} k_i(x)}{\sum_{x_i} k_i(x)} \left( \frac{g_\ell(x; \widehat{a}_{\ell,0}(x), \widehat{a}_{\ell,1}(x))}{\sum_{x_i < x} k_i(x)} \right) + \frac{\sum_{x_i \geq x} k_i(x)}{\sum_{x_i} k_i(x)} \left( \frac{g_r(x; \widehat{a}_{r,0}(x), \widehat{a}_{r,1}(x))}{\sum_{x_i \geq x} k_i(x)} \right) \\ &\geq \min(WRMS_\ell(x), WRMS_r(x)). \end{aligned}$$

This finishes the proof.  $\square$

Before discussing the next lemma, we first notice that the three estimates  $\widehat{a}_{c,0}(x)$ ,  $\widehat{a}_{r,0}(x)$  and  $\widehat{a}_{\ell,0}(x)$  have the following expressions:

$$\widehat{a}_{j,0}(x) = \sum_{i=1}^n Y_i K_j \left( \frac{x_i - x}{h_n} \right) \frac{w_{2,j} - w_{1,j}(x_i - x)}{w_{2,j}w_{0,j} - w_{1,j}^2}, \quad (\text{A.3})$$

for  $x \in [0, 1]$ ,  $j = c, r, \ell$  and

$$w_{k,j} = \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) (x_i - x)^k, \text{ for } k = 0, 1, 2, 3. \quad (\text{A.4})$$

Please notice that  $w_{3,j}$  does not appear in (A.3). But it will be used in the proof below. So we also give its definition here. Clearly  $w_{k,j}$  depends on  $x$ , which is not explicit in its notation for simplicity. For the three estimates, their interior and boundary regions, which are referred to several times below, are listed explicitly in Table A.1, for readers' convenience.

Table A.1: Interior and boundary regions of the three estimates.

Estimate	Interior region	Boundary region(s)
$\widehat{a}_{c,0}$	$(\frac{h_n}{2}, 1 - \frac{h_n}{2})$	$[0, \frac{h_n}{2}] \cup [1 - \frac{h_n}{2}, 1]$
$\widehat{a}_{r,0}$	$[0, 1 - \frac{h_n}{2})$	$[1 - \frac{h_n}{2}, 1]$
$\widehat{a}_{\ell,0}$	$(\frac{h_n}{2}, 1]$	$[0, \frac{h_n}{2}]$

**Lemma A.2** *If the kernel function  $K$  is uniformly Lipschitz continuous, then for any interior point  $x$ , we have*

$$\frac{w_{k,j}}{nh_n^{k+1}} = v_{k,j} + O\left(\frac{1}{nh_n}\right), \quad (\text{A.5})$$

where  $v_{k,j} = \int_{-1/2}^{1/2} u^k K_j(u) du$ ,  $j = c, r, \ell$  and  $k = 0, 1, 2, 3$ . For each  $j$ , this equation is uniformly true for all interior points.

The proof of Lemma A.2 is quite straightforward, by using the uniform Lipschitz continuity property of the kernel function  $K$ . It is therefore omitted. We are now ready to prove Theorems 3.1 and 3.2. The proof of Theorem 3.1 is based on some ideas from the proof of Theorem 3.1 in Qiu (2003) and some ideas from Masry (1996).

**Proof of Theorem 3.1:**

First, we can write

$$\widehat{a}_{j,0}(x) - f(x) = [\widehat{a}_{j,0}(x) - E\widehat{a}_{j,0}(x)] + [E\widehat{a}_{j,0}(x) - f(x)]. \quad (\text{A.6})$$



By (A.3) and Lemma A.2, we find

$$\begin{aligned}
E\widehat{a}_{j,0}(x) &= \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \frac{w_{2,j} - w_{1,j}(x_i - x)}{w_{2,j}w_{0,j} - w_{1,j}^2} f(x_i) \\
&= \frac{1}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \frac{v_{2,j} - v_{1,j} \left( \frac{x_i - x}{h_n} \right)}{v_{2,j}v_{0,j} - v_{1,j}^2} f(x_i) + O(1/nh_n) \\
&= \int K_j(u) \frac{v_{2,j} - v_{1,j}u}{v_{2,j}v_{0,j} - v_{1,j}^2} f(x + uh_n) du + O(1/nh_n) \\
&= \int K_j(u) \frac{v_{2,j} - v_{1,j}u}{v_{2,j}v_{0,j} - v_{1,j}^2} \left[ f(x) + uh_n f'(x) + \frac{(uh_n)^2}{2} f''(x) + o(h_n^2) \right] du + O(1/nh_n) \\
&= f(x) + \frac{(v_{2,j}^2 - v_{1,j}v_{3,j})f''(x)}{2(v_{2,j}v_{0,j} - v_{1,j}^2)} h_n^2 + o(h_n^2) + O(1/nh_n).
\end{aligned}$$

Since  $f''(x)$  is bounded uniformly, the above equation implies

$$\begin{aligned}
\sup_{x \in (h_n/2, 1-h_n/2)} |E\widehat{a}_{c,0}(x) - f(x)| &= O(h_n^2), \\
\sup_{x \in (h_n/2, 1]} |E\widehat{a}_{\ell,0}(x) - f(x)| &= O(h_n^2), \\
\sup_{x \in [0, 1-h_n/2)} |E\widehat{a}_{r,0}(x) - f(x)| &= O(h_n^2),
\end{aligned} \tag{A.7}$$

for  $n$  sufficiently large.

The first term on the right hand side of (A.6) can be written as:

$$\begin{aligned}
\widehat{a}_{j,0}(x) - E\widehat{a}_{j,0}(x) &= \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \frac{w_{2,j} - w_{1,j}(x_i - x)}{w_{2,j}w_{0,j} - w_{1,j}^2} \varepsilon_i \\
&= \frac{\frac{w_{2,j}}{nh_n^3}}{\frac{w_{2,j}}{nh_n^3} \frac{w_{0,j}}{nh_n} - \left( \frac{w_{1,j}}{nh_n^2} \right)^2} r_0(x) - \frac{\frac{w_{1,j}}{nh_n^2}}{\frac{w_{2,j}}{nh_n^3} \frac{w_{0,j}}{nh_n} - \left( \frac{w_{1,j}}{nh_n^2} \right)^2} r_1(x),
\end{aligned} \tag{A.8}$$

where for  $s = 0, 1$ ,

$$r_s(x) = \frac{1}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \left( \frac{x_i - x}{h_n} \right)^s \varepsilon_i. \tag{A.9}$$

By Lemma A.2, for  $k = 1, 2$ ,

$$\frac{\frac{w_{k,j}}{nh_n^{k+1}}}{\frac{w_{2,j}}{nh_n^3} \frac{w_{0,j}}{nh_n} - \left( \frac{w_{1,j}}{nh_n^2} \right)^2} = \frac{v_{k,j}}{v_{2,j}v_{0,j} - v_{1,j}^2} + O\left(\frac{1}{nh_n}\right). \tag{A.10}$$

Next we study the properties of  $r_0(x)$  and  $r_1(x)$ . Toward this end, we define, for any given  $\delta \in (0, 1)$ ,

$$\tilde{\varepsilon}_i = \varepsilon_i I[|\varepsilon_i| < t_i], \text{ where } t_i = \sqrt{i \ln i (\ln \ln i)^{1+\delta}} \text{ when } i \geq 3, \text{ and } t_1 = t_2 = t_3. \tag{A.11}$$

Let  $\tilde{r}_s(x)$  be the truncated version of  $r_s(x)$ , defined by

$$\tilde{r}_s(x) = \frac{1}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \left( \frac{x_i - x}{h_n} \right)^s \tilde{\varepsilon}_i.$$

Then  $r_s(x)$  can be written as

$$r_s(x) = [r_s(x) - \tilde{r}_s(x)] + [\tilde{r}_s(x) - E\tilde{r}_s(x)] + [E\tilde{r}_s(x) - Er_s(x)] \equiv A_s(x) + B_s(x) + C_s(x). \quad (\text{A.12})$$

Obviously,

$$A_s(x) = \frac{1}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \left( \frac{x_i - x}{h_n} \right)^s [\varepsilon_i - \tilde{\varepsilon}_i].$$

Since  $P(|\varepsilon_n| \geq t_n) \leq t_n^{-2} E\varepsilon_1^2 = \sigma^2 t_n^{-2}$ ,  $\sum_{n=1}^{\infty} P(|\varepsilon| \geq t_n) \leq 2 + \sigma^2 \sum_{n=3}^{\infty} t_n^{-2} < \infty$ . By the Borel-Cantelli lemma,  $P(|\varepsilon_n| \geq t_n \text{ i.o.}) = 0$ , or equivalently,  $P(\varepsilon_n \neq \tilde{\varepsilon}_n, \text{i.o.}) = 0$ . So there exists a full set  $\Omega_0$  (i.e.,  $P(\Omega_0) = 1$ ) such that for every  $\omega \in \Omega_0$  there exists a finite integer  $N(\omega) > 0$  with the property that:  $\varepsilon_n(\omega) = \tilde{\varepsilon}_n(\omega)$  when  $n \geq N(\omega)$ . So for any  $\omega \in \Omega_0$ ,

$$|A_s(x)| \leq \frac{1}{nh_n} \sum_{i=1}^{N(\omega)-1} K_j \left( \frac{x_i - x}{h_n} \right) \left( \frac{x_i - x}{h_n} \right)^s |\varepsilon_i(\omega) - \tilde{\varepsilon}_i(\omega)| \leq \frac{C(N(\omega), K)}{nh_n},$$

where  $C(N(\omega), K)$  is a constant depending on  $\omega$  and  $K$ , but not on  $x$ . Therefore

$$\limsup_{n \rightarrow \infty} \sup_x \sqrt{\frac{nh_n}{\ln n}} |A_s(x)| = 0, \text{ a.s.} \quad (\text{A.13})$$

For the term  $C_s(x)$  in (A.12), we have

$$\begin{aligned} |C_s(x)| &= \left| \frac{1}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \left( \frac{x_i - x}{h_n} \right)^s E(\tilde{\varepsilon}_i - \varepsilon_i) \right| \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) E(|-\varepsilon_i I(|\varepsilon_i| \geq t_i)|) \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \frac{1}{t_i} E(|\varepsilon_i|^2 I(|\varepsilon_i| \geq t_i)) \\ &\leq \frac{\sigma^2}{nh_n} \sum_{i=1}^n K_j \left( \frac{x_i - x}{h_n} \right) \frac{1}{t_i}, \text{ since } E(|\varepsilon_i|^2 I(|\varepsilon_i| \geq t_i)) \leq E(\varepsilon_i^2) = \sigma^2 \\ &= \frac{\sigma^2 \|K\|}{nh_n} \sum_{i=i_0+1}^{j_0} \frac{1}{t_i}, \end{aligned}$$

where  $\|K\| = \sup_{x \in [-1/2, 1/2]} |K(x)|$ , and where  $i_0 \leq j_0$  are two integers such that  $\{x_{i_0+1}, \dots, x_{j_0-1}\}$  is the longest sequence of the design points at which the weights  $K((x_i - x)/h_n)$  are non-zero. Obviously,  $j_0 - i_0 \leq nh_n + 2$ ,  $1/t_i < 1/\sqrt{i}$ , and  $\sum_{i=i_0+1}^{j_0} \frac{1}{\sqrt{i}} \leq \sum_{i=1}^{\lfloor nh_n+2 \rfloor} \frac{1}{\sqrt{i}}$ , where  $\lfloor x \rfloor$  denotes the integer part of  $x$ . So

$$\begin{aligned} |C_s(x)| &\leq \frac{\sigma^2 \|K\|}{nh_n} \sum_{i=1}^{\lfloor nh_n+2 \rfloor} \frac{1}{\sqrt{i}} \\ &\leq \frac{\sigma^2 \|K\|}{nh_n} \left[ 1 + \int_1^{\lfloor nh_n+2 \rfloor} \frac{1}{\sqrt{u}} du \right], \text{ since } \sum_{i=1}^a \frac{1}{\sqrt{i}} \leq 1 + \int_1^a \frac{1}{\sqrt{u}} du \\ &\leq \frac{2\sigma^2 \|K\|}{nh_n} [\sqrt{nh_n} + 2]. \end{aligned}$$

Therefore,

$$\lim_{n \rightarrow \infty} \sup_x \sqrt{\frac{nh_n}{\ln n}} |C_s(x)| \leq 2\sigma^2 \|K\| \lim_{n \rightarrow \infty} \left( \frac{nh_n}{nh_n \sqrt{\ln n}} + \frac{2}{\sqrt{nh_n \ln n}} \right) = 0. \quad (\text{A.14})$$

To handle the term  $B_s(x)$ , we define  $G_n = \{\frac{i}{d_n}; i = 1, \dots, d_n\}$  with  $d_n = \lceil \sqrt{\frac{n^2(\ln \ln n)^{1+\delta}}{h_n^3}} \rceil$ . Then for any  $x \in [0, 1]$ , there exists  $v(x) \in G_n$  such that  $|x - v(x)| \leq d_n^{-1}$ . We first write

$$\begin{aligned} \tilde{r}_s(x) - E\tilde{r}_s(x) &= [\tilde{r}_s(x) - \tilde{r}_s(v(x))] + [\tilde{r}_s(v(x)) - E\tilde{r}_s(v(x))] + [E\tilde{r}_s(v(x)) - E\tilde{r}_s(x)] \\ &\equiv Q_{1,s}(x) + Q_{2,s}(v(x)) + Q_{3,s}(x). \end{aligned} \quad (\text{A.15})$$

For  $Q_{1,s}(x)$ , by using the uniform Lipschitz continuity of  $K_j(x)x^s$  we have

$$\begin{aligned} |Q_{1,s}(x)| &\leq \frac{1}{nh_n} \sum_{i=1}^n |\tilde{\varepsilon}_i| \left| K_j \left( \frac{x_i - x}{h_n} \right) \left( \frac{x_i - x}{h_n} \right)^s - K_j \left( \frac{x_i - v(x)}{h_n} \right) \left( \frac{x_i - v(x)}{h_n} \right)^s \right| \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n t_n C_K \left| \frac{v(x) - x}{h_n} \right| \\ &\leq \frac{C_K}{nh_n} \sum_{i=1}^n \frac{t_n}{h_n d_n} \\ &\leq C_K \frac{1}{h_n} \frac{\sqrt{n \ln n (\ln \ln n)^{1+\delta} h_n^3}}{h_n \sqrt{n^2 (\ln \ln n)^{1+\delta}}} \\ &\leq C_K \sqrt{\frac{\ln n}{nh_n}}, \end{aligned}$$

where  $C_K$  is a constant depending on  $K$  only. The same result is true for  $Q_{3,s}(x)$ , since  $E|\tilde{\varepsilon}_i| \leq t_n$ . Therefore

$$\sup_x \sqrt{\frac{nh_n}{\ln n}} |Q_{1,s}(x)| = O(1), \text{ a.s.}, \text{ and } \sup_x \sqrt{\frac{nh_n}{\ln n}} |Q_{3,s}(x)| = O(1). \quad (\text{A.16})$$

The other term  $Q_{2,s}(v(x))$  can be written as

$$Q_{2,s}(v(x)) = \frac{1}{nh_n} \sum_{i=1}^n (\tilde{\varepsilon}_i - E\tilde{\varepsilon}_i) K_j \left( \frac{x_i - v(x)}{h_n} \right) \left( \frac{x_i - v(x)}{h_n} \right)^s \equiv \frac{1}{n} \sum_{i=1}^n U_{n,i}.$$

Let us divide the sequence  $\{U_{n,i}; i = 1, \dots, n\}$  into  $2q_n$  blocks of size  $\ell_n$  each, and a residual block of size  $< 2\ell_n$ . Then  $n = 2q_n \ell_n + \nu_n$  with  $\nu_n < 2\ell_n$ . Let  $\ell_n = \lfloor \frac{\sqrt{nh_n}}{t_n \sqrt{\ln n}} \rfloor \rightarrow 0$  as  $n \rightarrow \infty$ , and

$$\begin{aligned} V_n(m) &= \frac{1}{n} \sum_{t=(m-1)\ell_n+1}^{m\ell_n} U_{n,t}, \quad W'_n(v(x)) = \sum_{m=1}^{q_n} V_n(2m-1), \\ W''_n(v(x)) &= \sum_{m=1}^{q_n} V_n(2m), \quad W'''_n(v(x)) = \frac{1}{n} \sum_{t=2q_n \ell_n + 1}^n U_{n,t}. \end{aligned}$$

Then

$$Q_{2,s}(v(x)) = W'(v(x)) + W''(v(x)) + W'''(v(x)). \quad (\text{A.17})$$

For any  $\eta > 0$ , since  $\#G_n \leq d_n$ , we get

$$\begin{aligned}
P(\max_{a \in G_n} |Q_{2,s}(a)| > \eta) &\leq d_n [\max_{a \in G_n} P(|W'_n(a)| > \eta/3) + \max_{a \in G_n} P(|W''_n(a)| > \eta/3) + \\
&\quad \max_{a \in G_n} P(|W'''_n(a)| > \eta/3)] \\
&\equiv I_1 + I_2 + I_3.
\end{aligned} \tag{A.18}$$

We next study the three terms on the right-hand side of (A.18). First, notice that

$$|V_n(m)| \leq \frac{\ell_n 2t_n \|K\|}{nh_n} = \frac{2\|K\|}{\sqrt{nh_n \ln n}},$$

and hence by choosing  $\lambda_n = \frac{1}{4\|K\|} \sqrt{nh_n \ln n}$ , we have  $\lambda_n |V_n(m)| \leq 1/2$ . Using the inequalities  $\exp(u) \leq 1 + u + u^2$  for  $|u| \leq 1/2$ , and  $1 + x^2 \leq \exp(x^2)$ , and the fact that  $E(U_{n,t}) = 0$ , we get

$$\begin{aligned}
\exp(\pm \lambda_n V_n(m)) &\leq 1 \pm \lambda_n V_n(m) + \lambda_n^2 V_n^2(m) \\
E(\exp(\pm \lambda_n V_n(2m-1))) &\leq 1 + \lambda_n^2 E(V_n^2(2m-1)) \leq \exp(\lambda_n^2 E(V_n^2(2m-1))).
\end{aligned} \tag{A.19}$$

For any  $a \in G_n$ , by the Chebyshev's inequality, we have

$$\begin{aligned}
P(|W'_n(a)| > \eta/3) &= P(\lambda_n |W'_n(a)| > \lambda_n \eta/3) \\
&\leq \exp\left(\frac{-\lambda_n \eta}{3}\right) E\left(\exp\left(\lambda_n \left|\sum_{m=1}^{q_n} V_n(2m-1)\right|\right)\right) \\
&\leq \exp\left(\frac{-\lambda_n \eta}{3}\right) \left\{ E\left(\exp\left(\lambda_n \sum_{m=1}^{q_n} V_n(2m-1)\right)\right) + \right. \\
&\quad \left. E\left(\exp\left(-\lambda_n \sum_{m=1}^{q_n} V_n(2m-1)\right)\right) \right\} \\
&\leq 2 \exp\left(\frac{-\lambda_n \eta}{3}\right) \exp\left(\lambda_n^2 \sum_{m=1}^{q_n} E(V_n^2(2m-1))\right),
\end{aligned}$$

where (A.19) and the fact that  $V_n$ 's are independent random variables have been used. Furthermore,

$$\begin{aligned}
\sum_{m=1}^{q_n} E(V_n^2(2m-1)) &= \frac{1}{n^2} \sum_{m=1}^{q_n} E\left(\left(\sum_{t=(2m-1-1)\ell_n+1}^{(2m-1)\ell_n} U_{n,t}\right)^2\right) \\
&\leq \frac{1}{n^2} \sum_{t=1}^n E(U_{n,t}^2) \\
&\leq \frac{1}{n^2 h_n^2} \sum_{i=1}^n K_j \left(\frac{x_i - v(x)}{h_n}\right)^2 \left(\frac{x_i - v(x)}{h_n}\right)^{2s} \sigma^2 \\
&\leq \frac{\sigma^2 \int K_j(u)^2 u^{2s} du + O(1/nh_n)}{nh_n} \\
&= \frac{A_1 + o(1)}{nh_n},
\end{aligned}$$

where  $A_1$  is a constant depending on  $K$  and  $\sigma^2$ . So we obtain

$$I_1 \leq 2d_n \exp\left(\frac{-\lambda_n \eta}{3} + 2\frac{\lambda_n^2}{nh_n} A_1\right). \tag{A.20}$$

The same bound can be found for  $I_2$  in (A.18). It works for  $I_3$  as well since

$$|W_n'''(v(x))| \leq \frac{1}{n} \sum_{t=2q_n \ell_n + 1}^n |U_{n,t}| \leq \frac{2}{nh_n} \|K\| 2 \ell_n t_n = 4 \|K\| \frac{1}{\sqrt{nh_n \ln n}}.$$

By (A.18) and (A.20), if we take  $\eta = A_2 \sqrt{\frac{\ln n}{nh_n}}$  with  $A_2 > 0$ , then we have

$$\begin{aligned} P\left(\sqrt{\frac{nh_n}{\ln n}} \max_{a \in G_n} |Q_{2,s}(a)| > A_2\right) &\leq 6d_n \exp\left(\frac{-1}{12\|K\|} \sqrt{nh_n \ln n} A_2 \sqrt{\frac{\ln n}{nh_n}} + \frac{2A_1}{16\|K\|^2} \frac{nh_n \ln n}{nh_n}\right) \\ &= 6d_n \exp\left(\frac{-A_2}{12\|K\|} \ln n + \frac{A_1}{8\|K\|^2} \ln n\right) \\ &= 6d_n n^{\frac{-2A_2\|K\|+3A_1}{24\|K\|^2}}. \end{aligned}$$

By choosing  $A_2$  such that  $\sum_{n=1}^{\infty} P(\sqrt{\frac{nh_n}{\ln n}} \max_{a \in G_n} |Q_2(a)| > A_2) \leq \sum_{n=1}^{\infty} 6d_n n^{\frac{-2A_2\|K\|+3A_1}{24\|K\|^2}} < \infty$ , we have  $P(\max_{a \in G_n} \sqrt{\frac{nh_n}{\ln n}} |Q_2(a)| > A_2, i.o) = 0$ . This means that

$$\max_{a \in G_n} \sqrt{\frac{nh_n}{\ln n}} |Q_{2,s}(a)| = O(1), \text{ a.s.} \quad (\text{A.21})$$

Combining (A.15), (A.16) and (A.21), we have

$$\sup_x \sqrt{\frac{nh_n}{\ln n}} |B_s(x)| = O(1), \text{ a.s.} \quad (\text{A.22})$$

By (A.12), (A.13), (A.14) and (A.22), we get

$$\sup_x \sqrt{\frac{nh_n}{\ln n}} |r_s(x)| = O(1), \text{ a.s.} \quad (\text{A.23})$$

From (A.8), (A.10) and (A.23), we get

$$\begin{aligned} \sup_{x \in (h_n/2, 1-h_n/2)} |\widehat{a}_{c,0}(x) - E\widehat{a}_{c,0}(x)| &= O\left(\sqrt{\frac{\ln n}{nh_n}}\right), \text{ a.s.} \\ \sup_{x \in (h_n/2, 1]} |\widehat{a}_{\ell,0}(x) - E\widehat{a}_{\ell,0}(x)| &= O\left(\sqrt{\frac{\ln n}{nh_n}}\right), \text{ a.s.} \\ \sup_{x \in [0, 1-h_n/2]} |\widehat{a}_{r,0}(x) - E\widehat{a}_{r,0}(x)| &= O\left(\sqrt{\frac{\ln n}{nh_n}}\right), \text{ a.s.} \end{aligned} \quad (\text{A.24})$$

Finally, by (A.7), (A.24) and the condition that  $\sqrt{\frac{nh_n^5}{\ln n}} \rightarrow 0$ , we can conclude that

$$\left(\sqrt{\frac{nh_n}{\ln n}}\right) \sup_{x \in (h_n/2, 1-h_n/2)} |\widehat{a}_{c,0}(x) - f(x)| = O(1), \text{ a.s.},$$

and analogue statements for  $\widehat{a}_{\ell,0}(x)$  and  $\widehat{a}_{r,0}(x)$ . This completes the proof of Theorem 3.1.  $\square$

### Proof of Theorem 3.2.

We can rewrite the estimate  $\widehat{f}_3(x)$ , defined in (2.17), as:

$$\widehat{f}_3(x) = \widehat{a}_{c,0}(x)I(A_n(x)) + \widehat{a}_{\ell,0}(x)I(B_n(x)) + \widehat{a}_{r,0}(x)I(C_n(x)) + \frac{\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x)}{2}I(D_n(x)), \quad (\text{A.25})$$

where  $A_n(x), B_n(x), C_n(x)$  and  $D_n(x)$  are the corresponding inequalities in its definition. Clearly, for any  $x \in [0, 1]$ ,

$$I(A_n(x)) + I(B_n(x)) + I(C_n(x)) + I(D_n(x)) = 1. \quad (\text{A.26})$$

Therefore one and only one inequality is true in all cases.

Recall that the remainder terms  $R_{n,a,i}(x)$  in Propositions 2.1 and 2.2, for  $a = c, \ell, r$  and  $i = 1, 2, 3$ , tend to zero almost surely and uniformly with respect to  $x$ . So there exists a full set  $\Omega_1$  such that for every  $\omega \in \Omega_1$ ,  $\lim_{n \rightarrow \infty} \sup_{x \in [\rho, 1-\rho]; a=c,\ell,r; i=1,2,3} R_{n,a,i}(x, \omega) = 0$  (for convenience of explanation, we sometimes make  $\omega$  explicit in the notation).

The remaining part of the proof is divided into three parts, which correspond to the three regions defined in (3.1).

Part One:

First, let us consider  $x \in D_1$ . Then  $x$  is in a continuity region of  $f$  and it is at least  $h_n/2$  away from any jump point. By similar arguments to those in the proof of Theorem 3.1,

$$\sqrt{\frac{nh_n}{\ln n}} \sup_{x \in D_1} |\widehat{a}_{c,0}(x) - f(x)| \leq C, a.s.$$

where  $C$  is a constant. We find the same statement for  $\widehat{a}_{\ell,0}(x)$  and  $\widehat{a}_{r,0}(x)$  involving the same constant  $C$ . Then we have

$$\begin{aligned} \sup_{x \in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{f}_3(x) - f(x)| &= \sup_{x \in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{a}_{c,0}(x) - f(x)| I(A_n(x)) + \\ &\quad \sup_{x \in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{a}_{\ell,0}(x) - f(x)| I(B_n(x)) + \\ &\quad \sup_{x \in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{a}_{r,0}(x) - f(x)| I(C_n(x)) + \\ &\quad \sup_{x \in D_1} \sqrt{\frac{\ln n}{nh_n}} \left| \frac{\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x) - 2f(x)}{2} \right| I(D_n(x)) \\ &\leq 4C, a.s.. \end{aligned}$$

So

$$\sup_{x \in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{f}_3(x) - f(x)| = O(1), a.s.$$

Part Two:

In this part, we prove the uniform consistency of  $\widehat{f}_3$  in  $D_{2,\delta}$  with  $0 < \delta < 1/4$ , which consists of two mutually exclusive sets:  $D_{2,\delta,\ell}$  and  $D_{2,\delta,r}$ , defined by

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

We first focus on  $D_{2,\delta,\ell}$ . Any point  $x$  in this set has the expression  $x = s + \tau h_n$  where  $\tau \in [-1/2 + \delta, -\delta]$  and  $s$  is one of  $s_j$ 's. We know that in this region  $\sup_{x \in D_{2,\delta,\ell}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{a}_{\ell,0}(x) - f(x)| = O(1), a.s.$  So we can find a full set  $\Omega_2 \subset \Omega_1$  such that for any  $\omega \in \Omega_2$  there exists an integer

$n(\omega) =: n_1$  which has the property that when  $n \geq n_1$ , we have  $\sup_{x \in D_{2,\delta,\ell}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{a}_{\ell,0}(x, \omega) - f(x)| \leq C$ .

For any  $x \in D_{2,\delta,\ell}$  and  $\omega \in \Omega_2$ , the r.v.  $\text{diff}(x, \omega)$  defined in (2.18) becomes

$$\text{diff}(x, \omega) = \max(d^2 C_{\tau,c}^2 + R_{n,c,2}(x, \omega) - R_{n,\ell,2}(x, \omega), d^2 [C_{\tau,c}^2 - C_{\tau,r}^2] + R_{n,c,2}(x, \omega) - R_{n,r,2}(x, \omega)),$$

where  $d^2 C_{\tau,c}^2, d^2 C_{\tau,r}^2 > 0$  for  $\tau \in (-1/2, 0)$ . Since  $\omega \in \Omega_2 \subset \Omega_1$ , we have

$$\begin{aligned} \lim_{n \rightarrow \infty} [d^2 C_{\tau,c}^2 + R_{n,c,2}(x, \omega) - R_{n,\ell,2}(x, \omega)] &= d^2 C_{\tau,c}^2 =: a_\tau \\ \lim_{n \rightarrow \infty} [d^2 [C_{\tau,c}^2 - C_{\tau,r}^2] + R_{n,c,2}(x, \omega) - R_{n,r,2}(x, \omega)] &= d^2 [C_{\tau,c}^2 - C_{\tau,r}^2] =: a_\tau - b_\tau =: c_\tau, \end{aligned}$$

which implies that

$$\lim_{n \rightarrow \infty} \text{diff}(x, \omega) = \max(a_\tau, c_\tau) = a_\tau.$$

So for any  $\eta > 0$ , there exists  $n(\omega, \eta) > 0$  such that when  $n \geq n(\omega, \eta)$

$$|\text{diff}(x, \omega) - a_\tau| < \eta, \text{ or equivalently, } a_\tau - \eta < \text{diff}(x, \omega) < a_\tau + \eta.$$

Moreover, we have  $a_\tau \geq b_\tau \geq b$ , where  $b = \inf_{\tau \in (-1/2+\delta, -\delta)} b_\tau = \inf_{\tau \in (\delta, 1/2-\delta)} b_\tau > 0$ , which is demonstrated by Figure A.1. Let us take  $\eta = \frac{b}{2}$ . Then when  $n \geq n(\omega, \frac{b}{2}) =: n_2$ ,

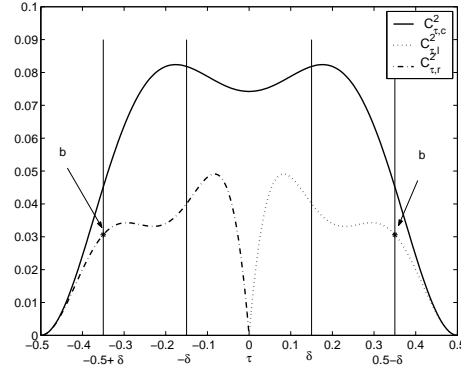


Figure A.1: A graphical illustration of the definition of  $b$  used in the proof of Theorem 3.2.

$$\text{diff}(x, \omega) > a_\tau - \frac{b}{2} \geq \frac{b}{2}.$$

Moreover, since  $u_n \rightarrow 0$ , for any  $\zeta > 0$  there exists  $n(\zeta) > 0$  such that when  $n \geq n(\zeta)$

$$-\zeta < u_n < \zeta.$$

If we take  $\zeta = \frac{b}{4}$ , we have, for  $n \geq n_3 = \max(n_2, n(\frac{b}{4}))$ ,

$$\text{diff}(x, \omega) - u_n > \frac{b}{2} - u_n > \frac{b}{2} - \frac{b}{4} = \frac{b}{4} > 0,$$

which is equivalent to

$$I(A_n(x, \omega)) = 0.$$

Now let us investigate whether the conditions  $C_n(x, \omega)$  and  $D_n(x, \omega)$  hold in such cases, which are equivalent to

$$\begin{aligned} \text{diff}(x, \omega) &> u_n \\ \text{WRMS}_\ell(x, \omega) &\geq \text{WRMS}_r(x, \omega). \end{aligned}$$

By the above argument, the first inequality is true when  $n \geq n_3$ . The second inequality can be written as :

$$b_\tau = d^2 C_{\tau,r}^2 \leq R_{n,\ell,2}(x, \omega) - R_{n,r,2}(x, \omega).$$

Since  $\omega \in \Omega_1$ , for any  $\kappa > 0$  there exists an integer  $n(\omega, \kappa) > 0$  such that for  $n \geq n(\omega, \kappa)$

$$|R_{n,\ell,2}(x, \omega) - R_{n,r,2}(x, \omega)| < \kappa.$$

Let  $\kappa = \frac{b}{2}$ . Then when  $n \geq n_4 := \max(n_3, n(\omega, \frac{b}{2}))$ , we have

$$\begin{aligned} \text{diff}(x, \omega) &> u_n \\ R_{n,\ell,2}(x, \omega) - R_{n,r,2}(x, \omega) &< \frac{b}{2} < b_\tau = d^2 C_{\tau,r}^2. \end{aligned}$$

Thus the conditions  $A_n(x, \omega)$ ,  $C_n(x, \omega)$  and  $D_n(x, \omega)$  cannot be satisfied, which implies that  $I(A_n(x, \omega)) = I(C_n(x, \omega)) = I(D_n(x, \omega)) = 0$ . By (A.26), we have  $I(B_n(x, \omega)) = 1$ . Therefore when  $n \geq n_5 := \max(n_4, n_1)$ ,

$$\sup_{D_{2,\delta,\ell}} \sqrt{\frac{nh_n}{\ln n}} |\hat{f}_3(x, \omega) - f(x)| = \sup_{D_{2,\delta,\ell}} \sqrt{\frac{nh_n}{\ln n}} |\hat{a}_{\ell,0}(x, \omega) - f(x)| \leq C.$$

Similarly, we can prove that

$$\sup_{D_{2,\delta,r}} \sqrt{\frac{nh_n}{\ln n}} |\hat{f}_3(x) - f(x)| = O(1), \text{ a.s.}$$

So

$$\sup_{D_{2,\delta}} \sqrt{\frac{nh_n}{\ln n}} |\hat{f}_3(x) - f(x)| = O(1), \text{ a.s.}$$

Part Three: When  $x \in D_2 \setminus D_{2,\delta}$ ,  $\hat{f}_3$  can be proved to be strong consistent in a similar way to the above arguments. But the consistency is not uniform with respect to  $x$  because we cannot find a unique, strictly positive, lower bound for  $a_\tau$  and  $b_\tau$  in such cases (cf. Figure A.1).  $\square$

## Acknowledgement

The authors are grateful to an anonymous referee for valuable remarks which led to an improved presentation of the paper.



## References

- Chu, C.-K., Glad, I.K., Godtliebsen, F. and Marron, J.S. (1998). Edge preserving smoothers for image processing (with discussion). *Journal of the American Statistical Association*, **93**, 526–556.
- Eubank, R.L. and Speckman, P.L. (1994). Nonparametric estimation of functions with jump discontinuities. In *Change-point Problems*, IMS–Lecture Notes, Eds. E. Carlstein, H.-G. Müller and D. Siegmund, **23**, 130–143.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, New York.
- 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.
- Grégoire, G. and Hamrouni, Z. (2002). Change-point estimation by local linear smoothing. *Journal of Multivariate Analysis*, **83**, 56–83.
- Hall, P. and Titterton D.M. (1992). Edge preserving and peak-preserving smoothing. *Technometrics*, **34**, 429–440.
- Hamrouni, Z. (1999). Inférence statistique par lissage linéaire local pour une fonction de régression présentant des discontinuités. *Doctoral Dissertation*, Université de Joseph Fourier, Grenoble, France.
- Kang, K.H., Koo, J.Y., and Park, C.W. (2000). Kernel estimation of discontinuous regression functions. *Statistics and Probability Letters*, **47**, 277–285.
- Koo, J.Y. (1997). Spline estimation of discontinuous regression functions. *Journal of Computational and Graphical Statistics*, **6**, 266–284.
- Lee, J.S. (1983). Digital image smoothing and the sigma filter. *Computer Vision, Graphics and image Processing*, **24**, 255–269.
- Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Series Analysis*, **17**, 571–599.
- Mc Donald J.A. and Owen A.B. (1986). Smoothing with linear fits. *Technometrics*, **28**, 195–208.
- Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics*, **20**, 737–761.
- Müller, H.-G. and Song, K.-S. (1997). Two-stage change-point estimators in smooth regression models. *Statistics & Probability Letters*, **34**, 323–335.
- Polzehl, J. and Spokoiny, V.G. (2000). Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society, Series B*, **2**, 335–354.
- Polzehl, J. and Spokoiny, V.G. (2003). Varying coefficient regression modeling by adaptive weights smoothing. *WIAS-Preprint 818*, Berlin.
- Qiu, P. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *J. Nonparametric Statist.*, **15**, 437–453.
- Qiu, P., and Yandell, B. (1998). A local polynomial jump detection algorithm in nonparametric regression. *Technometrics*, **40**, 141–152.

- Rue, H., Chu, C.-K., Godtlielsen, F. and Marron, J.S. (2002). M-smoother with local linear fit. *Journal of Nonparametric Statistics*, **14**, 155–168.
- Scott, D.W. (1992). *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley, New York.
- Speckman, P.L. (1994). Detection of change-points in nonparametric regression. *Unpublished manuscript*.
- Spokoiny, V.G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26**, 1356–1378.
- 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.

Irène Gijbels, University Centre of Statistics, University of Leuven, Belgium.

Email: irene.gijbels@wis.kuleuven.be

Alexandre Lambert, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Email: alambert@stat.ucl.ac.be

Peihua Qiu, School of Statistics, University of Minnesota, USA.

Email: qiu@stat.umn.edu