Applications of Mathematics

Irène Gijbels

Smoothing and preservation of irregularities using local linear fitting

Applications of Mathematics, Vol. 53 (2008), No. 3, 177–194

Persistent URL: http://dml.cz/dmlcz/140314

53 (2008) APPLICATIONS OF MATHEMATICS No. 3, 177–194

SMOOTHING AND PRESERVATION OF IRREGULARITIES USING LOCAL LINEAR FITTING*

Irène Gijbels, Leuven

(Invited)

Abstract. For nonparametric estimation of a smooth regression function, local linear fitting is a widely-used method. The goal of this paper is to briefly review how to use this method when the unknown curve possibly has some irregularities, such as jumps or peaks, at unknown locations. It is then explained how the same basic method can be used when estimating unsmooth probability densities and conditional variance functions.

Keywords: density estimation, irregularities, jumps, local linear fitting, mean, peaks, preservation, smoothing, variance

MSC 2010: 62G08, 62G05, 62G20

1. Introduction

In regression analysis the aim is to describe the functional influence that a covariate random variable X has on a random variable of interest Y. One is forced to rely on nonparametric methods if no prior knowledge is available on the functional form of the relationship. The available nonparametric methods are kernel methods, local polynomial fitting, spline estimation and wavelet-based methods, among others. All these methods have been studied in detail when the unknown curves possess some degree of smoothness (often in terms of differentiability). Special attention has to be paid when using smoothing methods to estimate functions that show certain irregularities, such as jumps, peaks, cusps, etc.

In this paper we focus on local linear fitting and review how this method can be adapted in a simple way to estimate a regression curve that possibly shows jumps at an unknown number of points (of unknown locations). This is done in Sections 2

* This research was supported by GOA/07/04-project of the Research Fund KULeuven. Support from the IAP research network nr. P6/03 of the Federal Science Policy, Belgium, is also acknowledged.

and 3. From this basic estimation procedure, one can further derive an appropriate estimator when the unknown curve shows peaks, as outlined in Section 4. Further extensions consist in estimating a non-smooth density and in estimating a conditional variance function in a regression setup. These are briefly discussed in Section 5.

2. Local linear curve fitting and smoothing

In this section we briefly review some basic properties of local linear fitting, in particular those that will have a link with results presented later on for non-smooth curves.

Consider the simplest form of a nonparametric regression model:

$$Y = m(X) + \varepsilon, \tag{2.1}$$

where a classical assumption on the error term ε is that the conditional expectation of ε given $X = x$ is zero and the conditional variance of ε given $X = x$ is $\sigma^2(x)$ (> 0 and finite). Consequently, under this model setup, we have $m(x) = E(Y|X = x)$, the conditional mean function (also referred to as the mean regression function). The aim is then to estimate this unknown function from the observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ from model (2.1):

$$Y_i = m(X_i) + \varepsilon_i, \tag{2.2}$$

with the error terms ε_i being independent random variables having the same distribution as ε.

The idea behind local linear fitting is very simple. If the interest is in estimating the regression function $m(\cdot)$ at a given point x then approximate the unknown curve $m(\cdot)$ locally—in a neighbourhood of x—by a linear function, and fit this linear function via least squares. In other words, using

$$m(z) \approx m(x) + m'(x)(z - x) \equiv a + b(z - x) \tag{2.3}$$

for z in a neighbourhood of the given point x, we arrive at the minimization problem

$$\text{minimize}_{a,b} \ \sum_{i=1}^{n} \{Y_i - (a + b(X_i - x))\}^2 K_h(X_i - x), \tag{2.4}$$

where $K_h(u) = h^{-1}K(u/h)$ is a rescaled version of a given kernel function K, and $h = h_n > 0$ is a bandwidth parameter. Commonly K is chosen to be a symmetric probability density function. The bandwidth parameter h controls the size of the

(approximating) neighbourhood while K determines the weight assigned to each observation within that neighbourhood. This is illustrated in Fig. 2.1. The weight function is shown as the dotted curve and boundaries of the neighbourhood are indicated as vertical dashed lines. The Epanechnikov kernel $K(u) = 0.75(1-u^2)1\{|u| \leqslant 1\}$ is used in the illustration.

The resulting estimators

$$(\hat{a}_0(x), \hat{a}_1(x)) = \arg\min_{a,b} \sum_{i=1}^{n} \{Y_i - (a + b(X_i - x))\}^2 K_h(X_i - x) \tag{2.5}$$

have nice and simple expressions

$$\hat{a}_0(x) = \sum_{i=1}^{n} K\Big(\frac{X_i - x}{h_n}\Big) \frac{w_{2,K} - w_{1,K}(X_i - x)}{w_{2,K} w_{0,K} - w_{1,K}^2} Y_i,$$
$$\hat{a}_1(x) = \sum_{i=1}^{n} K\Big(\frac{X_i - x}{h_n}\Big) \frac{w_{0,K}(X_i - x) - w_{1,K}}{w_{2,K} w_{0,K} - w_{1,K}^2} Y_i$$

where $w_{k,K} = \sum_{i=1}^{n} K((X_i - x)/h_n)(X_i - x)^k$, $k = 0, 1, 2$ is a sample version of the kth moment of the kernel function K.

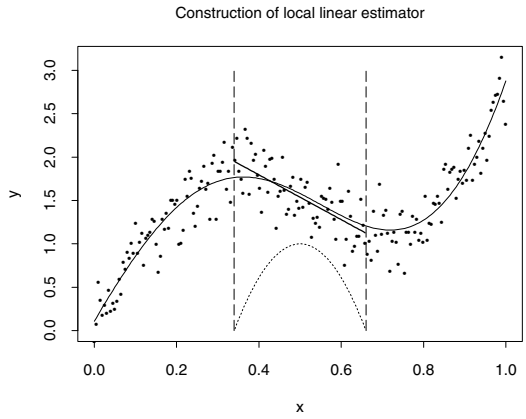

Figure 2.1. Some simulated data and construction of the local linear estimator at $x = 0.5$. Solid curve: true curve; dotted curve: the weights for estimation at $x = 0.5$.

Note that with $(\hat{a}_0(x), \hat{a}_1(x))$ one gets simultaneously an estimator for the regression function $m(x)$ as well as for its derivative $m'(x)$. Local linear fitting, and more generally local polynomial fitting (obtained by approximating in (2.3) by a polynomial function), have been studied in full detail during the last decades, and the merits of this smoothing method are very well understood. See for example Fan

and Gijbels [5] and references therein. Since the early nineties the method has been successfully applied in a variety of settings including more complex ones such as time series and multivariate analysis.

When studying the asymptotic behaviour of the local linear estimator $(\hat{a}_0(x), \hat{a}_1(x))$ one should make the distinction between the interior region and the boundary regions. Suppose, without loss of generality, that the design points X_i are in the interval $[0,1]$, i.e. that the density of X, denoted by $f_X(\cdot)$, has support $[0,1]$. To explain the consequences, suppose that the support of the kernel function is $[-1/2, 1/2]$. With x fixed this means that points y which are such that $(y-x)/h \in [-1/2, 1/2]$, or equivalently, $x - 1/2h \leqslant y \leqslant x + 1/2h$, will get a nonzero weight $K_h(y-x)$ assigned by the kernel function. If x is such that either $x - 1/2h < 0$ or $x + 1/2h > 1$ then this means that the kernel assigns weights to points y outside the interval $[0,1]$. These cases, $x < 1/2h$ and $x > 1 - 1/2h$, therefore deserve some special attention. This leads to the distinction between

$$\begin{aligned} &\text{interior region:} && [h_n/2, 1 - h_n/2]; \\ &\text{boundary region:} && [0, h_n/2) \cup (1 - h_n/2, 1]. \end{aligned}$$

One of the merits of local linear fitting is that the method leads to consistent estimates of the regression function and its derivative even in the boundary regions. Moreover, the rate of convergence of the regression estimator is the same in the boundary region as in the interior region. The latter is in contrast to, for example, the Nadaraya-Watson estimator (Nadaraya [22] and Watson [30]) that suffers from a slower convergence rate in the boundary regions. Although there is no difference in convergence rates in both regions, there are some small differences noticeable in the asymptotic behaviour in both type of regions.

As can be found e.g. in Ruppert and Wand [27] and Fan and Gijbels [5], the Mean Squared Error (MSE) of the local linear estimator $\hat{a}_0(x)$ in the interior region $x \in [h_n/2, 1 - h_n/2]$ is given by

$$\text{MSE}(\hat{a}_0(x)) = \Big[\frac{h_n^2}{2} m''(x) B_{c,K}\Big]^2 + \frac{\sigma^2(x)}{f_X(x)} \frac{1}{nh_n} V_{c,K} + o\Big(h_n^4 + \frac{1}{nh_n}\Big) \tag{2.6}$$

where

$$B_{c,K} = \int_{-1/2}^{1/2} u^2 K(u)\,\mathrm{d}u = v_{c,2}, \quad v_{c,j} = \int_{-1/2}^{1/2} u^j K(u)\,\mathrm{d}u \tag{2.7}$$

and

$$V_{c,K} = \int_{-1/2}^{1/2} K^2(u)\,\mathrm{d}u.$$

Here the index c stands for 'central' referring to the fact that the estimator uses datapoints that fall in the interval $[x-h_n/2, x+h_n/2]$, an interval 'centered' around x.

Consider now a fixed point x in the left boundary region $[0, h_n/2)$. We can write

$$x = \tau h_n \quad \text{with} \quad 0 \leqslant \tau < 1/2,$$

where the quantity τ is nothing but the distance of the point x from the left-endpoint 0. For this point the MSE of the local linear estimator is given by

$$\text{MSE}(\hat{a}_0(x)) = \Big[\frac{h_n^2}{2} m''(0+) B_{c,K}(\tau)\Big]^2 + \frac{\sigma^2(0+)}{f_X(0+)} \frac{1}{nh_n} V_{c,K}(\tau) + o\Big(h_n^4 + \frac{1}{nh_n}\Big) \tag{2.8}$$

where now

$$B_{c,K}(\tau) = \frac{v_{c,2}^2(\tau) - v_{c,1}(\tau) v_{c,3}(\tau)}{v_{c,2}(\tau) v_{c,0}(\tau) - v_{c,1}^2(\tau)} \tag{2.9}$$

and

$$V_{r,K}(\tau) = \int_{-\tau}^{1/2} \Big(\frac{v_{c,2}(\tau) - u\, v_{c,1}(\tau)}{v_{c,2}(\tau) v_{c,0}(\tau) - v_{c,1}^2(\tau)}\Big)^2 K^2(u)\, \mathrm{d}u,$$

where $v_{c,j}(\tau) = \int_{-\tau}^{1/2} u^j K(u)\, \mathrm{d}u$.

For right boundary points, i.e. for $x = 1 - \tau h_n$ with $-1/2 < \tau \leqslant 0$, there are similar kinds of expressions. See Fan and Gijbels [5].

From the mean squared error expressions in (2.6) and (2.8) it is clear that the differences in the first-order asymptotic behaviour are in these constants, depending on K and on τ appearing in the squared bias and variance. To give some idea about the differences in these constants we plot in Fig. 2.2 the ratios $B_{c,K}^2(\tau)/B_{c,K}^2$ as well as $V_{c,K}(\tau)/V_{c,K}$ as functions of τ for three different kernels: the Gaussian kernel, the Epanechnikov kernel and the uniform kernel. Clearly, all ratios of squared biases are below 1, whereas all ratios of variances are above 1. So at a boundary point, the bias tends to be smaller and the variance bigger. At the left-boundary point 0 these ratios become $B_{c,K}^2(0)/B_{c,K}^2$ and $V_{c,K}(0)/V_{c,K}$. In Tab. 2.1 we list these values for the three kernels in Fig. 2.2.

kernel	ratio $B_{c,K}^2(0)/B_{c,K}^2$	ratio $V_{c,K}(0)/V_{c,K}$
Gaussian	0.5659	6.3308
Epanechnikov	0.3352	7.4966
uniform	0.2500	8.0000

Table 2.1. The ratios $B_{c,K}^2(0)/B_{c,K}^2$ and $V_{c,K}(0)/V_{c,K}$ for the Gaussian, the Epanechnikov and the uniform kernel.

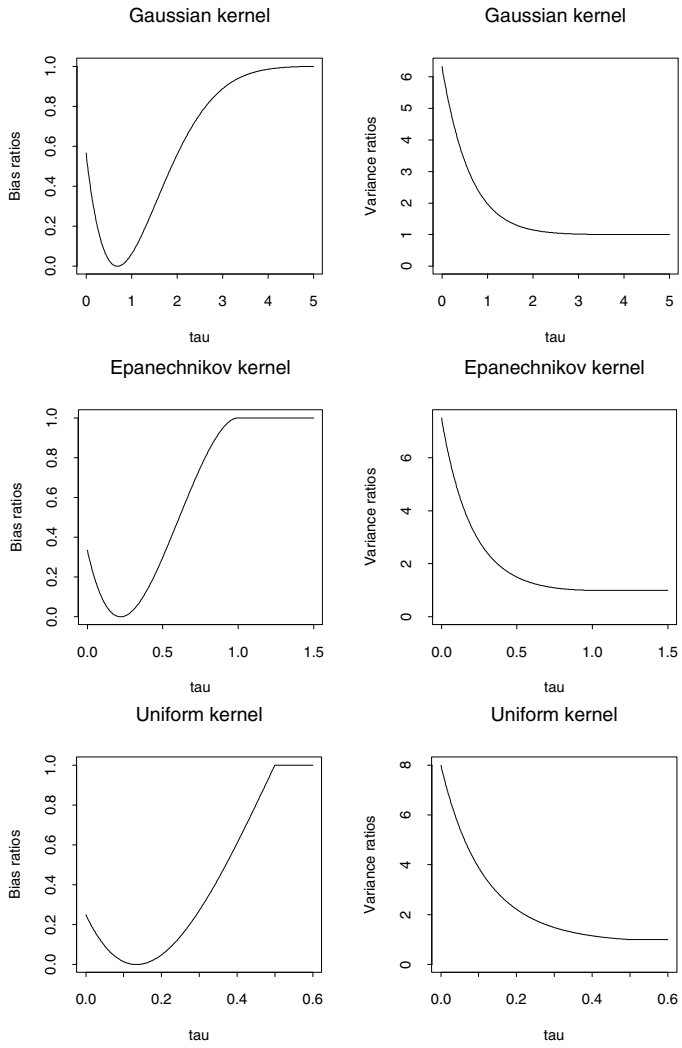

Figure 2.2. Behaviour of the ratios of the constant factors in squared bias (left) and in variance (right).

3. Smoothing and jump preservation

The asymptotic expressions in Section 2 are of course derived under some (sufficient) assumptions. One of them is that the unknown function m has the second order derivative.

What happens now if the unknown function has some irregularities? The simplest type of irregularities are discontinuities or jump points. So, suppose that the function m in the regression model (2.1) is such that it jumps at points s_j in $[0, 1]$,

with jump magnitudes d_j, $j = 1, \ldots, M$. The number of jump points M is finite but unknown.

For simplicity of presentation we will from now on concentrate on fixed and equispaced design, $x_i = i/n$ (and hence $f_X = 1$). We further act as if there were only one jump point. All these assumptions are imposed only for simplicity of presentation and can be relaxed (e.g. to random design).

Suppose that the unknown function m has a jump of size d at location $s \in (0, 1)$. When applying the local linear estimator to the data from model (2.1) with this regression function, the resulting estimator will be unsatisfactory since it will smooth out (or 'blur') completely the jump point. Intuitively this is clear since the estimator uses the data in the interval $[x - h_n/2, x + h_n/2]$, and when x is close to a jump point then this neighbourhood contains the data from the right-hand side of the jump-location as well as from the left-hand side of that location. See also the top panel of Fig. 3.1. This effect is also clear from the first-order asymptotic MSE-expressions provided in the first row of Tab. 3.1. These asymptotic results were established by Hamrouni [16] and Grégoire and Hamrouni [13].

How to get to a consistent estimator in this case? Suppose that the point x is close and to the left of the jump point s. In that case, we should avoid using the data from the right-hand side of x (since they would blur the effect of the jump). Similarly, if the point x is close to the jump point s but situated to the right of it, then we should avoid using the data from the left-hand side of x. This naturally leads to the consideration of a *left* local linear estimator and a *right* local linear estimator. More precisely, defining two one-sided kernels from the kernel K by putting

$$K_l(x) = K(x) \ \forall\, x \in [-1/2, 0), \quad K_r(x) = K(x) \ \forall\, x \in [0, 1/2],$$

one considers the *left* (l) and *right* (r) local linear estimates of m and m':

$$(\hat{a}_{l,0}(x), \hat{a}_{l,1}(x)) = \arg\min_{a,b} \sum_{i=1}^{n} [Y_i - a - b(x_i - x)]^2 K_l\Big(\frac{x_i - x}{h_n}\Big),$$

$$(\hat{a}_{r,0}(x), \hat{a}_{r,1}(x)) = \arg\min_{a,b} \sum_{i=1}^{n} [Y_i - a - b(x_i - x)]^2 K_r\Big(\frac{x_i - x}{h_n}\Big).$$

For each fixed point x we consider three local linear estimators: the central local linear estimator, the left local linear estimator and the right local linear estimator. Important to note is that they use the data from different neighbouring regions: the central estimator uses the data in $[x - h_n/2, x + h_n/2]$, whereas the left and right local linear estimators rely solely on the data in $[x - h_n/2, x)$ and $[x, x + h_n/2]$ respectively. Fig. 3.1 depicts the construction of the three estimators at the point $x = 0.49$, the neighbourhoods used and the kernel weights (dotted curves at the bottom).

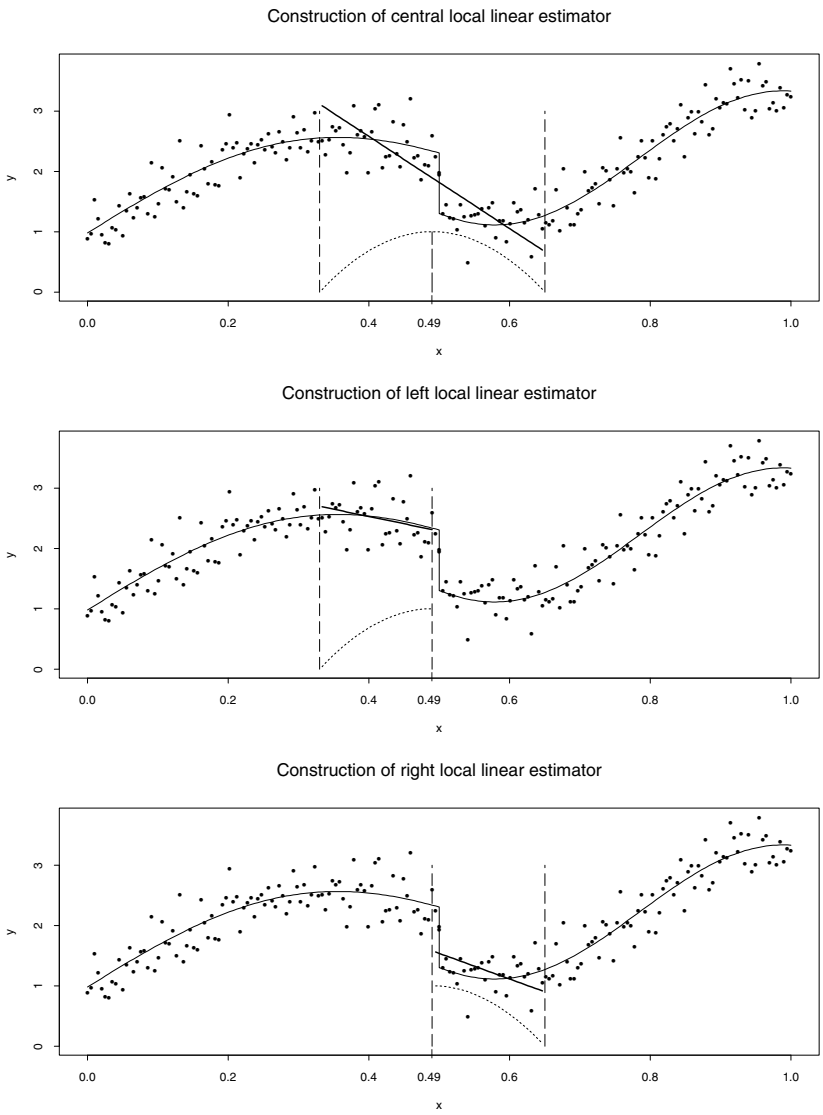

Figure 3.1. Construction of the central, left and right local linear estimators for a simulated dataset at the point $x = 0.49$ close and to the left of the jump point $s = 0.5$. Solid line: true regression function.

The asymptotic behaviour of these three estimators at a point x in the region of continuity of m is well known. Indeed, the behaviour of the left estimator corresponds to the behaviour of the central estimator at a left-boundary point (with $\tau = 0$), as given in (2.8). Note that we deliberately used the same notation τ in Section 2, although the setup is different. Further, the first-order asymptotic behaviour of the

three estimators in a neighbourhood of the jump point s is summarized in Tab. 3.1. For details see Hamrouni [16] and Gijbels et al. [12].

estim.	x continuity point	left neighbourhood of jump point $x = s + \tau h_n$ with $\tau \in [-\frac{1}{2}, 0)$	right neighbourhood of jump point $x = s + \tau h_n$ with $\tau \in [0, \frac{1}{2}]$
$\hat{a}_{c,0}(x)$	$[\frac{1}{2}h_n^2 m''(x)B_{c,K}]^2 + \frac{\sigma^2}{nh_n}V_{c,K}$	$\Big[d\int_{\lvert\tau\rvert}^{1/2} K(u)\,\mathrm{d}u\Big]^2 + \frac{\sigma^2}{nh_n}V_{c,K}$	$\Big[-d\int_{-1/2}^{-\lvert\tau\rvert} K(u)\,\mathrm{d}u\Big]^2 + \frac{\sigma^2}{nh_n}V_{c,K}$
$\hat{a}_{l,0}(x)$	$[\frac{1}{2}h_n^2 m''(x)B_{l,K}]^2 + \frac{\sigma^2}{nh_n}V_{l,K}$	$[\frac{1}{2}h_n^2 m''(s-)B_{l,K}]^2 + \frac{\sigma^2}{nh_n}V_{l,K}$	$\Big[-d\int_{-1/2}^{-\lvert\tau\rvert} K_l(u) \times \frac{v_{l,2}-v_{l,1}u}{v_{l,0}v_{l,2}-v_{l,1}^2}\,\mathrm{d}u\Big]^2 + \frac{\sigma^2}{nh_n}V_{l,K}$
$\hat{a}_{r,0}(x)$	$[\frac{1}{2}h_n^2 m''(x)B_{r,K}]^2 + \frac{\sigma^2}{nh_n}V_{r,K}$	$\Big[d\int_{\lvert\tau\rvert}^{1/2} K_r(u) \times \frac{v_{r,2}-v_{r,1}u}{v_{r,0}v_{r,2}-v_{r,1}^2}\,\mathrm{d}u\Big]^2 + \frac{\sigma^2}{nh_n}V_{r,K}$	$[\frac{1}{2}h_n^2 m''(s+)B_{r,K}]^2 + \frac{\sigma^2}{nh_n}V_{r,K}$

Table 3.1 First-order asymptotic Mean Squared Error expressions for the central, left and right local linear estimators.

The constants in the asymptotic bias and variance are

$$B_{l,K} = \frac{v_{l,2}^2 - v_{l,1}v_{l,3}}{v_{l,0}v_{l,2} - v_{l,1}^2}, \qquad V_{l,K} = \int_{-1/2}^{0} K^2(u)\Big[\frac{v_{l,2} - v_{l,1}u}{v_{l,2}v_{l,0} - v_{l,1}^2}\Big]^2\,\mathrm{d}u,$$

$$B_{r,K} = \frac{v_{r,2}^2 - v_{r,1}v_{r,3}}{v_{r,0}v_{r,2} - v_{r,1}^2}, \qquad V_{r,K} = \int_{0}^{1/2} K^2(u)\Big[\frac{v_{r,2} - v_{r,1}u}{v_{r,2}v_{r,0} - v_{r,1}^2}\Big]^2\,\mathrm{d}u,$$

$$B_{c,K} = \int_{-1/2}^{1/2} u^2 K(u)\,\mathrm{d}u = v_{c,2}, \qquad V_{c,K} = \int_{-1/2}^{1/2} K^2(u)\,\mathrm{d}u$$

with

$$v_{l,j} = \int_{-1/2}^{0} u^j K_l(u)\,\mathrm{d}u, \quad v_{r,j} = \int_{0}^{1/2} u^j K_r(u)\,\mathrm{d}u, \quad v_{c,j} = \int_{-1/2}^{1/2} u^j K(u)\,\mathrm{d}u.$$

For a continuity point the ratios of these constants were provided in Tab. 2.1, indicating a smaller bias but a bigger variance for the left and right estimators. From the expressions in Tab. 3.1 it is clear that for x in a continuous region, it is preferable to use the central estimator, and for x near a jump point s one should use the left or right estimator. The crucial question is then how to decide upon these three estimators in a data-driven way (we do not know whether and where a jump occurs). A very simple idea here is to use the Weighted Residual Sums of Squares (WRSS)

as goodness-of-fit measures, namely

$$\text{WRMS}_{\text{j}}(x) = \frac{\sum_i [Y_i - \hat{a}_{\text{j},0} - \hat{a}_{\text{j},1}(x_i - x)]^2 K_{\text{j}}(\frac{x_i - x}{h})}{\sum_i K_{\text{j}}(\frac{x_i - x}{h})} \tag{3.1}$$

for $\text{j} = c, l, r$ where we denote $K_c = K$.

The first order asymptotic behaviour of these diagnostic quantities is given in Tab. 3.2 and can be found in Qiu [24], Lambert [18] and Gijbels et al. [12]. It has been shown that the remainder terms in these asymptotic results tend to zero uniformly in x. The quantities $C^2_{\tau,c}$, $C_{\tau,l}$ and $C_{\tau,r}$ are defined in terms of K and of τ, the distance from the jump point s. See the previous cited work for their definitions.

quantity	x continuity point	left nhd of jump point $x = s + \tau h_n$ with $\tau \in [-\frac{1}{2}, 0)$	right nhd of jump point $x = s + \tau h_n$ with $\tau \in [0, \frac{1}{2}]$
$\text{WRMS}_c(x)$	σ^2	$\sigma^2 + d^2 C^2_{\tau,c}$	$\sigma^2 + d^2 C^2_{\tau,c}$
$\text{WRMS}_l(x)$	σ^2	σ^2	$\sigma^2 + d^2 C^2_{\tau,l}$
$\text{WRMS}_r(x)$	σ^2	$\sigma^2 + d^2 C^2_{\tau,r}$	σ^2
$\text{diff}(x)$	0	$d^2 C^2_{\tau,c}$	$d^2 C^2_{\tau,c}$

Table 3.2. First-order asymptotic behaviour of the Weighted Residual Sum of Squares quantities.

Important is to look at the maximum of the pairwise differences between the three estimators:

$$\text{diff}(x) = \max(\text{WRMS}_c(x) - \text{WRMS}_l(x), \text{WRMS}_c(x) - \text{WRMS}_r(x)). \tag{3.2}$$

The first-order asymptotic behaviour of this difference type of diagnostic is provided in the last line of Tab. 3.2. This also serves as a motivation for considering the diagnostic quantities in (3.1). For details see Lambert [18] and Gijbels et al. [12]. Clearly the difference quantity in (3.2) allows for making a data-driven choice between the central local linear estimator on the one hand and on the left and right local linear estimator on the other. This all together motivated the estimator

$$\hat{m}(x) = \begin{cases} \hat{a}_{c,0}(x) & \text{if } \text{diff}(x) \leqslant u \\ \hat{a}_{l,0}(x) & \text{if } \text{diff}(x) > u \text{ and} \\ & \text{WRMS}_l(x) < \text{WRMS}_r(x) \\ \hat{a}_{r,0}(x) & \text{if } \text{diff}(x) > u \text{ and} \\ & \text{WRMS}_l(x) > \text{WRMS}_r(x) \\ (\hat{a}_{l,0}(x) + \hat{a}_{r,0}(x))/2 & \text{if } \text{diff}(x) > u \text{ and} \\ & \text{WRMS}_l(x) = \text{WRMS}_r(x), \end{cases} \tag{3.3}$$

where $u > 0$ is a threshold parameter, and where the last line is added for finite-sample situations. The chance of the last event to happen tends to zero as n tends to infinity.

McDonald and Owen [20] and Hall and Titterington [15] also considered the three types of estimators but did not get to direct estimation of the unknown curve with a data-driven decision rule. Qiu [24] only considered the left and right estimators and combined them using similar diagnostic quantities. See also Qiu [25].

A theoretical study (including strong consistency with rate and asymptotic normality) as well as simulation studies (including comparisons with wavelet-based procedures, among others) for the estimator in (3.3) can be found in Lambert [18] and Gijbels et al. [12]. Practical choices of h and u are provided there. Lambert [18] also gives evidence that the estimation method is applicable to random design and to heteroscedastic regression models, with $\sigma^2(x)$ instead of a constant σ^2. The estimation procedure is quite appealing due to its simplicity and its theoretical foundations. A generalization of the method to the case of estimation of non-smooth regression surfaces is given in Gijbels et al. [11].

4. Regression smoothing and peak preservation

Jump discontinuities are of course not the only type of irregularities that can occur in a function. The method reviewed in Section 3 can also be used for nonparametric estimation of a regression curve that shows possibly a spiky-behaviour. A spike can be formalized as a discontinuity in the derivative of the function.

So suppose we have a regression model (2.1) in which m is differentiable but possibly has an unknown number of points M_p at which the derivative function is discontinuous. For simplicity of presentation we focus on the case of one point s at which the derivative function m' has a discontinuity of size d'. An example of such a spiky function and a simulated dataset from model (2.1) is given in Fig. 4.1 (solid curve). A question is whether one can improve upon the local linear estimator for the derivative function m', namely $\hat{a}_{c,1}(x)$ as defined in (2.5)? In Fig. 4.1 the local linear estimator for the displayed dataset is plotted for two values of the bandwidth parameter. Note the considerable underestimation in the peak area.

We recall the asymptotic MSE for the estimator $\hat{a}_{1,c}$ for points in the interior region $[x - h_n/2, x + h_n/2]$,

$$\text{MSE}(\hat{a}_1(x)) = \left[\frac{h_n^2}{6}\Big\{m^{(3)}(x) + 3m''(x)\frac{f_X'(x)}{f_X(x)}\Big\}B_{c,K}^1\right]^2 + \frac{\sigma^2(x)}{f_X(x)}\frac{1}{nh_n^3}V_{c,K}^1 + o\Big(h_n^4 + \frac{1}{nh_n^3}\Big) \tag{4.1}$$

where

$$B^1_{c,K} = \left(\int_{-1/2}^{1/2} u^2 K(u)\,\mathrm{d}u\right)\left(\int_{-1/2}^{1/2} u^4 K(u)\,\mathrm{d}u\right)^{-1} = v_{c,2}\, v_{c,4}^{-1}$$

and

$$V^1_{c,K} = \left(\int_{-1/2}^{1/2} u^2 K^2(u)\,\mathrm{d}u\right)\left(\int_{-1/2}^{1/2} u^2 K(u)\,\mathrm{d}u\right)^{-2}.$$

For points in the boundary region the rate of convergence is lower, since in that region the bias will be of order h_n instead of h_n^2. See Fan and Gijbels [5]. Hence, this is also the order that we can hope to obtain for an estimator of m', based on local linear fitting, in a neighbourhood of a spike.

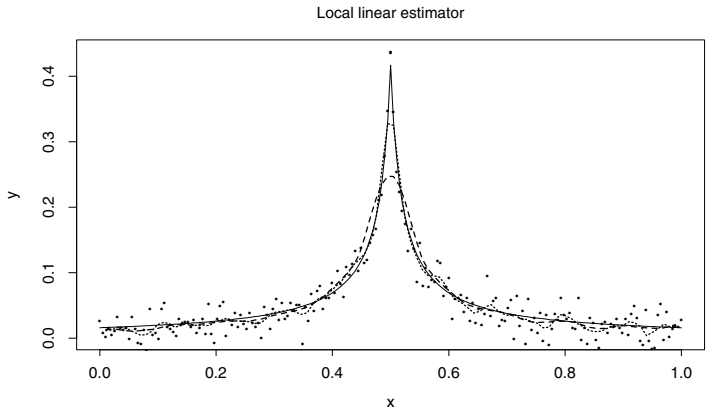

Figure 4.1. A function with a spike and simulated data ($n = 200$). The true curve (thin solid curve) and local linear fits for two values of h (dotted and dashed curves).

From the asymptotic properties of the local linear estimator one can see that the central local linear estimator is a consistent estimator at s, but that the rate of convergence of the estimator in a neighbourhood of the spike is of lower order than in regions where the function m' is differentiable. Indeed, by analogy with what happens in boundary regions for local polynomial fitting, the central local linear estimator has a bias of order h_n there, instead of h_n^2. Hence there is room for improvement. Tab. 4.1 indicates the rate of the first-order asymptotic bias of the central, left and right estimators. The quantities $Q^0_{\mathrm{j}}(K,\tau)$, $\mathrm{j} = c, r, l$ and alike, used and denoted in a generic way, are defined in terms of the kernel K and the distance τ to the location s of the spike. See Desmet and Gijbels [4] for detailed expressions.

estimator	left neighbourhood of spike	right neighbourhood of spike
$\mathrm{Bias}(\hat{a}_{r,0}(x))$	$h_n\, d'\, Q_r^0(K,\tau)$	$h_n^2\, Q_r^0(K)$
$\mathrm{Bias}(\hat{a}_{c,0}(x))$	$h_n\, d'\, Q_c^0(K,\tau)$	$h_n\, d'\, Q_c^0(K,\tau)$
$\mathrm{Bias}(\hat{a}_{l,0}(x))$	$h_n^2\, Q_l^0(K)$	$h_n\, d'\, Q_l^0(K,\tau)$
$\mathrm{Bias}(\hat{a}_{r,1}(x))$	$d'\, Q_r^1(K,\tau)$	$h_n\, Q_r^1(K)$
$\mathrm{Bias}(\hat{a}_{c,1}(x))$	$d'\, Q_c^1(K,\tau)$	$d'\, Q_c^1(K,\tau)$
$\mathrm{Bias}(\hat{a}_{l,1}(x))$	$h_n\, Q_l^1(K)$	$d'\, Q_l^1(K,\tau)$

Table 4.1. Behaviour of the bias of the estimators in various regions.

The basic ideas of the previous section can be applied, but an appropriate diagnostic quantity is needed. A good candidate is

$$\widetilde{\mathrm{WRSS}}_{\mathrm{j}}(x) = \frac{\sum_i [Y_i - \hat{a}_{\mathrm{j},0} - \hat{a}_{\mathrm{j},1}(x_i - x)]^2 K_{\mathrm{j}}\Big(\dfrac{x_i - x}{h}\Big)}{\sum_i (x_i - x)^2 K_{\mathrm{j}}\Big(\dfrac{x_i - x}{h}\Big)}, \quad \mathrm{j} = c, r, l. \tag{4.2}$$

A theoretical study of the behaviour of this diagnostic quantity and its use in an estimation procedure is established in Desmet and Gijbels [4]. With these $\widetilde{\mathrm{WRSS}}_{\mathrm{j}}(x)$'s we indeed have quantities that are such that the differences of them are, in the first order, independent of n, h_n and σ^2 but dependent on the size of the jump in the derivative (i.e. d') and of the distance of x from the location of the spike. Hence, these quantities allow for detecting locations at or near a spike, analogously to the quantities $\mathrm{WRMS}_{\mathrm{j}}(x)$'s in the case of discontinuities. Further research is done for combining the above ideas with the details of the method in Section 3 to come to an estimation procedure that is applicable for curves that show both jumps and peaks.

5. Further applications

In this section we briefly discuss two other applications of the method exposed in Section 3.

5.1. Estimating a non-smooth density

We now consider the situation of having i.i.d. observations $X_1, X_2, \ldots, X_n$ from X with density function f_X unknown. A well-known and widely used nonparametric estimator of f_X is the Parzen-Rosenblatt kernel density estimator (Parzen [23] and Rosenblatt [26])

$$\tilde{f}_X(x) := \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{X_i - x}{h_n}\Big), \tag{5.1}$$

where K and $h = h_n > 0$ are respectively a given symmetric probability density function and a bandwidth parameter.

When the support of f_X is bounded from below or above, then it is well known that the classical estimator $\tilde{f}_X(x)$ is not a consistent estimator at the boundary point. This is illustrated in Fig. 5.1 for exponential density.

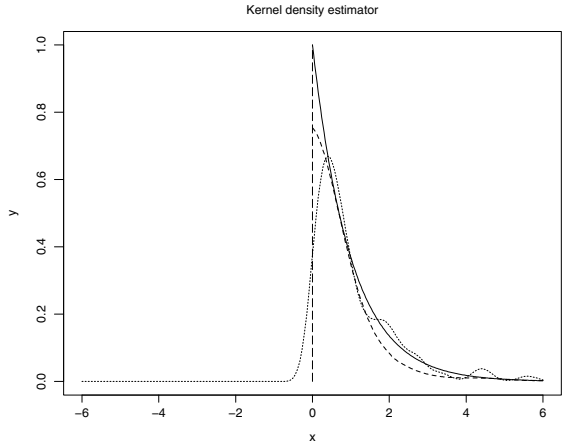

Figure 5.1. Kernel density estimator (dotted line) for exponential density, for $n = 200$. Dashed curve: estimator obtained by applying the reflection method.

Several types of remedies have been proposed in this case. There are reflection-type methods proposed by Schuster [29], transformation methods such as those discussed in Marron and Ruppert [19] and boundary-kernel type methods as discussed in, for example, Jones and Foster [17]. In Fig. 5.1 we plot (as a dashed curve) a kernel density estimate using the reflection method by Schuster [29]. For this method one either needs to know the support of the density, or to estimate it in the first step. This is in contrast with the method proposed below, where estimation is done in one single step.

The aim of this application part is to show that the method of Section 3 can easily be used to produce a consistent estimator for f_X when this function has a discontinuity at a boundary point. Moreover, and more importantly, the method can also be applied when having a density that possibly shows some discontinuities in the interior region as well. Fig. 5.2 depicts such a density, namely $f_X(x) = 0.5 \exp(x) I\{x < 0\} + \exp(-2x) I\{x \geqslant 0\}$.

The problem of density estimation is first transformed into a regression problem. Consider an interval $[a, b]$ that contains all observations $X_1, X_2, \ldots, X_n$. We partition this interval into N subintervals $\{I_k;\, k = 1, \ldots, N\}$ of equal length $\delta = (b - a)/N$.

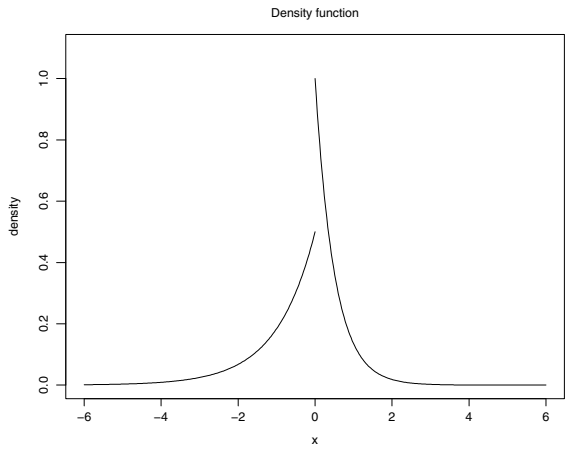

Figure 5.2. A density with a jump point at 0 of size 0.5.

For each of the intervals I_k we take

$$x_k = a + \frac{b-a}{N}\Big(k - \frac{1}{2}\Big) \quad \text{center of } I_k, \tag{5.2}$$
$$y_k = \frac{1}{n\delta}\sum_{i=1}^{n} I\{x_k - \delta/2 \leqslant X_i \leqslant x_k + \delta/2\},$$

where y_k is the proportion of the data falling in the interval I_k divided by δ. We have that $n\delta y_k$ is binomially distributed with parameters n and $p_k = \int_{I_k} f_X(x)\,\mathrm{d}x$. Therefore

$$E(y_k) = \frac{1}{n\delta}\,np_k = \frac{1}{\delta}\int_{I_k} f_X(x)\,\mathrm{d}x = \frac{1}{\delta}\int_{x_k-\delta/2}^{x_k+\delta/2} f_X(x)\,\mathrm{d}x \approx f_X(x_k),$$
$$\mathrm{Var}(y_k) = \frac{1}{n^2\delta^2}np_k(1-p_k) = \frac{1}{n\delta}\frac{p_k}{\delta}(1-p_k) \approx \frac{1}{n\delta}f_X(x_k).$$

Hence, estimation of f_X can be viewed as a nonparametric heteroscedastic regression problem with $m(x) = f_X(x)$ and $\sigma^2(x) \approx (1/n\delta)f_X(x)$, with available data $\{(x_k, y_k);\, k = 1, \ldots, N\}$.

An important remark here is that if f_X has a jump discontinuity, then this will also be the case with the variance function. This is an extra complication which one cannot ignore. Strictly speaking we have no theoretical justification for the method (yet). A possibility is to use Anscombe's variance stabilizing transformation (see Anscombe [1]) which replaces y_k by

$$y_k^* = \sqrt{2n\delta y_k + \frac{3}{8}},$$

and then, based on the data (x_k, y_k^*), one estimates $g_X(x) \approx 2\sqrt{f_X(x)n\delta + \frac{3}{8}}$. Taking the inverse then leads to an estimate for f_X. This rather naive approach seems to lead to an estimator with a satisfactory finite-sample behaviour. See also Lambert [18]. Further research is needed to establish the impact of the variance transformation on the heteroscedasticy of the problem and on the estimation procedure.

As an illustration we apply this method to the density in Fig. 5.2. We simulated 200 samples of size $n = 500$ from this density, and applied the above method with the Epanechnikov kernel taking $\delta = 0.556$. Depicted in Fig. 5.3 is the mean estimated curve (from the 200 samples) and the 5th and 95th percentile curves (based on the sample Mean Integrated Squared Error performance). For showing more details we present a zoom-in of the results around the jump area. We clearly see that the proposed estimator improves upon the conventional density estimator.

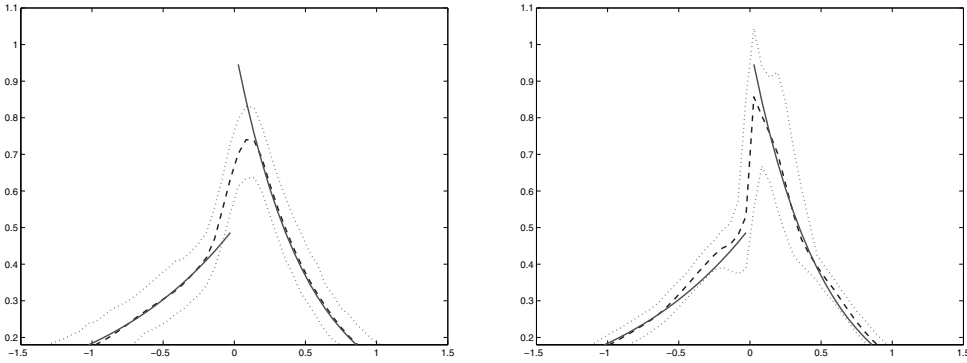

Figure 5.3. Simulation results for the density in Fig. 5.2 for $n = 500$. Left: Conventional kernel density estimator; right: Proposed estimator. Mean curve (dashed curve), 5th and 95th percentile curves (dashed curves) from 100 samples.

The results on the exponential density in Fig. 5.1 are shown in Fig. 5.4.

5.2. Estimating non-smooth conditional mean and variance function

We now return to the regression model in (2.1). Sections 2, 3 and 4 focused on the estimation of the mean regression function $m(\cdot)$ in the smooth and the non-smooth case, respectively. Often the interest is not only in estimation of the conditional mean function but also of the conditional variance function, i.e. in $\sigma^2(x) = \mathrm{Var}(Y|X = x)$. There is quite numerous literature on nonparametric estimation of smooth variance functions. See, for example, Gasser et al. [8], Hall et.al. [14], Müller and Stadtmüller [21], Ruppert et al. [28] and Fan and Yao [6], among others. All these papers however assume that both the conditional mean and variance function are smooth functions. It is of interest to look at the situation where possibly $m(\cdot)$ and/or $\sigma^2(\cdot)$ could show irregularities, say jump points. Testing for breakpoints simultaneously in m and σ^2 has been the subject of the recent work by Gao at al. [7]. There

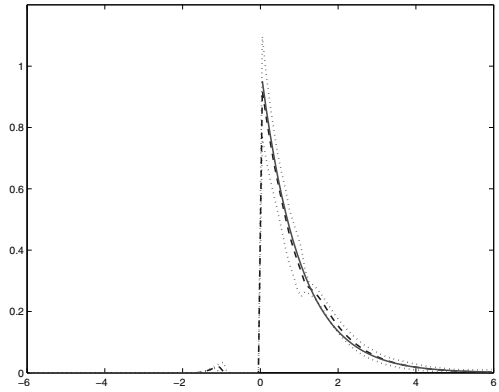

Figure 5.4. Simulation results for the exponential density in Fig. 5.1 for $n = 200$: proposed estimator. Mean curve (dashed curve), 5th and 95th percentile curves (dashed curves) from 100 samples.

are many papers concerning testing procedures for testing for continuity of $m(\cdot)$. See for example Gijbels and Goderniaux [9], [10] among others, and a recent paper by Antoch et al. [2] where local linear estimation is used for testing for continuity of $m(\cdot)$ allowing also for a possibly non-smooth conditional variance function.

When $m(\cdot)$ and $\sigma^2(\cdot)$ both might have jump points, a possible approach would be to use the estimation procedure exposed in Section 3 together with the variance estimation procedure of Fan and Yao [6]. This would then result in using the estimator in (3.3) to get an estimator $\hat{m}(\cdot)$ and then to use the residuals

$$\hat{r}_i = (Y_i - \hat{m}(X_i))^2$$

in the second step proceeding with local linear fitting of these pseudo-observations. This is the approach followed by Casas and Gijbels [3].

References

[1] *F. J. Anscombe*: The transformation of Poisson, binomial and negative-binomial data. Biometrika *35* (1948), 246–254.

[2] *J. Antoch, G. Grégoire, M. Hušková*: Tests for continuity of regression function. J. Stat. Plann. Inference *137* (2007), 753–777.

[3] *I. Casas, I. Gijbels*: Estimation of smooth and non-smooth variance functions. Manuscript (2007).

[4] *L. Desmet, I. Gijbels*: Peak preserving regression using local linear fitting. Submitted (2007).

[5] *J. Fan, I. Gijbels*: Local Polynomial Modelling and its Applications. Chapman & Hall, London, 1996.

[6] *J. Fan, Q. Yao*: Efficient estimation of conditional variance functions in stochastic regression. Biometrika *85* (1998), 645–660.

[7] *J. Gao, I. Gijbels, S. Van Bellegem*: Nonparametric simultaneous testing for structural breaks. J. Econom. Special Issue on Specification Testing *143* (2008), 123–142.

[8] *T. Gasser, L. Sroka, C. Jennen-Steinmetz*: Residual variance and residual pattern in nonlinear regression. Biometrika *73* (1986), 625–633.
[9] *I. Gijbels, A.-C. Goderniaux*: Bandwidth selection for change point estimation in nonparametric regression. Technometrics *46* (2004), 76–86.
[10] *I. Gijbels, A.-C. Goderniaux*: Bootstrap test for change-points in nonparametric regression. J. Nonparametric Stat. *16* (2004), 591–611.
[11] *I. Gijbels, A. Lambert, P. Qiu*: Edge-preserving image denoising and estimation of discontinuous surfaces. IEEE Trans. Pattern Anal. Mach. Intell. *28* (2006), 1075–1087.
[12] *I. Gijbels, A. Lambert, P. Qiu*: Jump-preserving regression and smoothing using local linear fitting: A compromise. Ann. Inst. Stat. Math. *59* (2007), 235–272.
[13] *G. Grégoire, Z. Hamrouni*: Change-point estimation by local linear smoothing. J. Multivariate Anal. *83* (2002), 56–83.
[14] *P. Hall, J. W. Kay, D. M. Titterington*: Asymptotically optimal difference-based estimation of variance in nonparametric regression. Biometrika *77* (1990), 521–528.
[15] *P. Hall, D. M. Titterington*: Edge preserving and peak-preserving smoothing. Technometrics *34* (1992), 429–440.
[16] *Z. Hamrouni*: Inférence statistique par lissage linéaire local pour une fonction de régression présentant des discontinuités. Doctoral Dissertation. Université de Joseph Fourier, Grenoble, 1999.
[17] *M. C. Jones, P. Foster*: A simple nonnegative boundary correction method for kernel density estimation. Stat. Sin. *6* (1996), 1005–1013.
[18] *A. Lambert*: Nonparametric estimation of discontinuous curves and surfaces. PhD. Dissertation. Université catholique de Louvain, Louvain-la-Neuve, 2005.
[19] *J. S. Marron, D. Ruppert*: Transformations to reduce boundary bias in kernel density estimation. J. R. Stat. Soc., Ser. B *56* (1994), 653–671.
[20] *J. A. McDonald, A. B. Owen*: Smoothing with split linear fits. Technometrics *28* (1986), 195–208.
[21] *H.-G. Müller, U. Stadtmüller*: On variance function estimation with quadratic forms. J. Stat. Plann. Inference *35* (1993), 213–231.
[22] *E. A. Nadaraya*: On estimating regression. Theory Probab. Appl. *9* (1964), 141–142.
[23] *E. Parzen*: On estimation of a probability density function and mode. Ann. Math. Stat. *33* (1962), 1065–1076.
[24] *P. Qiu*: A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. J. Nonparametric Stat. *15* (2003), 437–453.
[25] *P. Qiu*: Image Processing and Jump Regression Analysis. J. Wiley & Sons, Hoboken, 2005.
[26] *M. Rosenblatt*: Remarks on some nonparametric estimates of a density function. Ann. Inst. Math. Stat. *27* (1956), 832–837.
[27] *D. Ruppert, M. P. Wand*: Multivariate locally weighted least squares regression. Ann. Stat. *22* (1994), 1346–1370.
[28] *D. Ruppert, M. P. Wand, U. Holst, O. Hössjer*: Local polynomial variance-function estimation. Technometrics *39* (1997), 262–273.
[29] *E. F. Schuster*: Incorporating support constraints into nonparametric estimators of densities. Commun. Stat., Theory Methods *14* (1985), 1123–1136.
[30] *G. S. Watson*: Smooth regression analysis. Sankhya, Ser. A *26* (1964), 359–372.

Author's address: *I. Gijbels*, Katholieke Universiteit Leuven, Department of Mathematics and Leuven Statistics Research Centre (LStat), Celestijnenlaan 200B, KULeuven Postbox 2400, B-3001 Leuven (Heverlee), Belgium, e-mail: `irene.gijbels@wis.kuleuven.be`.