INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0401

JUMP-PRESERVING REGRESSION AND SMOOTHING USING LOCAL LINEAR FITTING: A COMPROMISE

I. GIJBELS, A. LAMBERT and P. QIU

http://www.stat.ucl.ac.be

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

Irène Gijbels[1]
Institut de Statistique
Université catholique de Louvain,

Alexandre Lambert[1]
Institut de Statistique
Université catholique de Louvain

Peihua Qiu[2]
School of Statistics
University of Minnesota

February 2, 2004

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 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] 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 (Belgian Science Policy). [2] It was also supported by a grant from National Security Agency of USA.

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 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 the 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 simultaneouly.

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.

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 Titterington (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 Titterington (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 simplicity, and by a data-driven choice of all parameters involved. Other relevant references are the papers 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 various 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 defered 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

$$Y_i = f(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where the $x_i = i/n$ are n equally-spaced design points in the design interval $[0,1]$, ε_i are n *iid* random errors with zero mean and finite variance σ^2, and f is a nonparametric regression function. In this paper, we consider the case when f has jumps at points s_j in $[0,1]$ with jump magnitudes d_j, for $j = 1, \ldots, m$. Without loss of generality, f is assumed to be right-continuous at all jump locations. The number of jumps m, the jump locations s_j's and the jump magnitudes d_j'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, which is confirmed by some numerical examples in Section 5. 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_ℓ 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 defined by:

$$\begin{aligned}
(\widehat{a}_{\ell,0}(x), \widehat{a}_{\ell,1}(x)) &= \arg\min_{a,b} \sum_{i=1}^{n} [Y_i - a - b(x_i - x)]^2 K_\ell\left(\frac{x_i - x}{h_n}\right), \\
(\widehat{a}_{r,0}(x), \widehat{a}_{r,1}(x)) &= \arg\min_{a,b} \sum_{i=1}^{n} [Y_i - a - b(x_i - x)]^2 K_r\left(\frac{x_i - x}{h_n}\right), \\
(\widehat{a}_{c,0}(x), \widehat{a}_{c,1}(x)) &= \arg\min_{a,b} \sum_{i=1}^{n} [Y_i - a - b(x_i - x)]^2 K\left(\frac{x_i - x}{h_n}\right), \qquad (2.2)
\end{aligned}$$

where $h_n > 0$ is a bandwidth parameter. Obviously the third set of estimators $(\widehat{a}_{c,0}(x), \widehat{a}_{c,1}(x))$ 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, which is referred to as the *conventional* local linear estimators below. 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 $[x, x + \frac{h_n}{2}]$. These two sets of estimates are called the *left* and the *right* local linear estimates, respectively.

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

$$\begin{aligned}
\mathrm{WRMS}_\ell(x) &= \frac{\sum_i [Y_i - \widehat{a}_{\ell,0} - \widehat{a}_{\ell,1}(x_i - x)]^2 K_\ell(\frac{x_i - x}{h})}{\sum_i K_\ell(\frac{x_i - x}{h})}, \\
\mathrm{WRMS}_r(x) &= \frac{\sum_i [Y_i - \widehat{a}_{r,0} - \widehat{a}_{r,1}(x_i - x)]^2 K_r(\frac{x_i - x}{h})}{\sum_i K_r(\frac{x_i - x}{h})}, \\
\mathrm{WRMS}_c(x) &= \frac{\sum_i [Y_i - \widehat{a}_{c,0} - \widehat{a}_{c,1}(x_i - x)]^2 K(\frac{x_i - x}{h})}{\sum_i K(\frac{x_i - x}{h})}. \qquad (2.3)
\end{aligned}$$

If f is continuous around a given point x, then all three estimators $\widehat{a}_{\ell,0}(x), \widehat{a}_{r,0}(x)$ and $\widehat{a}_{c,0}(x)$ 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 σ^2 in such cases.

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

$$\begin{aligned}
WRMS_\ell(x) &= \sigma^2 + R_{n,\ell,1}(x), \\
WRMS_r(x) &= \sigma^2 + R_{n,r,1}(x), \\
WRMS_c(x) &= \sigma^2 + R_{n,c,1}(x), \qquad (2.4)
\end{aligned}$$

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 $x \in [\frac{h_n}{2}, 1 - \frac{h_n}{2}] \setminus \bigcup_{j=1}^{m} \left[s_j - \frac{h_n}{2}, s_j + \frac{h_n}{2}\right]$.

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

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

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 $\mathrm{WRMS}_\ell(x)$ provides a consistent estimator for σ^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 $\mathrm{WRMS}_r(x)$ provides a consistent estimator for σ^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.

$$\begin{aligned}
C^2_{\tau,\ell} &= \Big(\frac{1}{v_{\ell,0}v_{\ell,2} - v^2_{\ell,1}}\Big)^2 \int_{-\tau}^{0} \left[\int_{-1/2}^{-\tau} (v_{\ell,2} - v_{\ell,1}x) K_\ell(x)\,dx + u \int_{-1/2}^{-\tau} (v_{\ell,0}x - v_{\ell,1}) K_\ell(x)\,dx\right]^2 K_\ell(u)\,du \\
&+ \Big(\frac{1}{v_{\ell,0}v_{\ell,2} - v^2_{\ell,1}}\Big)^2 \int_{-1/2}^{-\tau} \left[\int_{-\tau}^{0} (v_{\ell,2} - v_{\ell,1}x) K_\ell(x)\,dx - u \int_{-1/2}^{-\tau} (v_{\ell,0}x - v_{\ell,1}) K_\ell(x)\,dx\right]^2 K_\ell(u)\,du \\
C^2_{\tau,r} &= \Big(\frac{1}{v_{r,0}v_{r,2} - v^2_{r,1}}\Big)^2 \int_{0}^{-\tau} \left[\int_{-\tau}^{1/2} (v_{r,2} - v_{r,1}x) K_r(x)\,dx + u \int_{-\tau}^{1/2} (v_{r,0}x - v_{r,1}) K_r(x)\,dx\right]^2 K_r(u)\,du \\
&+ \Big(\frac{1}{v_{r,0}v_{r,2} - v^2_{r,1}}\Big)^2 \int_{-\tau}^{1/2} \left[\int_{0}^{-\tau} (v_{r,2} - v_{r,1}x) K_r(x)\,dx - u \int_{-\tau}^{1/2} (v_{r,0}x - v_{r,1}) K_r(x)\,dx\right]^2 K_r(u)\,du \\
C^2_{\tau,c} &= \Big(\frac{1}{v_{c,0}v_{c,2} - v^2_{c,1}}\Big)^2 \int_{-1/2}^{-\tau} \left[\int_{-\tau}^{1/2} (v_{c,2} - v_{c,1}x) K(x)\,dx + u \int_{-\tau}^{1/2} (v_{c,0}x - v_{c,1}) K(x)\,dx\right]^2 K(u)\,du \\
&+ \Big(\frac{1}{v_{c,0}v_{c,2} - v^2_{c,1}}\Big)^2 \int_{-\tau}^{1/2} \left[\int_{-1/2}^{-\tau} (v_{c,2} - v_{c,1}x) K(x)\,dx - u \int_{-\tau}^{1/2} (v_{c,0}x - v_{c,1}) K(x)\,dx\right]^2 K(u)\,du,
\end{aligned} \tag{2.6}$$

where

$$v_{\ell,j} = \int_{-1/2}^{0} u^j K(u)\,du, \quad v_{r,j} = \int_{0}^{1/2} u^j K(u)\,du, \quad v_{c,j} = \int_{-1/2}^{1/2} u^j K(u)\,du, \tag{2.7}$$

denote some moments of the kernel functions K_ℓ, K_r and K.

Proposition 2.2 *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; $h_n \to 0$ and $nh_n \to \infty$. Then, we have* *(i). for any $x = s + \tau h_n$ with $\tau \in [-\frac{1}{2}, 0)$:*

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

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 $\tau \in [0, \frac{1}{2}]$:

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

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 $\mathrm{WRMS}_\ell(x)$ and $\mathrm{WRMS}_r(x)$ in (2.8) and (2.9) were derived by Qiu (2003). The expressions for $\mathrm{WRMS}_c(x)$ 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 $C^2_{\tau,\ell}$, $C^2_{\tau,r}$ and $C^2_{\tau,c}$, we explore these quantities a bit further.

From their expressions, the quantities $C^2_{\tau,\ell}$, $C^2_{\tau,r}$ and $C^2_{\tau,c}$ depend on τ and the kernel function K. When K is one of the Uniform, Epanechnikov and Triangular kernels, which are presented in Figure 2.1, the three quantities are shown in Figure 2.2 as functions of τ. From Figure 2.2, we can discover some of their common properties: all three quantities seem to be bimodal functions of τ and this bimodality is stronger for kernels with smaller values at 0. Furthermore, the quantities $C^2_{\tau,\ell}$ and $C^2_{\tau,r}$ are symmetric versions of one another and $C^2_{\tau,c}$ is always symmetric about $\tau = 0$. Moreover, we have $C^2_{\tau,r} \le C^2_{\tau,c}$ for all values of τ. All these properties have been proved formally. Expressions for finding the local maxima of the three functions of τ are also available.

2.2 Estimation procedures

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

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

which is defined by one of the left and right estimates with smaller WRMS value. Qiu (2003) proved that $\widehat{f}_1(x)$ is a consistent estimator of $f(x)$ in the entire design interval. In practice it

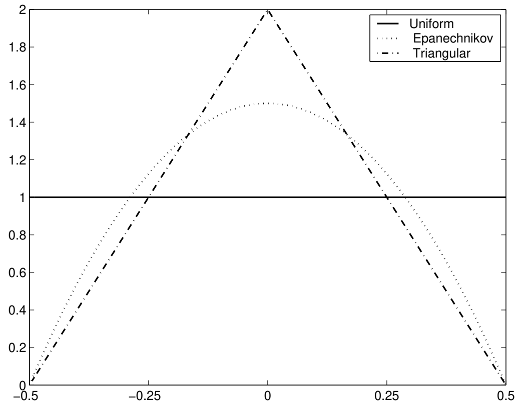
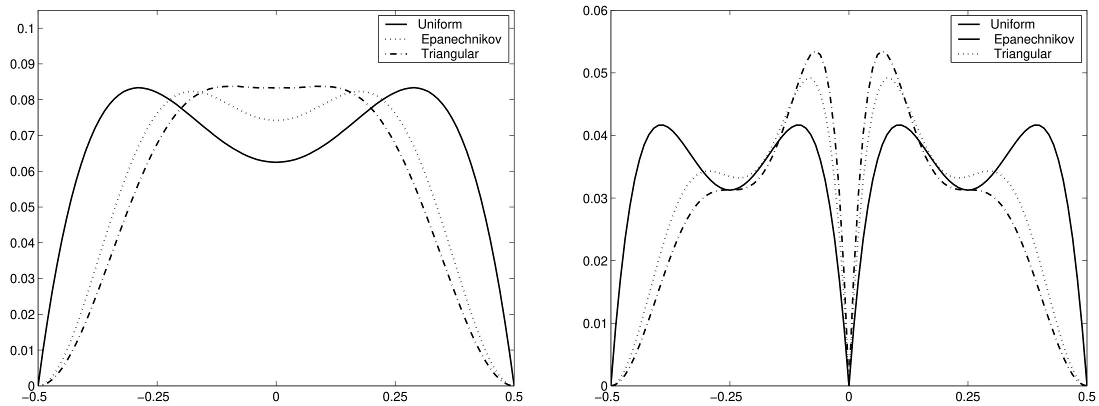


Figure 2.1: The Uniform, Epanechnikov and Triangular kernel functions.



(a) (b)

Figure 2.2: (a) Plots of $C^2_{\tau,c}$; (b) Plots of $C^2_{\tau,\ell}$ and $C^2_{\tau,r}$, for the three kernel functions shown in Figure 2.1.

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 $\widehat{a}_{\ell,0}(x)$, $\widehat{a}_{r,0}(x)$ and $\widehat{a}_{c,0}(x)$, 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 $[0,1]\setminus\bigcup_{j=1}^{m}\left[s_j-\frac{h_n}{2},\ s_j+\frac{h_n}{2}\right]$; the kernel function K is uniformly Lipschitz continuous; $h_n\to 0$ and $nh_n\to\infty$. Then* *at any* $x \in [\frac{h_n}{2}, 1-\frac{h_n}{2}] \setminus \bigcup_{j=1}^{m} \left[s_j - \frac{h_n}{2}, s_j + \frac{h_n}{2}\right]$*, we have*

$$
\begin{aligned}
MSE(\widehat{a}_{\ell,0}(x)) &= [\frac{h_n^2}{2} f''(x) B_{\ell,K}]^2 + \frac{\sigma^2}{nh_n} V_{\ell,K} + o(h^4 + \frac{1}{nh_n}),\\
MSE(\widehat{a}_{r,0}(x)) &= [\frac{h_n^2}{2} f''(x) B_{r,K}]^2 + \frac{\sigma^2}{nh_n} V_{r,K} + o(h^4 + \frac{1}{nh_n}),\\
MSE(\widehat{a}_{c,0}(x)) &= [\frac{h_n^2}{2} f''(x) B_{c,K}]^2 + \frac{\sigma^2}{nh_n} V_{c,K} + o(h^4 + \frac{1}{nh_n}),
\end{aligned}
\tag{2.11}
$$

where

$$
\begin{aligned}
B_{\ell,K} &= \frac{v_{\ell,2}^2 - v_{\ell,1}v_{\ell,3}}{v_{\ell,0}v_{\ell,2} - v_{\ell,1}^2}, & V_{\ell,K} &= \int_{-0.5}^{0} K^2(u)[\frac{v_{\ell,2} - v_{\ell,1}u}{v_{\ell,2}v_{\ell,0} - v_{\ell,1}^2}]^2\, du,\\
B_{r,K} &= \frac{v_{r,2}^2 - v_{r,1}v_{r,3}}{v_{r,0}v_{r,2} - v_{r,1}^2}, & V_{r,K} &= \int_{0}^{0.5} K^2(u)[\frac{v_{r,2} - v_{r,1}u}{v_{r,2}v_{r,0} - v_{r,1}^2}]^2\, du,\\
B_{c,K} &= \int_{-0.5}^{0.5} u^2 K(u)\, du = v_{c,2}, & V_{c,K} &= \int_{-0.5}^{0.5} K^2(u)\, du,
\end{aligned}
\tag{2.12}
$$

with $v_{\ell,j}$ *and* $v_{r,j}$ *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* $[s - h_n, s + \frac{h_n}{2}]$ *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;* $h_n \to 0$ *and* $nh_n \to \infty$*. Then, we have*

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

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

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

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

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 τh_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 ∞, the Mean Integrated Squared Error of the three estimates all tends to zero as n tends to ∞. See Hamrouni (1999) for details about this result.

As an illustration of these asymptotic expressions, we plot in Figure 2.3 the asymptotic bias functions, as function of x, 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 2.3 it can be seen that all three estimates have relatively low biases in continuity regions, and one of the two one-sided estimates has a considerably smaller bias than the other one near a jump point.

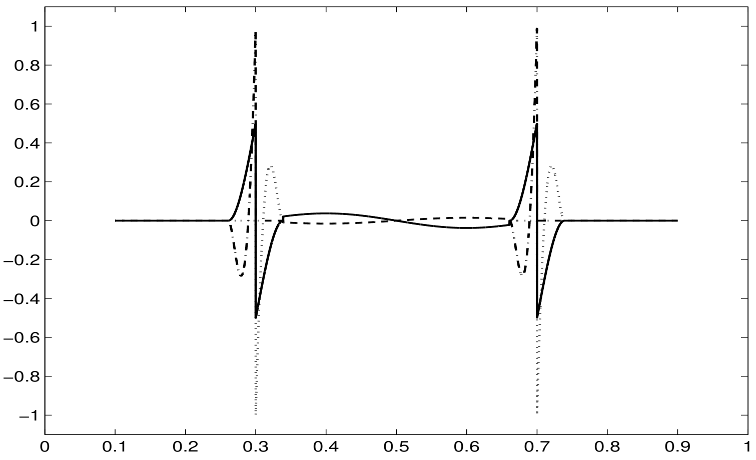


Figure 2.3: Asymptotic bias functions of $\widehat{a}_{c,0}(x)$ (solid curve), $\widehat{a}_{\ell,0}(x)$ (dotted curve) and $\widehat{a}_{r,0}(x)$ (dashed-dotted curve) for the regression function $f = g_1$ defined in (5.1), using $h_n = 0.078$.

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 several curve estimation procedures which try to balance jump-preservation and local smoothing.

As mentioned above, f_1 is quite noisy in continuity regions of f. To overcome this limitation, the first natural improvement is to define the following estimate, which tries to use the conventional local linear estimate in continuity regions.

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

However it can be checked that $\mathrm{WRMS}_c(x) > \min(\mathrm{WRMS}_\ell(x), \mathrm{WRMS}_r(x))$ 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.

To overcome this problem, we propose two different modifications. By the first modification, $\mathrm{WRMS}_c(x)$ is replaced by $\frac{\mathrm{WRMS}_{c(x)}}{C}$ in (2.15) where $C > 1$ is a constant, which gives the following estimate:

$$\widehat{f}_3(x) = \begin{cases} \widehat{a}_{c,0}(x) & \text{if} \quad \frac{\mathrm{WRMS}_{c(x)}}{C} \le \min(\mathrm{WRMS}_\ell(x), \mathrm{WRMS}_r(x)) \\ \widehat{a}_{\ell,0}(x) & \text{if} \quad \frac{\mathrm{WRMS}_{c(x)}}{C} > \min(\mathrm{WRMS}_\ell(x), \mathrm{WRMS}_r(x)) = \mathrm{WRMS}_\ell(x) \\ \widehat{a}_{r,0}(x) & \text{if} \quad \frac{\mathrm{WRMS}_{c(x)}}{C} > \min(\mathrm{WRMS}_\ell(x), \mathrm{WRMS}_r(x)) = \mathrm{WRMS}_r(x) \\ (\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x))/2 & \text{if} \quad \frac{\mathrm{WRMS}_{c(x)}}{C} > \mathrm{WRMS}_\ell(x) = \mathrm{WRMS}_r(x). \end{cases} \tag{2.16}$$

The constant C should be chosen such that the conventional estimate is selected in continuity regions of f by (2.16) (i.e., $\frac{\mathrm{WRMS}_{c(x)}}{C} \le \min(\mathrm{WRMS}_\ell(x), \mathrm{WRMS}_r(x))$ in such cases), and one of the two one-sided estimates is selected near a jump point (i.e., $\frac{\mathrm{WRMS}_{c(x)}}{C} > \min(\mathrm{WRMS}_\ell(x), \mathrm{WRMS}_r(x))$). By Propositions 2.1 and 2.2, these two requirements are asymptotically equivalent to

$$1 < C < 1 + \frac{d^2}{\sigma^2} C_{\tau,c}^2, \tag{2.17}$$

which depends on the signal-to-noise ratio d/σ.

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:

$$\widehat{f}_4(x) = \begin{cases} \widehat{a}_{c,0}(x) & \text{if} \quad \mathrm{diff}(x) \le u \\ \widehat{a}_{\ell,0}(x) & \text{if} \quad \mathrm{diff}(x) > u \text{ and } \mathrm{WRMS}_\ell(x) < \mathrm{WRMS}_r(x) \\ \widehat{a}_{r,0}(x) & \text{if} \quad \mathrm{diff}(x) > u \text{ and } \mathrm{WRMS}_\ell(x) > \mathrm{WRMS}_r(x) \\ (\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x))/2 & \text{if} \quad \mathrm{diff}(x) > u \text{ and } \mathrm{WRMS}_\ell(x) = \mathrm{WRMS}_r(x), \end{cases} \tag{2.18}$$

where u is a threshold value and

$$\mathrm{diff}(x) = \max(\mathrm{WRMS}_c(x) - \mathrm{WRMS}_\ell(x), \mathrm{WRMS}_c(x) - \mathrm{WRMS}_r(x)). \tag{2.19}$$

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

$$0 < u < d^2C^2_{\tau,c}. \tag{2.20}$$

In contrast to the (asymptotic) constraint on C (cf. (2.17)), we can see that the (asymptotic) constraint on u does not have the error variance σ^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 (2.17) and (2.20) should be replaced by

$$1 < C < 1 + \frac{d_j^2}{\sigma^2}C^2_{\tau_j,c} \text{ and } 0 < u < d_j^2 C^2_{\tau_j,c},$$

for $j = 1, 2, \ldots, m$, where τ_j is defined by $\tau_j = \frac{x-s_j}{h_n}$.

The two modified estimates described above overcome the limitation of f_1 that the conventional estimate is not selected in continuity regions. But they might suffer another slight drawback, although it is secondary compared to the drawback of f_1, as shown by the simulation study in Section 5. Consider for example the second modified estimate $\widehat{f}_4(\cdot)$ with a threshold value u. Suppose that x is a point close to a jump point s, i.e., $x = s + \tau h$ with $\tau \in [-1/2, 1/2]$. For simplicity of notation, we dropped the dependence of h_n on n here. Then we can find $0 < \tau(u) < 1/2$ such that $d^2C^2_{\tau(u),c} = u$. Then in regions $\{s + \tilde{\tau}h : \tilde{\tau} \in [-1/2, -\tau(u)] \cup [\tau(u), 1/2]\}$, the conventional estimate will be selected as $\widehat{f}_4$, because asymptotically $d^2C^2_{\tilde{\tau},c} \le u$ in such cases, as demonstrated by Figure 2.4 (a).

To get ride of this drawback, we propose to compute another conventional estimate with a somewhat bigger bandwidth $h_c = k \times h$ for $k \ge 1$, besides the three estimates used in (2.18). This conventional estimate and its WRMS value are denoted by $\widehat{a}_{c,0,h_c}$ and WRMS_{c,h_c}, respectively. The rationale of this proposal is demonstrated by Figure 2.4 (b). From the plot it can be seen that when $s + \tau(u)h_c \ge s + h/2$, if the final curve estimate is defined by $\widehat{a}_{r,0}$ in $[s, s + \tau(u)h_c)$ and by $\widehat{a}_{c,0}$ in $[s + \tau(u)h_c, s + h_c/2]$, then it is consistent in the whole right-sided neighborhood, although it is a little less efficient than $\widehat{a}_{c,0}$ in $[s + h/2, s + \tau(u)h_c]$. It can be similarly defined in the left-sided neighborhood. The resulting estimate is

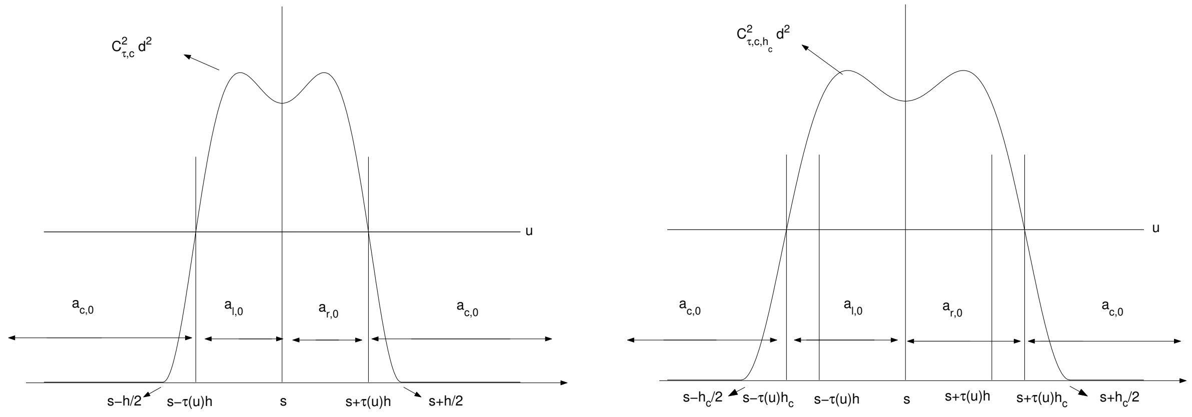
$$\widehat{f}_5(x) = \begin{cases} \widehat{a}_{c,0}(x) & \text{if} \quad \mathrm{diff}(x) \le u \\ \widehat{a}_{\ell,0}(x) & \text{if} \quad \mathrm{diff}(x) > u \text{ and } \mathrm{WRMS}_\ell(x) < \mathrm{WRMS}_r(x) \\ \widehat{a}_{r,0}(x) & \text{if} \quad \mathrm{diff}(x) > u \text{ and } \mathrm{WRMS}_\ell(x) > \mathrm{WRMS}_r(x) \\ (\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x))/2 & \text{if} \quad \mathrm{diff}(x) > u \text{ and } \mathrm{WRMS}_\ell(x) = \mathrm{WRMS}_r(x), \end{cases} \tag{2.21}$$

where the enlarged bandwidth h_c and the regular bandwidth h should be such that $s + \tau(u)h_c \ge s + h/2$, or equivalently,

$$h_c \ge \frac{h}{2\tau(u)} \quad \Longleftrightarrow \quad k \ge \frac{1}{2\tau(u)}, \tag{2.22}$$

and

$$\mathrm{diff}(x) = \max(\mathrm{WRMS}_{c,h_c}(x) - \mathrm{WRMS}_\ell(x), \mathrm{WRMS}_{c,h_c}(x) - \mathrm{WRMS}_r(x)). \tag{2.23}$$



(a) (b)

Figure 2.4: (a): Asymptotic behaviour of diff(x) defined in (2.19) near a jump point s, and the different estimates selected by $\widehat{f}_4$; (b): Asymptotic behaviour of diff(x) defined in (2.23) near a jumps point s, and the different estimates selected by $\widehat{f}_5$ (with $k = 1/(2\tau(u))$).

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

3 Consistency

From our extensive simulation study in Section 5, it appears that $\widehat{f}_4$ has some preferable properties compared to the other estimates $\widehat{f}_1$, $\widehat{f}_3$ and $\widehat{f}_5$. 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 $\widehat{f}_4$ 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 $h_n \to 0$, $nh_n^3 \to \infty$ and $\sqrt{\frac{nh_n^5}{\ln n}} \to 0$ as n goes to ∞, 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 $\widehat{f}_4$ is uniformly strong consistent on the interval $[0, 1]$ excluding all neighbourhoods of jumps 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:

$$
\begin{aligned}
D_1 &= [\rho, 1-\rho] \backslash \bigcup_{j=1}^{m} [s_j - h_n/2, s_j + h_n/2] \\
D_2 &= \bigcup_{j=1}^{m} [s_j - h_n/2, s_j + h_n/2] \\
D_{2,\delta} &= \bigcup_{j=1}^{m} \left\{ [s_j - (1/2-\delta)h_n, s_j - \delta h_n] \bigcup [s_j + \delta h_n, s_j + (1/2-\delta)h_n] \right\}, \qquad (3.1)
\end{aligned}
$$

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 $\{s_j, j = 1, \ldots, m\}$ 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 $h_n \to 0$, $nh_n^3 \to \infty$ and $\sqrt{\frac{nh_n^5}{\ln n}} \to 0$, and the threshold $u = u_n \to 0$ as n goes to ∞. Then we have:*

(i)

$$\sup_{x \in D_1} \sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_4(x) - f(x)| = O(1), a.s.;$$

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

$$\sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_4(x) - f(x)| = O(1), a.s.;$$

(iii)

$$\sup_{x \in D_{2,\delta}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_4(x) - f(x)| = O(1), a.s.,$$

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 three estimates $\widehat{f}_1, \widehat{f}_3$ and $\widehat{f}_5$ 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, $\mathrm{MISE} = E\left[\int_0^1 (\widehat{f}(x) - f(x))^2\, f_X(x)\, dx\right]$, 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:

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

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

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

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 $\widehat{f}_3$, $\widehat{f}_4$ and $\widehat{f}_5$ can be selected, together with the bandwidth h_n, by the cross-validation procedure:

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

where $\widehat{f}^{-i}$ is one of $\widehat{f}_3^{-i}$, $\widehat{f}_4^{-i}$ and $\widehat{f}_5^{-i}$.

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 (2.17) needs to be imposed on C_n, which has two unknown quantities $|d|$ and σ^2 involved. The quantity $|d|$ can be estimated by $\max_{x \in [\rho, 1-\rho]} |\widehat{a}_{\ell,0}(x) - \widehat{a}_{r,0}(x)|$ where $\rho > 0$ is a small constant, and σ^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 (2.17). We know that when $C = 1$, $\widehat{f}_3(x) = \widehat{f}_2(x)$ and both estimates are close to $\widehat{f}_1(x)$ in such a case, which is good in preserving the jumps. If C is large, then $\widehat{f}_3(x)$ 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 (2.20) does not have σ^2 involved.

4.2.3 Factor k

For $\widehat{f}_5$, the cross-validation procedure can be used for selecting h_n, u_n and h_c (or k, since $h_c = k \times h$) simultaneouly. We discuss three possible scenarios here.

Suppose for a moment that the jump size d is known. Then the theoretical condition (2.22) requires that

$$h_c \geq \frac{h_n}{2\tau(u_n)}, \tag{4.4}$$

where $\tau(u_n)$ is chosen such that $C^2_{\tau(u_n),c} = u_n/d^2$. Given the bimodality of $C^2_{\tau,c}$ (see Figure 2.2), the quantity $\tau(u_n)$ is uniquely defined only if $u_n < d^2 C^2_{0,c}$, where $C_{0,c}$ denotes the value of $C_{\tau,c}$ when $\tau = 0$. For such a u_n, we take h_c to be the lower bound of (4.4). The corresponding curve estimate is denoted by $\widehat{f}_{5,A}$.

Next let us consider the more realistic situation that the jump size $|d|$ is unknown. One can proceed similarly as above, by replacing $|d|$ with a preliminary estimator. For given h_n, u_n and at a given point x, $|d|$ can be estimated by

$$\widehat{|d|}(x) = \sup_{z \in [x \pm h_n/2]} |\widehat{a}_{r,0}(z) - \widehat{a}_{\ell,0}(z)|.$$

Substituting $\widehat{|d|}(x)$ into $C^2_{\tau(u_n),c} = u_n/d^2$ leads to $\widehat{\tau}(u_n, x)$, and this on its turn leads to $\widehat{h}_c(x) = \frac{h_n}{2\widehat{\tau}(u_n,x)}$. If x is in a continuity region of f, then $\widehat{|d|}(x)$ would be close to zero, $u_n/\widehat{d}^2(x)$ would be large and consequently $\widehat{\tau}(u_n, x)$ is not well defined. In such a case, we let $k = 1$ and hence the bandwidth h_c is the same as the bandwidth h, which is reasonable because the extra bandwidth h_c is not needed in continuity regions. If x is near a jump point, then $\widehat{|d|}(x)$ is close to the true jump size $|d|$ and the resulting $\widehat{h}_c$ would be close to the theoretical h_c. We denote the corresponding curve estimate by $\widehat{f}_{5,B}$.

Finally, $k = k_n$ can be selected by the following cross-validation procedure, along with h_n and u_n:

$$(\widehat{h}_n, \widehat{u}_n, \widehat{k}_n) = \arg \min_{h_n, u_n, k_n} \sum_{i=1}^{n} [Y_i - \widehat{f}_5^{-i}(x_i)]^2,$$

where it often suffices to search for optimal k_n in $[1, 2]$. The corresponding curve estimate is denoted by $\widehat{f}_{5,C}$. This last procedure for choosing k_n is computationally expensive, and hence should only be used for small to moderate sample sizes. Table 4.1 summarizes the parameter selection procedures discussed above.

Table 4.1: Proposed parameter selection methods for various curve estimates.

method	parameters	selection procedure
$\widehat{a}_{c,0}$	h_n	CV(h_n)
$\widehat{f}_1$	h_n	CV(h_n)
$\widehat{f}_3$	h_n, C_n	CV(h_n, C_n)
$\widehat{f}_4$	h_n, u_n	CV(h_n, u_n)
$\widehat{f}_{5,A}$	h_n, u_n, k	CV(h_n, u_n) + asymptotic k, when d is known
$\widehat{f}_{5,B}$	h_n, u_n, k	CV(h_n, u_n) + data-driven $k(x)$, when d is unknown
$\widehat{f}_{5,C}$	h_n, u_n, k	CV(h_n, u_n, k)

5 Simulation Study

In this section we investigate the performance of the curve estimation procedures discussed in Section 2.2 by a simulation study. All numerical examples are organized in two subsections. Most of them are for the case of fixed design and are presented in Section 5.1. Examples in cases when the design points are random or when the errors are heteroscedastic are given in Section 5.2.

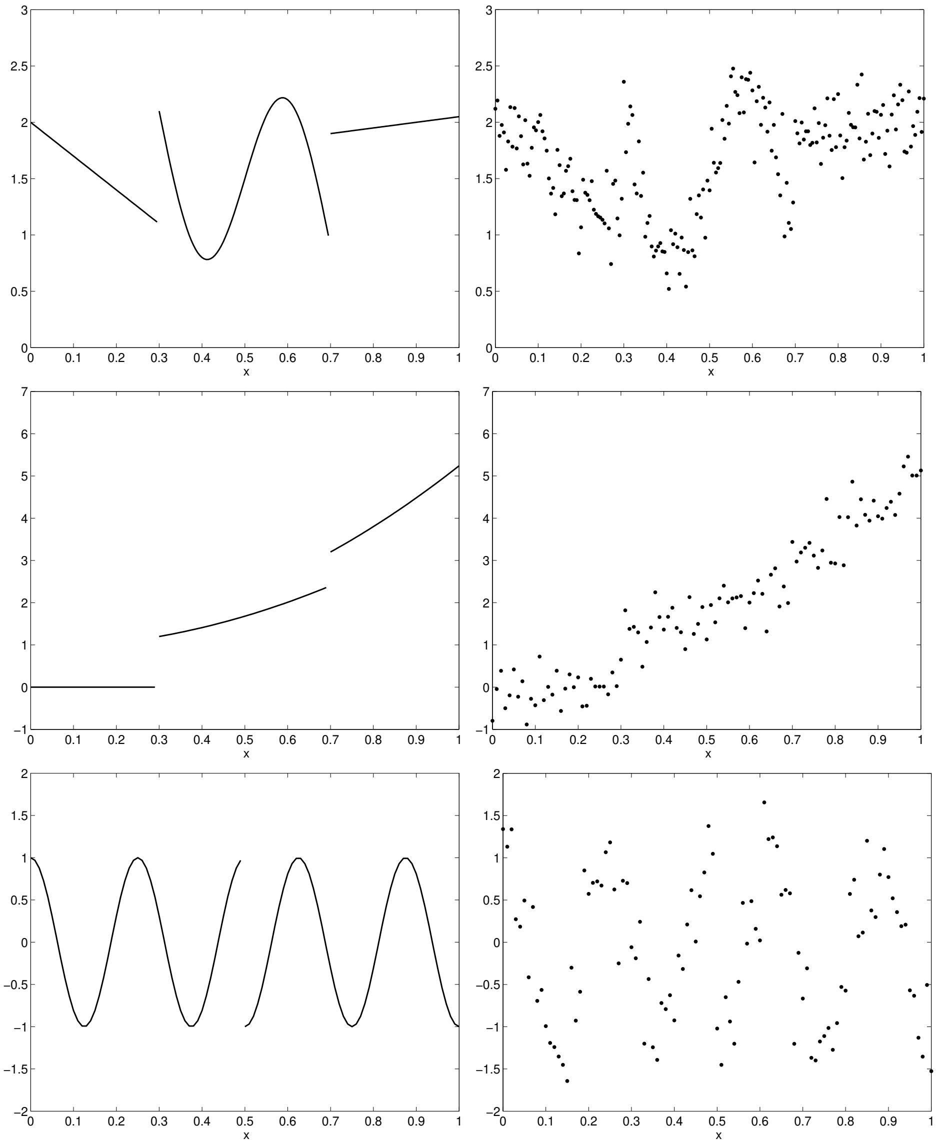


Figure 5.1: Left panel: graphs of the functions g_1, g_2 and g_3. Right panel: simulated data from model (2.1) with $f = g_1$ ($n = 200, \varepsilon \sim N(0; 0.2^2)$); $f = g_2$ ($n = 100, \varepsilon \sim N(0; 0.4^2)$), and $f = g_3$ ($n = 100, \varepsilon \sim N(0; 0.4^2)$).

5.1 Fixed design case

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

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

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 random versions of the three regression functions with n and σ values as specified in the figure caption.

For a given regression function f, $N = 200$ samples are then generated from model (2.1) with $\varepsilon_1 \sim N(0; \sigma^2)$. The MISE value of a curve estimate $\widehat{f}$ is estimated by

$$\widehat{\mathrm{MISE}} = \frac{1}{N} \sum_{j=1}^{N} \widehat{\mathrm{ISE}}_j, \tag{5.2}$$

where

$$\widehat{\mathrm{ISE}}_j = \sum_{i=1}^{n-1} \frac{\mathrm{SE}_j(x_i) + \mathrm{SE}_j(x_{i+1})}{2n},$$

$\mathrm{SE}_j(x_i) = (f(x_i) - \widehat{f}^j(x_i))^2$ and $\widehat{f}^j$ denotes the curve estimate $\widehat{f}$ constructed from the j-th simulated sample.

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

$$\widehat{\mathrm{MISE}}_s = \frac{1}{N} \sum_{j=1}^{N} \widehat{\mathrm{ISE}}_{s,j}, \tag{5.3}$$

where

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

When $f = g_1$, 6 different values of σ^2 are considered, with a fixed sample size $n = 200$. The performance of the curve estimation procedures is evaluated by their $\widehat{\mathrm{MISE}}$ and $\widehat{\mathrm{MISE}}_{s=0.3} + \widehat{\mathrm{MISE}}_{s=0.7}$ 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 $\widehat{a}_{c,0}$ has decent $\widehat{\mathrm{MISE}}$ 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 $\widehat{f}_1$ has opposite behaviour, namely, it has relatively small $\widehat{\mathrm{MISE}}_{s=0.3} + \widehat{\mathrm{MISE}}_{s=0.7}$ values (i.e., good jump-preserving), but relatively large $\widehat{\mathrm{MISE}}$ 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. It is further noticed that the estimates $\widehat{f}_{5,A}$, $\widehat{f}_{5,B}$ and $\widehat{f}_{5,C}$ perform similarly to $\widehat{f}_4$. So it seems that the additional improvement which we tried to incorporate into $\widehat{f}_4$ does not have a big impact on the performance of the

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	σ=0.1		σ=0.15		σ=0.2		σ=0.25		σ= 0.3		σ= 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
$\widehat{f}_3$	0.0025	0.0011	0.0065	0.0040	0.0108	0.0070	0.0146	0.0093	0.0180	0.0110	0.0243	0.0143
$\widehat{f}_4$	0.0027	0.0014	0.0055	0.0030	0.0088	0.0048	0.0127	0.0068	0.0168	0.0090	0.0248	0.0126
$\widehat{f}_{5,A}$	0.0030	0.0018	0.0058	0.0031	0.0091	0.0049	0.0131	0.0070	0.0169	0.0091	0.0248	0.0126
$\widehat{f}_{5,B}$	0.0032	0.0019	0.0060	0.0034	0.0091	0.0051	0.0129	0.0071	0.0169	0.0091	0.0253	0.0126
$\widehat{f}_{5,C}$	0.0027	0.0015	0.0054	0.0029	0.0087	0.0047	0.0127	0.0068	0.0168	0.0089	0.0253	0.0127

estimate. In other words, the correction by introducing a larger bandwidth h_c in $\widehat{f}_5$ does not seem to be worth the effort. Further, we can see that for all estimates $\widehat{\text{MISE}}$ increases with σ^2. For a large value of σ^2 all estimates (except $\widehat{f}_1$) show similar results. This means that for such a large value of σ^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 $\widehat{f}_4$ behaves better than $\widehat{f}_3$ for median values of σ. This is no surprise since from their definitions provided in Section 2.2, it is already clear that σ^2 is less involved in $\widehat{f}_4$ (also see the constraints (2.17) and (2.20)).

Figure 5.2 gives a graphical display of the performance of the four estimates $\widehat{a}_{c,0}$, $\widehat{f}_1$, $\widehat{f}_3$ and $\widehat{f}_4$ when $f = g_1$. In each plot, the true regression function g_1 is represented by the solid curve. The dotted curves represent the 5^{th} percentile, the average and the 95^{th} percentile of $N = 200$ replicated fits. The curve of the averaged fit can be seen as an estimator of $\text{Bias}(\widehat{f}(x))$, and the difference between the upper and lower dotted curves can be regarded as a measure of $\text{Var}(\widehat{f}(x))$. From the plots, it can be seen that the conventional estimator $\widehat{a}_{c,0}$ blurs the two jumps, but its variance is quite small. The estimate $\widehat{f}_1$ preserves the jumps well, but its variance is quite large in continuity regions. For $\widehat{f}_3$ and $\widehat{f}_4$, their variances are close to that of $\widehat{a}_{c,0}$, and they preserve the jumps much better than the conventional estimate. 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.

When $f = g_2$, we consider three different sample sizes $n = 100, 200, 300$ and one error variance value $\sigma^2 = 0.16$. The global and local MISE values of various estimates are presented in Table 5.2 (first, second and third columns respectively). We can see that for all estimates $\widehat{\text{MISE}}$ decreases as n increases. Furthermore, the estimate $\widehat{f}_1$ 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 σ^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 $\widehat{a}_{c,0}$ has the largest $\widehat{\text{MISE}}_{s=0.3}$ values (due to jump blurring) and $\widehat{f}_1$ has the smallest $\widehat{\text{MISE}}_{s=0.3}$ values (i.e., good jump preserving). The other estimates behave in between

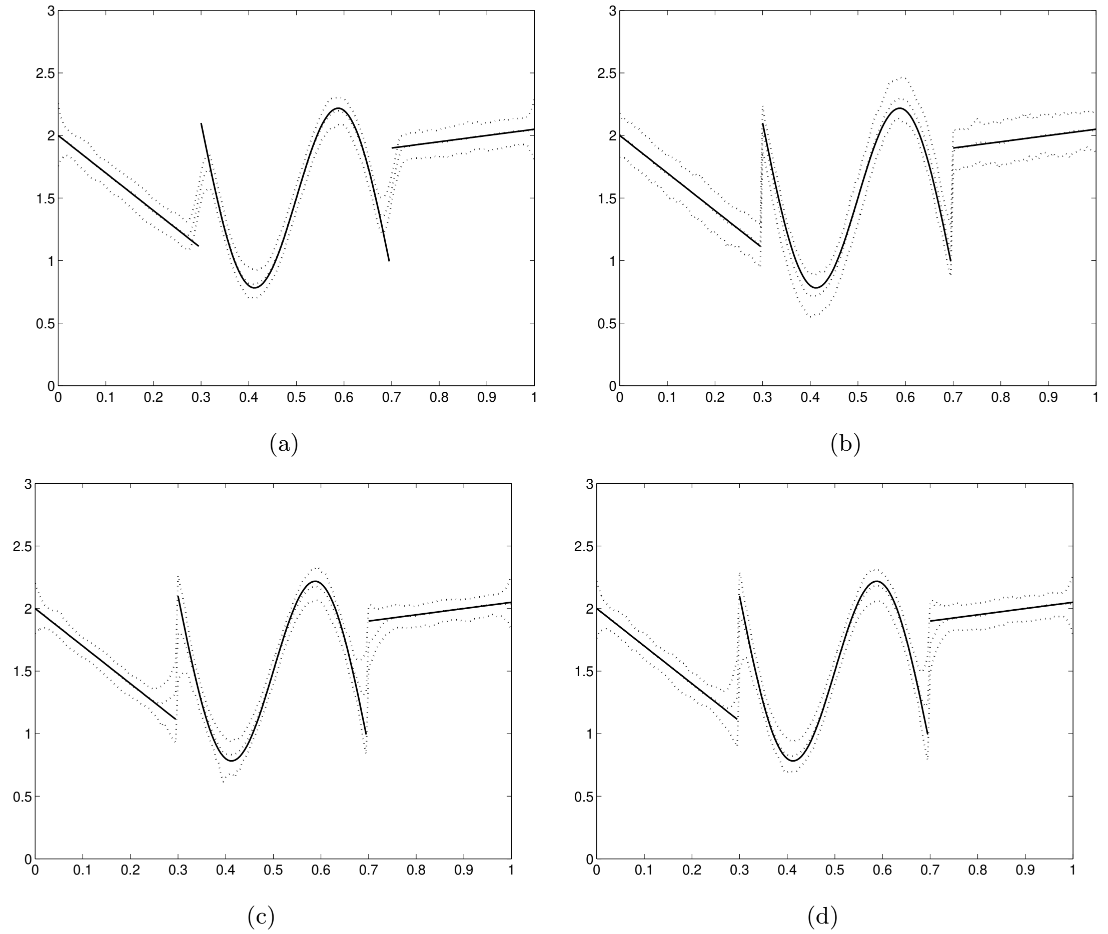


Figure 5.2: Plots of four curve estimates when $f = g_1$, $n = 200$, $\sigma = 0.2$ and $N = 200$. In each plot, the true regression function is denoted by the solid curve. The 5^{th} percentile, the average and the 95^{th} percentile of N replicated fits are denoted by dotted curves. Estimates: (a) $\widehat{a}_{c,0}$; (b) $\widehat{f}_1$; (c) $\widehat{f}_3$; and (d) $\widehat{f}_4$.

Table 5.2: $\widehat{\text{MISE}}$, $\widehat{\text{MISE}}_{s=0.3}$ and $\widehat{\text{MISE}}_{s=0.7}$ values when $f = g_2$, $N = 200$ and $\varepsilon_1 \sim N(0; 0.16)$.

Method	$n = 100$			$n = 200$			$n = 300$		
	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_{s=0.3}$	$\widehat{\text{MISE}}_{s=0.7}$	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_{s=0.3}$	$\widehat{\text{MISE}}_{s=0.7}$	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_{s=0.3}$	$\widehat{\text{MISE}}_{s=0.7}$
$\widehat{a}_{c,0}$	0.0363	0.0129	0.0064	0.0242	0.0091	0.0047	0.0194	0.0075	0.0037
$\widehat{f}_1$	0.0529	0.0083	0.0072	0.0311	0.0054	0.0052	0.0234	0.0042	0.0047
$\widehat{f}_3$	0.0365	0.0133	0.0074	0.0235	0.0090	0.0055	0.0178	0.0058	0.0045
$\widehat{f}_4$	0.0362	0.0110	0.0070	0.0235	0.0076	0.0054	0.0176	0.0051	0.0048
$\widehat{f}_{5,B}$	0.0360	0.0113	0.0068	0.0240	0.0076	0.0055	0.0184	0.0061	0.0046
$\widehat{f}_{5,C}$	0.0369	0.0110	0.0069	0.0239	0.0078	0.0056	0.0180	0.0057	0.0049

them. Around the second jump point, it can be seen that $\widehat{a}_{c,0}$ performs the best and all other estimates have similar performance.

Figure 5.3 depicts the performance of the four estimates $\widehat{a}_{c,0}$, $\widehat{f}_1$, $\widehat{f}_3$ and $\widehat{f}_4$ when $f = g_2$.

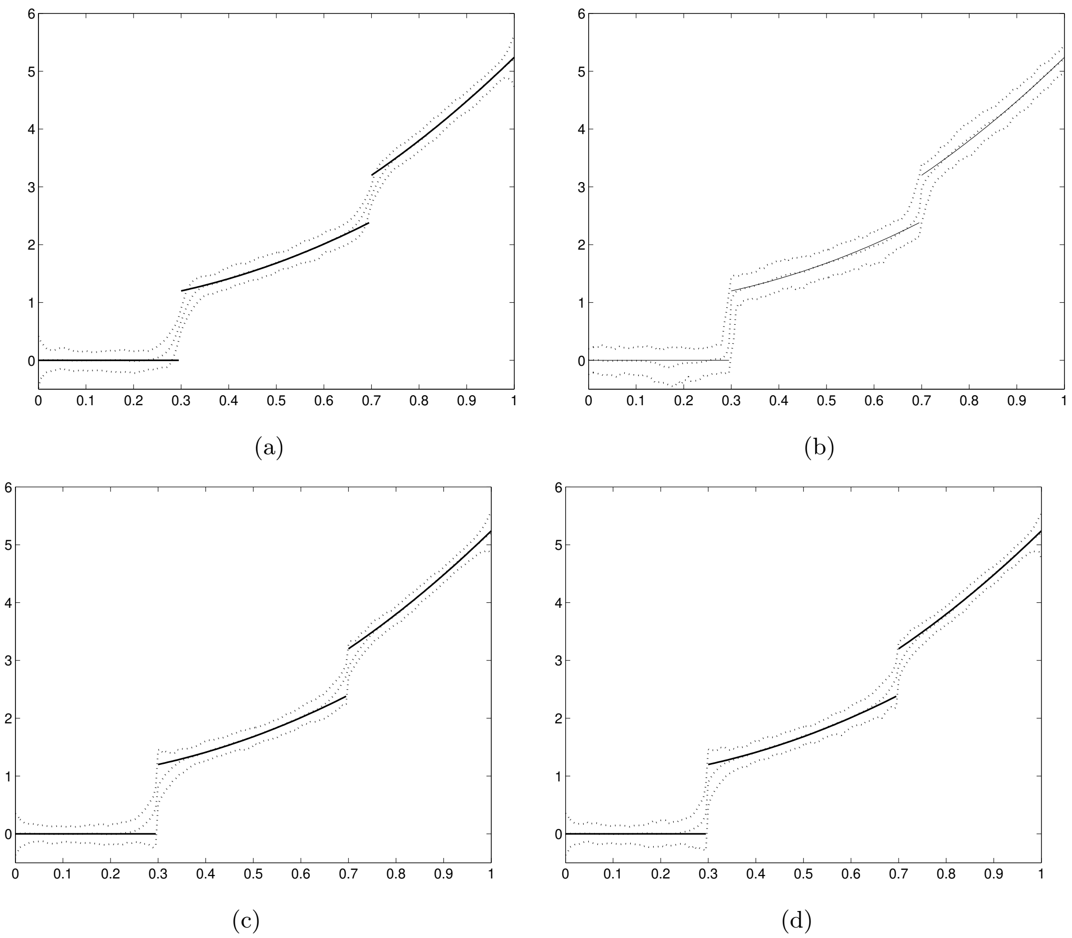


Figure 5.3: Plots of four curve estimates when $f = g_2$, $n = 200$, $\sigma = 0.4$ and $N = 200$. In each plot, the true regression function is denoted by the solid curve. The 5^{th} percentile, the average and the 95^{th} percentile of N replicated fits are denoted by dotted curves. Estimates: (a) $\widehat{a}_{c,0}$; (b) $\widehat{f}_1$; (c) $\widehat{f}_3$; and (d) $\widehat{f}_4$.

As before the true regression function is denoted by the solid curve in each plot. The dotted curves represent the 5^{th} percentile, the average and the 95^{th} percentile of $N = 200$ replicated estimates. Similar conclusions to those from Figure 5.2 can be drawn here, except that the first jump seems to be better preserved by $\widehat{f}_1, \widehat{f}_3$ and $\widehat{f}_4$.

When $f = g_3$, we also use three different sample sizes $n = 100, 200, 300$ and 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 $\widehat{f}_4$ performs considerably better than $\widehat{f}_3$ in this case.

Figure 5.4 depicts the true regression function g_3 and its four estimates $\widehat{a}_{c,0}$, $\widehat{f}_1$, $\widehat{f}_3$ and $\widehat{f}_4$.

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
$\widehat{f}_3$	0.0743	0.0226	0.0355	0.0155	0.0252	0.0115
$\widehat{f}_4$	0.0484	0.0168	0.0277	0.0094	0.0210	0.0073
$\widehat{f}_{5,B}$	0.0497	0.0181	0.0287	0.0102	0.0217	0.0080
$\widehat{f}_{5,C}$	0.0487	0.0166	0.0276	0.0093	0.0212	0.0073

Similar conclusions to those from Figures 5.2 and 5.3 can be drawn here.

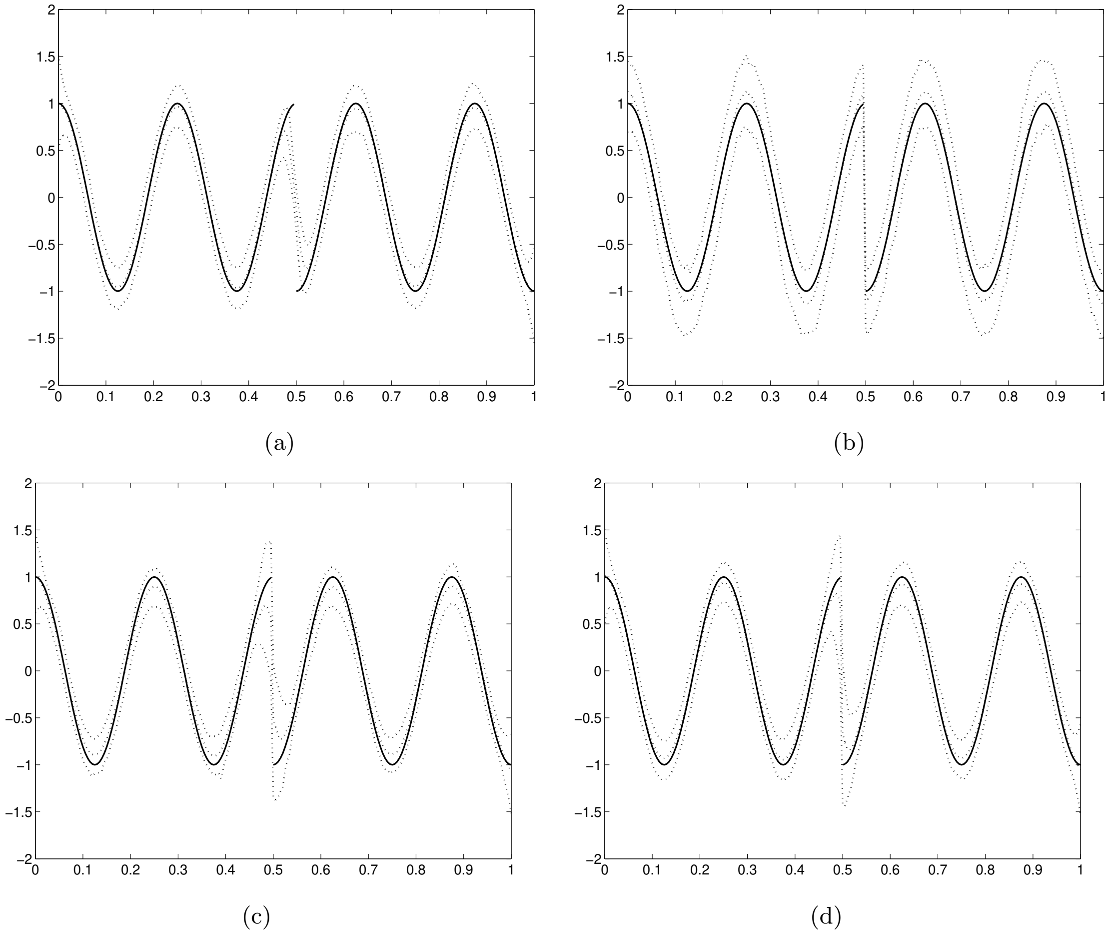


Figure 5.4: Plots of four curve estimates when $f = g_3$, $n = 200$, $\sigma = 0.4$ and $N = 200$. In each plot, the true regression function is denoted by the solid curve. The 5^{th} percentile, the average and the 95^{th} percentile of N replicated fits are denoted by dotted curves. Estimates: (a) $\widehat{a}_{c,0}$; (b) $\widehat{f}_1$; (c) $\widehat{f}_3$; and (d) $\widehat{f}_4$.

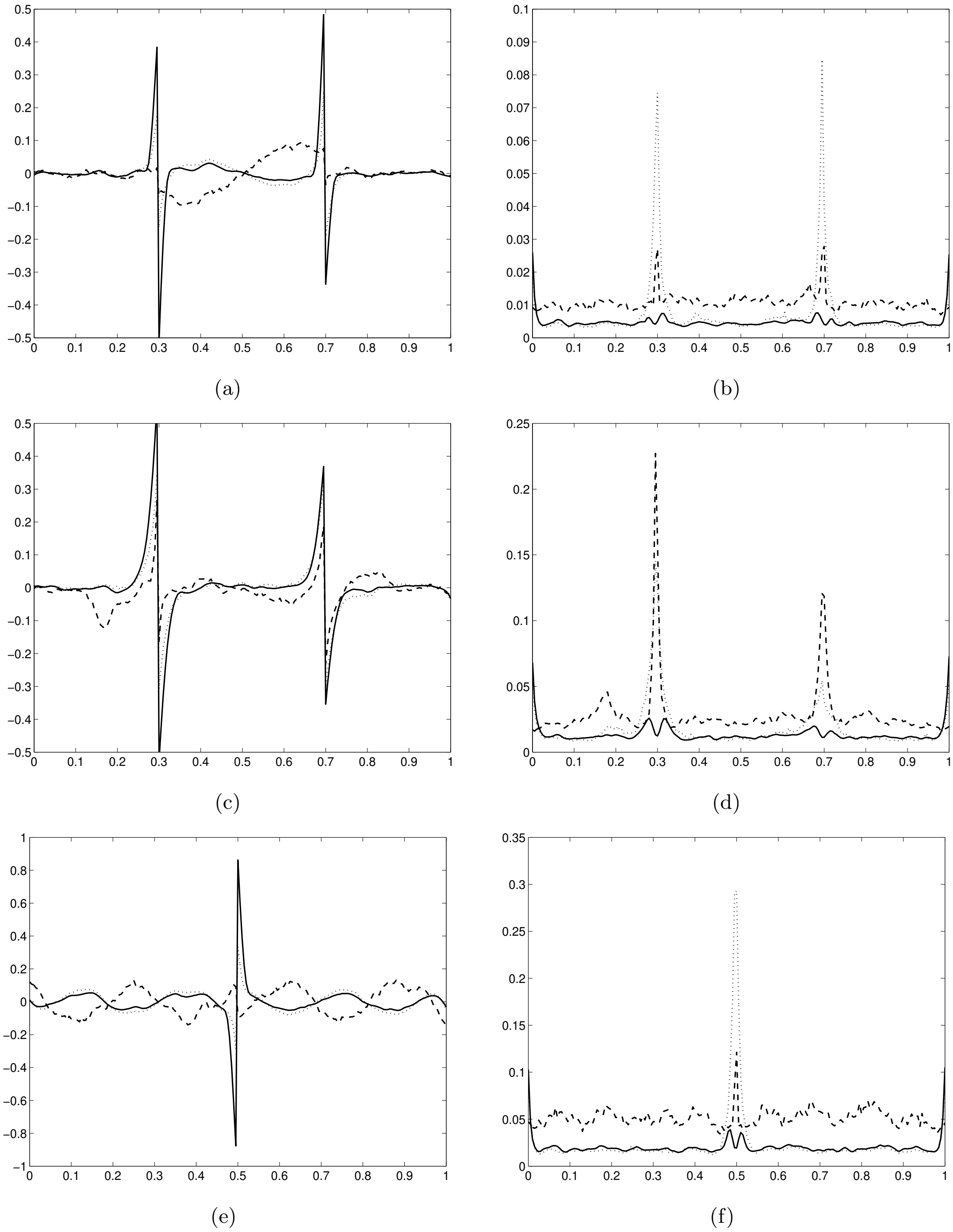


Figure 5.5: Plots of the estimated bias functions (left panels) and variance functions (right panels) of the estimates $\widehat{a}_{c,0}$ (solid curves), $\widehat{f}_1$ (dashed curves) and $\widehat{f}_4$ (dotted curves). (a) and (b): $f = g_1$, $n = 200$ and $\sigma = 0.2$; (c) and (d): $f = g_2$, $n = 200$ and $\sigma = 0.4$; and (e) and (f): $f = g_3$, $n = 200$ and $\sigma = 0.4$. In all cases: $N = 200$.

Figure 5.5 presents the estimated bias (left panels) and variance (right panels) functions of $\widehat{a}_{c,0}, \widehat{f}_1$ and $\widehat{f}_4$ when f equals g_1, g_2 and g_3, respectively. 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 2.3, 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 $\widehat{f}_4$ and $\widehat{a}_{c,0}$ are similar and small, due to the fact that $\widehat{f}_4$ 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 $\widehat{f}_1$ and $\widehat{f}_4$ are bigger around the jump points because they have to choose among three estimates at such places, which increases their variances.

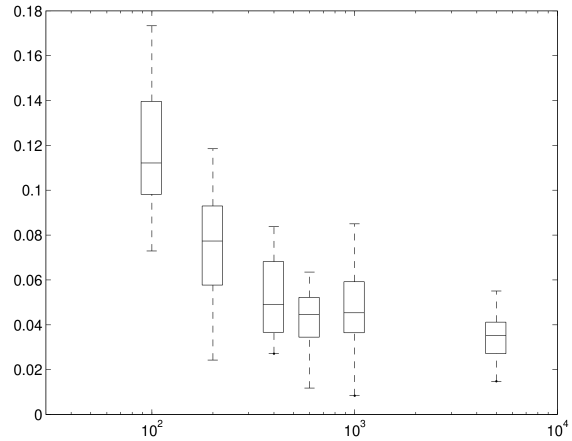


Figure 5.6: Boxplots of $\widehat{u}_n$ values obtained from 20 replicated simulations in cases when $f = g_1$, $\sigma^2 = 0.04$ and $n = 100, 200, 400, 600, 1000$ and 5000.

In Section 4 we proved the consistency of $\widehat{f}_4$, in which the threshold u_n is required to tend to zero as n tends to infinity. In the practical implementation of $\widehat{f}_4$, 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 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 $\widehat{f}_4$ 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 Random design and heteroscedastic errors

In this section we explore the performance of the estimates $\widehat{a}_{0,c}$, $\widehat{f}_1$, $\widehat{f}_3$ and $\widehat{f}_4$ in cases when the design points are random and the errors are heteroscedastic. For simplicity, our discussion is restricted to $f = g_1$.

5.2.1 Random design with homoscedastic errors

Consider the regression model

$$Y = g_1(X) + \varepsilon,$$

where X has a Uniform distribution on $[0,1]$ and $\varepsilon \sim N(0, 0.2^2)$. We generated $N = 200$ samples of size $n = 200$ each from this model. The estimates $\widehat{a}_{0,c}$, $\widehat{f}_1$, $\widehat{f}_3$ and $\widehat{f}_4$ are then applied to each sample and their performance is evaluated by the global and local MISE values, as described in Section 4.1. The estimation of these MISE values in the current case is discussed below.

Let $X^j_{(i)}$ be the i^{th} ordered design point of the j^{th} sample, for $1 \le i \le n$ and $1 \le j \le N$. The global MISE value can then be estimated via (5.2) with $\widehat{\mathrm{ISE}}_j$ obtained by

$$\widehat{\mathrm{ISE}}_j = \sum_{i=1}^{n-1} \left(\frac{\mathrm{SE}_j(X^j_{(i)}) + \mathrm{SE}_j(X^j_{(i+1)})}{2} \right) \times (X^j_{(i+1)} - X^j_{(i)})$$

and $\mathrm{SE}_j((X^j_{(i)})) = \{\widehat{f}^j(X^j_{(i)}) - f(X^j_{(i)})\}^2$. This estimation is obtained by applying the trapezoidal rule for the integration. Similarly the local MISE_s can be estimated via (5.3) with

$$\widehat{\mathrm{ISE}}_{s,j} = \sum_{i=1}^{n-1} \left(\frac{\mathrm{SE}_j(X^j_{(i)}) + \mathrm{SE}_j(X^j_{(i+1)})}{2} \right) \times (X^j_{(i+1)} - X^j_{(i)}) I \left[s - 0.05 < X^j_{(i)} < X^j_{(i+1)} < s + 0.05 \right].$$

The global and local MISE values of the four estimates are summarized in the first and second columns of Table 5.4, respectively. By comparing these results with the corresponding fixed-design results in Table 5.1 (cf. the columns marked by $\sigma = 0.2$), we can see that the MISE values in the random-design case are a little bigger than those in the fixed-design case. Other than that, the four estimates perform similarly in both cases.

5.2.2 Heteroscedastic errors with fixed design

Next we investigate the performance of the four estimates in the case when the design points are fixed but the errors are heteroscedastic. Toward this end, let us consider the following regression model

$$Y = g_1(x) + \sigma(x)\varepsilon,$$

where $\varepsilon \sim \mathrm{N}(0;1)$ and

$$\sigma^2(x) = \begin{cases} \frac{2}{3}, & \text{on } [0, 0.5) \\ \frac{2}{3}(\cos(4\pi x - \pi) + 2), & \text{on } [0.5, 1]. \end{cases}$$

It can be checked that $\int_0^1 \sigma^2(x)\, dx = 0.04$. Therefore we have the same value of the integrated variance in this case as in the simulation study described above for the random-design case. The variance function is shown in Figure 5.7.

Then $N = 200$ replicated simulations were performed. The $\widehat{\mathrm{MISE}}$ and $\widehat{\mathrm{MISE}}_{s=0.3} + \widehat{\mathrm{MISE}}_{s=0.7}$ values are computed in the way as described in Section 5.1. They are presented in the two

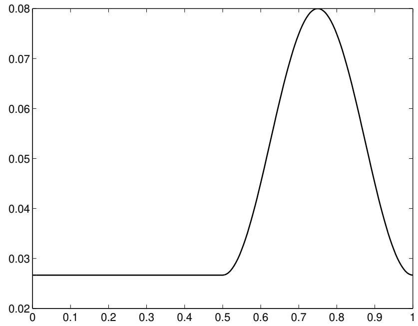


Figure 5.7: Plot of $\sigma^2(x)$.

Table 5.4: $\widehat{\text{MISE}}$ (first columns) and $\widehat{\text{MISE}}_{s=0.3} + \widehat{\text{MISE}}_{s=0.7}$ (second columns) values in the random-design case (2 left-most columns) and heteroscedastic errors case (2 right-most columns).

Method	$X \sim U[0,1], \sigma^2 = 0.04$		Fixed design and $\int_0^1 \sigma^2(x)dx = 0.04$	
$\widehat{a}_{c,0}$	0.0130	0.0096	0.0114	0.0071
$\widehat{f}_1$	0.0227	0.0042	0.0129	0.0036
$\widehat{f}_3$	0.0131	0.0085	0.0106	0.0069
$\widehat{f}_4$	0.0098	0.0060	0.0097	0.0068

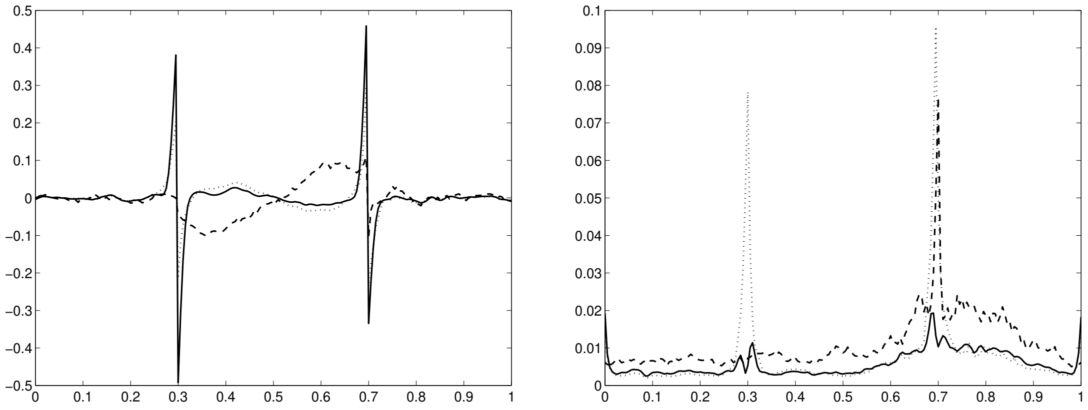
right-most columns of Table 5.4. Note that the results of $\widehat{a}_{c,0}$, $\widehat{f}_1$ and $\widehat{f}_3$ here are similar to the corresponding results reported in Section 5.1 when the errors are homoscedastic. For $\widehat{f}_4$, although its global and local MISE values are a little bit larger here than the corresponding values reported in Section 5.1, they are still the smallest compared the $\widehat{\text{MISE}}$ values of the other estimates. Hence $\widehat{f}_4$ also outperforms its competitors in cases with heteroscedastic errors.

To further illustrate the impact of the variance function $\sigma^2(x)$, we plot in Figure 5.8 the estimated bias and variance functions for the estimates $\widehat{a}_{c,0}$, $\widehat{f}_1$ and $\widehat{f}_4$. We can see that the bias functions look quite similar to those in Figure 5.5(a). The estimated variance functions are most affected in regions with large $\sigma^2(x)$ values (cf. Figures 5.5(b) and 5.8(b)).

For facilitating a further comparison among the estimates $\widehat{a}_{c,0}$, $\widehat{f}_1$, $\widehat{f}_3$ and $\widehat{f}_4$ in all three cases with (i) fixed-design and homoscedastic errors; (ii) random-design and homoscedastic errors; and (iii) fixed-design and heteroscedastic errors, we present in Figure 5.9 the boxplots of $\widehat{\text{ISE}}$'s for the four estimates in the three cases. We can see that all estimates are affected by the random design although $\widehat{f}_1$ seems to be affected most dramatically. Overall the effect of heteroscedastic errors on the estimates is small.

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



(a) (b)

Figure 5.8: Plots of the estimated bias functions (left panel) and variance functions (right panel) for the esitmates $\widehat{a}_c$ (solid curves), $\widehat{f}_1$ (dashed curves) and $\widehat{f}_4$ (dotted curves) in the case when $f = g_1$ and the error variance equals $\sigma^2(x)$.

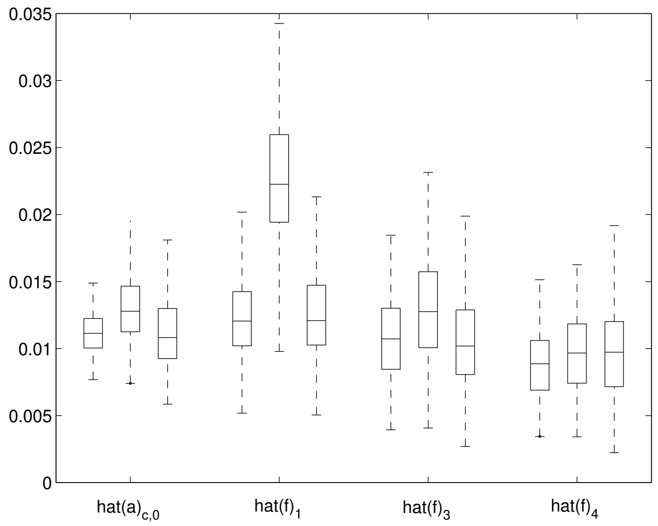


Figure 5.9: Boxplots of $\widehat{\mathrm{ISE}}_j, j = 1, \ldots, N$ for the four estimates when $f = g_1$ and $n = 200$. For each estimate we present three boxplots. First boxplot: fixed design, homoscedastic errors and $\sigma^2 = 0.04$; Second boxplot: random-design, $X \sim \mathrm{U}[0, 1]$, homoscedastic errors and $\sigma^2 = 0.04$; Third boxplot: fixed-design, heteroscedastic errors and $\int_0^1 \sigma^2(x)\, dx = 0.04$.

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 $\widehat{f}_4$. 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

$\widehat{f}_4$ (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 $\widehat{f}_4$ preserves well the discontinuities and removes the noise efficiently in continuity regions.

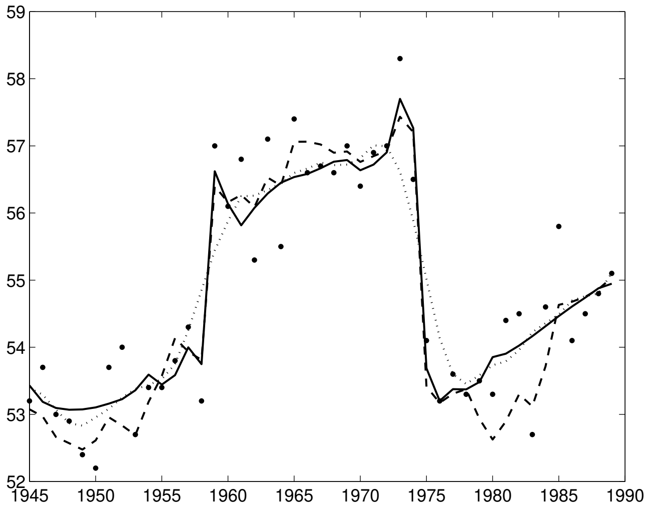


Figure 6.1: The data (small dots) and the estimates $\widehat{a}_{c,0}(x)$ (dotted curve), $\widehat{f}_1(x)$ (dashed curve) and $\widehat{f}_4(x)$ (solid curve). The cross-validation scores of the three estimates are 0.8475, 1.0568 and 0.7380, respectively.

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

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

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

$$\begin{aligned} \text{WRMS}_\ell(x) &= g_\ell(x;\widehat{a}_{\ell,0}(x),\widehat{a}_{\ell,1}(x))/\sum_{x_i<x}\alpha_i(x) \\ \text{WRMS}_r(x) &= g_r(x;\widehat{a}_{r,0}(x),\widehat{a}_{r,1}(x))/\sum_{x_i\geq x}\alpha_i(x) \\ \text{WRMS}_c(x) &= g_c(x;\widehat{a}_{c,0}(x),\widehat{a}_{c,1}(x))/\sum_{x_i}\alpha_i(x), \end{aligned} \tag{A.2}$$

where $\alpha_i(x) = K((x_i - x)/h_n) \geq 0$ for all i. By (A.1) and (A.2), we find

$$\begin{aligned}
\mathrm{WRMS}_c(x) &= \{g_\ell(x; \widehat{a}_{c,0}(x), \widehat{a}_{c,1}(x)) + g_r(x; \widehat{a}_{c,0}(x), \widehat{a}_{c,1}(x))\} / \sum_{x_i} \alpha_i(x) \\
&\geq \{g_\ell(x; \widehat{a}_{\ell,0}(x), \widehat{a}_{\ell,1}(x)) + g_r(x; \widehat{a}_{r,0}(x), \widehat{a}_{r,1}(x))\} / \sum_{x_i} \alpha_i(x) \\
&= \frac{\sum_{x_i<x} \alpha_i(x)}{\sum_{x_i} \alpha_i(x)} \left(\frac{g_\ell(x; \widehat{a}_{\ell,0}(x), \widehat{a}_{\ell,1}(x))}{\sum_{x_i<x} \alpha_i(x)} \right) + \frac{\sum_{x_i\geq x} \alpha_i(x)}{\sum_{x_i} \alpha_i(x)} \left(\frac{g_r(x; \widehat{a}_{r,0}(x), \widehat{a}_{r,1}(x))}{\sum_{x_i\geq x} \alpha_i(x)} \right) \\
&= \frac{\sum_{x_i<x} \alpha_i(x)}{\sum_{x_i} \alpha_i(x)} \mathrm{WRMS}_\ell(x) + \frac{\sum_{x_i\geq x} \alpha_i(x)}{\sum_{x_i} \alpha_i(x)} \mathrm{WRMS}_r(x) \\
&\geq \min(\mathrm{WRMS}_\ell(x), \mathrm{WRMS}_r(x)).
\end{aligned}$$

This finishes the proof. □

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:

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

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

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

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)
$\widehat{a}_{c,0}$	$\left(\frac{h_n}{2}, 1 - \frac{h_n}{2}\right)$	$\left[0, \frac{h_n}{2}\right] \bigcup \left[1 - \frac{h_n}{2}, 1\right]$
$\widehat{a}_{r,0}$	$\left[0, 1 - \frac{h_n}{2}\right)$	$\left[1 - \frac{h_n}{2}, 1\right]$
$\widehat{a}_{\ell,0}$	$\left(\frac{h_n}{2}, 1\right]$	$\left[0, \frac{h_n}{2}\right]$

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

$$\frac{w_{k,j}}{n h_n^{k+1}} = v_{k,j} + O(\frac{1}{n h_n}), \tag{A.5}$$

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

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

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

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

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:

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

where for $s = 0, 1$,

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

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

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

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

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

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, $P(|\varepsilon_n|\geq t_n\ i.o.)=0$, or equivalently. $P(\varepsilon_n\neq\tilde{\varepsilon}_n, i.o.)=0$. So there exists a full set Ω_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 ω and K, but not on x. Therefore

$$\lim_{n\to\infty}\sup_x\sqrt{\frac{nh_n}{\ln n}}|A_s(x)|=0, a.s. \tag{A.13}$$

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

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

where $\|K\| = \sup_{x\in[-1/2,1/2]} |K(x)|$, and where $i_0 \le j_0$ 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, $j_0 - i_0 \le nh_n + 2$, $1/t_i < 1/\sqrt{i}$, and $\sum_{i=i_0+1}^{j_0} \frac{1}{\sqrt{i}} \le \sum_{i=1}^{\lfloor nh_n+2 \rfloor} \frac{1}{\sqrt{i}}$, where $\lfloor x \rfloor$ denotes the integer part of x. So

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

Therefore,

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

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 $|x - v(x)| \le d_n^{-1}$. We first write

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

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

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

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

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

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 $\{U_{n,i}; i = 1, \ldots, n\}$ into $2q_n$ blocks of size ℓ_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 $\ell_n = \lfloor \frac{\sqrt{nh_n}}{t_n\sqrt{\ln n}} \rfloor \to 0$ as $n \to \infty$, and

$$V_n(m) = \frac{1}{n} \sum_{t=(m-1)\ell_n+1}^{m\ell_n} U_{n,t}, \quad W_n'(v(x)) = \sum_{m=1}^{q_n} V_n(2m-1),$$
$$W_n''(v(x)) = \sum_{m=1}^{q_n} V_n(2m), \quad W_n'''(v(x)) = \frac{1}{n} \sum_{t=2q_n\ell_n+1}^{n} U_{n,t}.$$

Then

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

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

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

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

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

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

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

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

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

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

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

where A_1 is a constant depending on K and σ^2. So we obtain

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

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

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

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

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

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

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

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

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

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

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

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

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

Finally, by (A.7) and (A.24), we can conclude that

$$\sup_{x\in(h_n/2,1-h_n/2)} |\widehat{a}_{,0}(x) - f(x)| = O(h_n^2) + O(\sqrt{\frac{\ln n}{nh_n}}),\ a.s.$$
$$\sup_{x\in(h_n/2,1]} |\widehat{a}_{\ell,0}(x) - f(x)| = O(h_n^2) + O(\sqrt{\frac{\ln n}{nh_n}}),\ a.s.$$
$$\sup_{x\in[0,1-h_n/2)} |\widehat{a}_{r,0}(x) - f(x)| = O(h_n^2) + O(\sqrt{\frac{\ln n}{nh_n}}),\ a.s.$$

By the condition that $\sqrt{\frac{nh_n^5}{\ln n}} \to 0$, this leads to

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

This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2.
We can rewrite the estimate $\widehat{f_4}(x)$, defined in (2.18), as:

$$\widehat{f_4}(x) = \widehat{a}_{c,0}(x)I(A_n(x)) + \widehat{a}_{\ell,0}(x)I(B_n(x)) + \widehat{a}_{r,0}(x)I(C_n(x)) + \frac{\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x)}{2}I(D_n(x)), \tag{A.25}$$

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

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

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 Ω_1 such that for every $\omega \in \Omega_1$, $\lim_{n\to\infty} \sup_{x\in[\rho,1-\rho];\ a=c,\ell,r;\ i=1,2,3} R_{n,a,i}(x,\omega) = 0$ (for convenience of explanation, we sometimes make ω 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,

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

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

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

where C is a constant. Then we have

$$\begin{aligned}
\sup_{x\in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{f}_4(x) - f(x)| &= \sup_{x\in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{a}_{c,0}(x) - f(x)| I(A_n(x)) + \\
&\quad \sup_{x\in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{a}_{\ell,0}(x) - f(x)| I(B_n(x)) + \\
&\quad \sup_{x\in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{a}_{r,0}(x) - f(x)| I(C_n(x)) + \\
&\quad \sup_{x\in D_1} \sqrt{\frac{\ln n}{nh_n}} \left| \frac{\widehat{a}_{\ell,0}(x) + \widehat{a}_{r,0}(x) - 2f(x)}{2} \right| I(D_n(x)) \\
&\leq 4C, a.s..
\end{aligned}$$

So

$$\sup_{x\in D_1} \sqrt{\frac{\ln n}{nh_n}} |\widehat{f}_4(x) - f(x)| = O(1), \ a.s.$$

Part Two:

In this part, we prove the uniform consistency of $\widehat{f}_4$ 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. $\mathrm{diff}(x,\omega)$ defined in (2.19) becomes

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

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

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

which implies that

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

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

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

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 $n \geq n(\omega, \frac{b}{2}) =: n_2$,

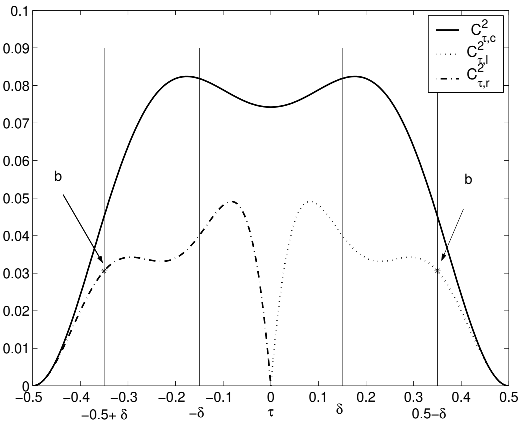


Figure A.1: A graphical illustration of the definition of b used in the proof of Theorem 3.2.

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

Moreover, since $u_n \to 0$, 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 $n \geq n_3 = \max(n_2, n(\frac{b}{4}))$,

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

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

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

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

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

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 $n \geq n_5 := \max(n_4, n_1)$,

$$\sup_{D_{2,\delta,\ell}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_4(x,\omega) - f(x)| = \sup_{D_{2,\delta,\ell}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{a}_{\ell,0}(x,\omega) - f(x)| \leq C.$$

Similarly, we can prove that

$$\sup_{D_{2,\delta,r}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_4(x) - f(x)| = O(1), \ \ a.s.$$

So

$$\sup_{D_{2,\delta}} \sqrt{\frac{nh_n}{\ln n}} |\widehat{f}_4(x) - f(x)| = O(1), \ \ a.s.$$

Part Three: When $x \in D_2 \backslash D_{2,\delta}$, $\widehat{f}_4$ 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_τ and b_τ in such cases (cf. Figure A.1). □

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*, to appear.

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 Titterington 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.

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.

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., Godtliebsen, 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.