A simple nonparametric estimator of a monotone regression function

Holger Dette , Natalie Neumeyer, Kay F. Pilz
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@ruhr-uni-bochum.de

September 18, 2003

Abstract

In this paper a new method for monotone estimation of a regression function is proposed. The estimator is obtained by the combination of a density and a regression estimate and is appealing to users of conventional smoothing methods as kernel estimators, local polynomials, series estimators or smoothing splines. The main idea of the new approach is to construct a density estimate from the estimated values $\hat{m}(i/N)$ $(i = 1, \ldots, N)$ of the regression function to use these "data" for the calculation of an estimate of the inverse of the regression function. The final estimate is then obtained by a numerical inversion. Compared to the conventially used techniques for monotone estimation the new method is computationally more efficient, because it does not require constrained optimization techniques for the calculation of the estimate. We prove asymptotic normality of the new estimate and compare the asymptotic properties with the unconstrained estimate. In particular it is shown that for kernel estimates or local polynomials the monotone estimate is first order asymptotically equivalent to the unconstrained estimate. We also illustrate the performance of the new procedure by means of a simulation study.

AMS Subject Classification: 62G05, 62G20
Keywords and Phrases: isotonic regression, order restricted inference, Nadaraya-Watson estimator, local linear regression

1 Introduction

Smoothing as a means of modeling nonlinear structure in data has become increasingly popular in numerous applications. However, in many cases monotone estimates of the regression function are required, because physical considerations suggest that the response is a monotone function of the explanatory variable. There exists a vast amount of literature on the problem of estimating a regression function m which is believed to be monotone. Brunk (1955) proposed a modified maximum likelihood method to construct an estimate of a monotone regression function. Because this estimate is not smooth in general Mukerjee (1988) modified this method to obtain a monotone estimate with properties similar to those of nonparametric regression estimators. Mammen (1991) investigated the order of isotonisating and smoothing in a two step procedure, where one step consists in the construction of an isotonic estimate and the other in the construction of a classical kernel estimate [see also Cheng and Lin (1981), Wright (1982) and Friedman and Tibshirani (1984) for similar procedures]. Monotone nonparametric regression estimators based on constrained spline smoothing have been proposed by Ramsay (1988), Kelly and Rice (1990), Mammen and Thomas-Agnam (1999), while Mammen, Marron, Turlach and Wand (2001) suggested projection - based techniques for constrained smoothing. Recently, Hall and Huang (2001) proposed a new method for monotonizing a general kernel type estimator, which modifies the weights in a kernel estimator such that the modified function is monotone. This approach is particular appealing to users of conventional kernel methods.

In the present paper we propose an alternative construction of monotone regression functions. Similarly as in Hall and Huang (2001) our work is motivated by the search of a monotone estimate, which enjoys the same level of smoothness as its unconstrained counterpart and is additionally applicable to general smoothing methods. In contrast to the procedure proposed by the lastnamed authors (which is only able to monotonize kernel type estimators) the method suggested in this paper applies to arbitrary regression estimates and is computationally more efficient because it does not require a constrained optimization. Our approach constructs a density estimate from the estimated regression function and uses this additional smoothing step to obtain a monotone estimate of the inverse regression function. The monotone regression estimate is finally obtained by reflecting this function at the line $y = x$.

The method can easily be motivated by considering an i.i.d. sample of uniformly distributed random variables, say $U_1, \dots, U_N \sim \mathcal{U}([0,1])$. If m is a strictly increasing function on the interval $[0,1]$ with positive derivative, K_d is a kernel function and h_d a bandwidth, then

$$\frac{1}{Nh_d}\sum_{i=1}^{N} K_d\Big(\frac{m(U_i)-u}{h_d}\Big) \tag{1.1}$$

is the classical kernel estimate of the density of the random variable $m(U_1)$, that is

$$(m^{-1})'(u) I_{[m(0),m(1)]}(u).$$

Consequently

$$\frac{1}{Nh_d}\int_{-\infty}^{t}\sum_{i=1}^{N} K_d\Big(\frac{m(U_i)-u}{h_d}\Big)du \tag{1.2}$$

is a consistent estimate of the function m^{-1} at the point t. In the context of nonparametric regression $m(X) = E[Y \mid X]$ is the regression of Y with respect to X and the function m can be estimated by any standard method (kernel type, local polynomial, series or spline estimator), which yields an estimate of the inverse m^{-1} of the strictly increasing function m. The corresponding estimate of m is finally obtained by inversion of this estimate.

The estimate is carefully described in Section 2, where we also discuss some of its main properties as a monotone approximation of a given function. In Section 3 we study some of the statistical properties of the new estimate and prove asymptotic normality of the estimates for m^{-1} and m if kernel type or local polynomial estimators are used for the preliminary estimation of the regression function. In particular we show that for local linear estimators the new estimate is asymptotically first order equivalent to the unconstrained estimate. The choice of the smoothing parameters is also investigated from an asymptotic point of view. In Section 4 we discuss the finite sample properties of the new estimator by means of a simulation study. Finally, some of the technical details are given in the appendix. The main advantages of the new procedure are the computational simplicity (because it does not require any constrained optimization techniques) and the asymptotic equivalence to the unconstrained estimate. This makes the new method attractive to users of conventional kernel methods.

2 Monotone smoothing by inversion

Consider the nonparametric regression model

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $\{(X_i, Y_i)\}_{i=1}^n$ is a bivariate sample of i.i.d. observations such that X_i has a positive two times continuously differentiable density f with compact support, say $[0, 1]$. The variance function $\sigma : [0, 1] \to \mathbb{R}^+$ and the regression function $m : [0, 1] \to \mathbb{R}$ are assumed to be continuous and two times continuously differentiable, respectively. If there is evidence that the regression function m is (strictly) increasing we define for $N \in \mathbb{N}$

$$\hat{m}_I^{-1}(t) := \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{t} K_d\Big(\frac{\hat{m}(\frac{i}{N}) - u}{h_d}\Big) du \tag{2.2}$$

as an estimate of $m^{-1}(t)$, where

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_r\left(\frac{X_i - x}{h_r}\right) Y_i}{\sum_{i=1}^n K_r\left(\frac{X_i - x}{h_r}\right)} \tag{2.3}$$

is the classical Nadaraya-Watson estimate [see Nadaraya (1964) or Watson (1964)], K_d and K_r denote symmetric kernels with compact support, say $[-1, 1]$, existing second moment and h_d, h_r are the corresponding bandwidths converging to 0 with increasing sample size n. We assume that K_d is two times continuously differentiable on its support and that the kernel K_r has been appropriately modified at the boundary [see Müller (1984)]. For the sake of transparency we restrict ourselves to the Nadaraya-Watson estimate, but it is notable that all results of the paper remain valid (subject to an appropriate modification of constants) for other types of kernel estimators as the Gasser-Müller estimator [see Gasser and Müller (1979)] or local polynomials [see Fan and Gijbels (1996) or Wand and Jones (1995)].

Note that the indices "r" and "d" correspond to the phrase "regression" and "density" because we combine a regression with a density estimate to define the estimator in (2.2). Comparing this estimate with the motivation in equation (1.2) we see that the uniformly distributed random variables have been replaced by an equidistant design. Note that it is not necessary (and in many case not desirable) that the number N of design points coincides with the sample size n. Finally, we note that the estimate $\hat{m}_I^{-1}$ is isotonic if the kernel K_d is positive which will be assumed throughout this paper. In this case an isotonic estimate of the regression function $\hat{m}_I$ is simply obtained by reflection of the function $\hat{m}_I^{-1}$ at the line $y = x$. Note that the estimator $\hat{m}_I^{-1}$ is

equal to 1 and 0 if $t > \max_{i=1}^N \hat{m}(\frac{i}{N}) + h_d$ and $t < \min_{i=1}^N \hat{m}(\frac{i}{N}) - h_d$, respectively, and only for $t \in [\min_{i=1}^N \hat{m}(\frac{i}{N}), \max_{i=1}^N \hat{m}(\frac{i}{N})]$ the inverse of the function $\hat{m}_I$ is calculated.
If the regression function m is supposed to be strictly decreasing the estimate can easily be modified as

$$\hat{m}_A^{-1}(t) := \frac{1}{Nh_d} \sum_{i=1}^N \int_t^\infty K_d\Big(\frac{\hat{m}(\frac{i}{N}) - u}{h_d}\Big) du \tag{2.4}$$

and the antitonic estimate is obtained by the inversion of this function. Throughout this paper we restrict ourselves to the case of an isotonic regression function and the estimate (2.2). Corresponding results for the antitonic case are very similar and obtained by the same reasoning. Because $\hat{m}$ converges uniformly to the unknown regression function m, it is heuristically clear that the estimate $\hat{m}_I^{-1}$ is in some sense close to the function

$$m_N^{-1}(t) \;=\; \frac{1}{Nh_d} \int_{-\infty}^t \sum_{i=1}^N K_d\Big(\frac{m(\frac{i}{N}) - u}{h_d}\Big) du, \tag{2.5}$$

which is an approximation of the integral

$$\frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\Big(\frac{m(x) - u}{h_d}\Big) du dx = \int_0^1 I\{m(x) \le t\} dx + o(1) \tag{2.6}$$

(note that the kernel K_d has compact support). In other words, the statistic $\hat{m}_I^{-1}(t)$ is a consistent estimate of the quantity $\int_0^1 I\{m(x) \le t\}dx$ and this property does not depend on the particular estimate $\hat{m}$ used in the regression step provided that $\hat{m}$ is a uniformly consistent nonparametric estimate for the regression function m. The leading term on the right hand side of equation (2.6) is equal to $m^{-1}(t) = \inf\{u \mid m(u) > t\}$ if the regression function is increasing and the following lemma gives the precise order of this approximation if the regression function is strictly increasing.

Lemma 2.1. *If the regression function is strictly increasing and the assumptions stated at the beginning of this section are satisfied then we have for any $t \in (m(0), m(1))$ with $m'(m^{-1}(t)) > 0$*

$$m_N^{-1}(t) = m^{-1}(t) + \kappa_2(K_d) h_d^2 (m^{-1})''(t) + o(h_d^2) + O\Big(\frac{1}{Nh_d}\Big) \;, \tag{2.7}$$

where the constant $\kappa_2(K)$ is given by

$$\kappa_2(K) = \frac{1}{2} \int_{-1}^1 v^2 K(v) dv. \tag{2.8}$$

It is easy to see that the functions m_N^{-1} and $\hat{m}_I^{-1}$ are strictly increasing, if

$$\begin{aligned} \max_{i=1}^N \min_{j=1}^N |m(\frac{i}{N}) - m(\frac{j}{N})| \;&<\; 2h_d \\ \max_{i=1}^N \min_{j=1}^N |\hat{m}(\frac{i}{N}) - \hat{m}(\frac{j}{N})| \;&<\; 2h_d, \end{aligned} \tag{2.9}$$

respectively. If the number of design points N (which is not necessarily equal to the sample size n) is chosen sufficiently large, the first inequality is satisfied because of the continuity of the regression function m. Similarly, if the sample size $n \to \infty$, the second inequality holds for any estimate $\hat{m}$ almost surely, which is uniformly strong consistent with a rate $o(h_d)$. Throughout this paper m_N denotes the inverse of the function m_N^{-1}. Because m_N^{-1} is expected to be an approximation of the function m^{-1}, it is intuitively clear that the inverse m_N of m_N^{-1} is an approximation of the function m. The following lemma makes this statement precise and is proved in the appendix.

Lemma 2.2. *If the regression function m is strictly increasing and the assumptions stated at the beginning of this section are satisfied, then we have for any $t \in (0,1)$ with $m'(t) > 0$*

$$m_N(t) = m(t) + \kappa_2(K_d)h_d^2 \frac{m''(t)}{(m'(t))^2} + o(h_d^2) + O\Big(\frac{1}{Nh_d}\Big).$$

If m is not necessarily increasing, the function

$$g : t \to \int_0^1 I\{m(x) \le t\}dx \tag{2.10}$$

or its approximation

$$g_{h_d} : t \to \int_0^1 \int_{-\infty}^t \frac{1}{h_d} K_d\Big(\frac{m(x)-u}{h_d}\Big)dudx \tag{2.11}$$

is still well defined and nondecreasing. Note that the function g is not necessarily differentiable [see also the examples presented below] and g_{h_d} can be considered as a smooth version of g which converges to g if $h_d \to 0$. The (generalized) inverse $g_{h_d}^{-1}$ can be considered as an approximation of the function m by a nondecreasing smooth function. The properties of this function are important for the behaviour of our estimate and will be briefly described in the following. For the sake of brevity we restrict ourselves to the function g and mention that the properties of g_{h_d} are similar. If for a fixed t_0 the set $m^{-1}(\{t_0\}) = \{x_0\}$ is a singleton and $m'(x_0) > 0$, then, obviously, $g(t_0) = x_0$ and $g^{-1}(x_0) = m(x_0)$. Now let $x_0 \in [0,1]$ denote the infimum of all points such that there exists a t_0 with this property (note that the case $x_0 = 0$ is not excluded) and define $x_1 \ge 0$ as the maximal point such that this property is satisfied for all $x \in (x_0, x_1)$ with corresponding value $t_1 = m(x_1)$. In this case we have for all $t = m(x) \in [t_0, t_1]$

$$g(t) = x_0 + (x - x_0) = x$$

which proves $g^{-1}(x) = m(x)$ for all $x \in [x_0, x_1]$. If $x_1 < 1$, the function m is decreasing in a neighbourhood $(x_1, x_1 + \varepsilon)$ and there may exist a second interval, say (x_2, x_3), such that m is strictly increasing on (x_2, x_3) and such that for all $t \in (m(x_2), m(x_3))$ the set $m^{-1}(\{t\})$ is a singleton. For this interval the same argument shows $g^{-1}(x) = m(x)$ for all $x \in [x_2, x_3]$. The repetition of this argument shows that on any interval $[a, b]$, where m is strictly increasing such that $m^{-1}(\{a\})$ and $m^{-1}(\{b\})$ are singletons the inverse of the function g coincides with the regression function m. We will illustrate this behaviour in the following examples.

Example 2.3. Consider the function

$$m(x) = \frac{11}{3}x - 8x^2 + \frac{16}{3}x^3$$

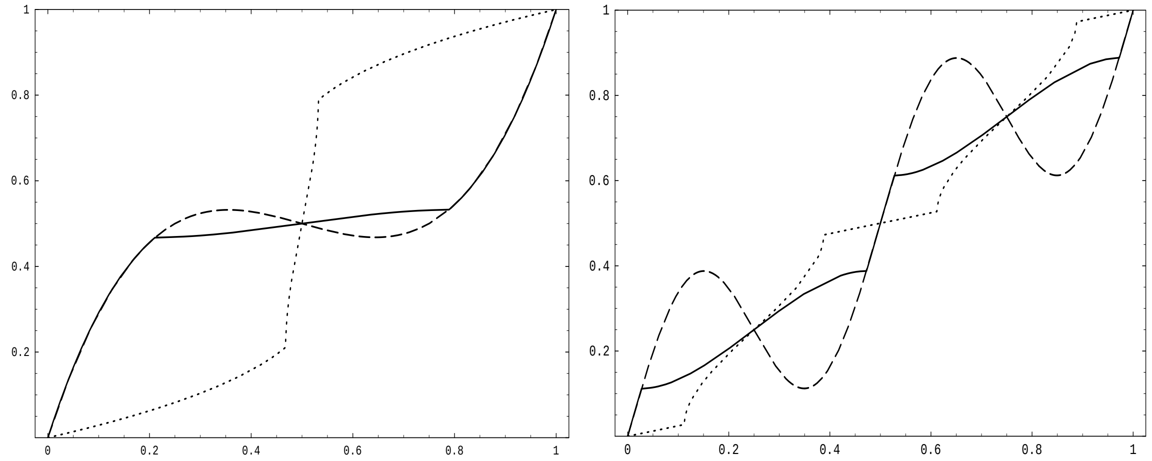

Figure 2.1: *Approximation of a non-monotone regression function m (dashed line) by its monotone approximation* g^{-1} *(solid line); left panel:* $m(x) = \frac{11}{3}x - 8x^2 + \frac{16}{3}x^3$*; right panel:* $m(x) = x + \frac{1}{4}\sin(4\pi x)$*. The figures show also the function g defined by (2.10) (dotted line).*

which is strictly increasing on the intervals $[0,(6-\sqrt{3})/12], [(6+\sqrt{3})/12, 1]$. The functions m, g and g^{-1} are depicted in the left panel of Figure 2.1, where the function g^{-1} coincides with the function m whenever $m^{-1}(\{t\})$ is a singleton. Note that there are two points, where the function g is not differentiable. Our second example illustrates the approximation of the oscillating function

$$m(x) = x + \frac{1}{4}\sin(4\pi x)$$

by the monotone increasing function g^{-1} [see the right panel of Figure 2.1] while the convex case,

$$m(x) = \frac{16}{9}(x - \frac{1}{4})^2,$$

and concave case,

$$m(x) = 1 - 4(x - \frac{1}{2})^2,$$

are shown in the left and right panel of Figure 2.2, respectively.

3 Main results – asymptotic behaviour

In this section we investigate some of the asymptotic properties of the estimates $\hat{m}_I^{-1}$ and $\hat{m}_I$. It turns out that both estimates are (appropriately centered) asymptotically normal distributed, where the asymptotic variance depends on the limit

$$\lim_{h_r \to 0, h_d \to 0} h_r/h_d =: c \in (0, \infty]$$

of the ratio of the smoothing parameters. In the case $c = \infty$ we show that the new monotone estimate $\hat{m}_I$ is first order asymptotically equivalent to the unconstrained estimate $\hat{m}$, if the Nadaraya-Watson estimator with a uniform design or a local linear estimator is used for the estimation of the regression function.

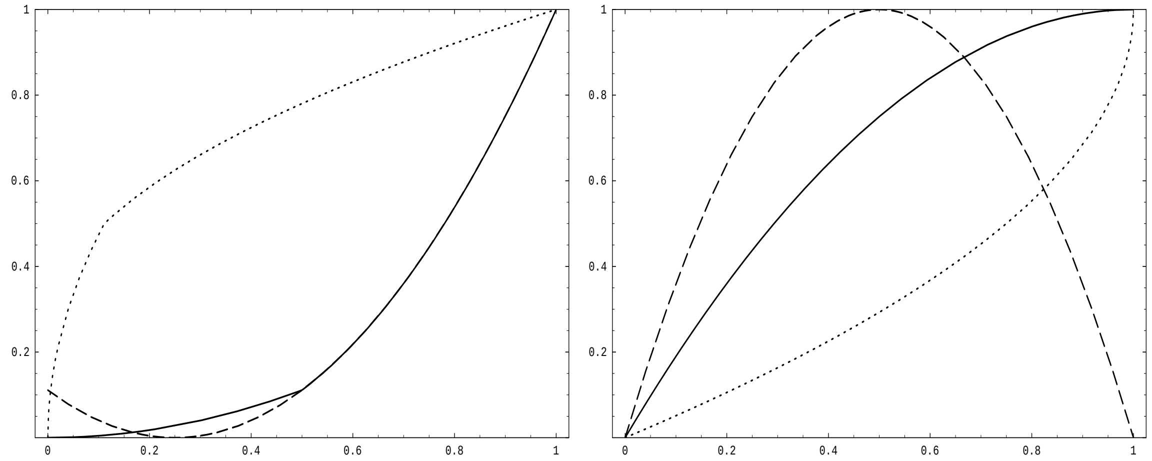

Figure 2.2: *Approximation of a non-monotone regression function m (dashed line) by its monotone approximation g^{-1}(solid line); left panel: $m(x) = \frac{16}{9}(x - \frac{1}{4})^2$; right panel: $m(x) = 1 - 4(x - \frac{1}{2})^2$. The figures show also the function g defined by (2.10) (dotted line).*

3.1 Asymptotic normality

We assume that the smoothness conditions regarding the density, variance and regression function stated at the beginning of Section 2 are satisfied. For the bandwidths h_r and h_d in the regression and density estimate we require $h_r \to 0, h_d \to 0, nh_r \to \infty, nh_d \to \infty$ and additionally

$$(3.1) \qquad nh_r^5 = O(1), \qquad n = O(N),$$

$$(3.2) \qquad \frac{\log h_r^{-1}}{nh_r h_d^3} = o(1).$$

Note that for the "optimal" rate in regression estimation $h_r = \gamma n^{-1/5}$ the latter assumption reduces to $h_d n^{4/15}/(\log n)^{1/3} \to \infty$. Our first result shows the asymptotic normality of the estimate $\hat{m}_I^{-1}$ and only requires the estimate

$$(3.3) \qquad \frac{1}{nh_r h_d^2} = o(1)$$

which gives $h_d n^{2/5} \to \infty$ in the case where the optimal bandwidth $h_r = \gamma n^{-1/5}$ is used for the estimation of the regression function.

Theorem 3.1. *If the assumptions (3.1) and (3.3) are satisfied, $\lim_{n\to\infty} \frac{h_r}{h_d} = c \in (0, \infty)$ exists and m is strictly increasing, then it follows that for all $t \in (m(0), m(1))$ with $m'(m^{-1}(t)) > 0$*

$$(3.4) \qquad \sqrt{nh_d}\Big(\hat{m}_I^{-1}(t) - m_N^{-1}(t) + \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{fm'}\Big)(m^{-1}(t))\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, g^2(t)),$$

where the constant $\kappa_2(K_r)$ is defined in (2.8) and the asymptotic variance is given by

$$(3.5) \qquad g^2(t) = \frac{\sigma^2(m^{-1}(t))}{m'(m^{-1}(t))f(m^{-1}(t))} \times \int\int\int K_d(w + cm'(m^{-1}(t))(v-u))K_d(w)K_r(u)K_r(v)dwdudv.$$

If $\lim_{n\to\infty} \frac{h_r}{h_d} = \infty$, *then we have for all* $t \in (m(0), m(1))$ *with* $m'(m^{-1}(t)) > 0$

$$\sqrt{nh_r}\Big(\hat{m}_I^{-1}(t) - m_N^{-1}(t) + \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{fm'}\Big)(m^{-1}(t))\Big) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \tilde{g}^2(t)), \tag{3.6}$$

where the asymptotic variance is given by

$$\tilde{g}^2(t) = \frac{\sigma^2(m^{-1}(t))}{\{m'(m^{-1}(t))\}^2 f(m^{-1}(t))} \int K_r^2(u)du. \tag{3.7}$$

Note that for sufficiently large n and N the functions $\hat{m}_I^{-1}$ and m_N^{-1} are strictly increasing independent of the monotonicity of the "true" regression function m [see the inequalities (2.9)]. The following result shows that the corresponding inverse functions $\hat{m}_I$ and m_N also satisfy an asymptotic normal law.

Theorem 3.2. *Assume that the assumptions of Theorem 3.1 are satisfied and let* $\hat{m}_I$ *and* m_N *denote the inverse functions of the functions* $\hat{m}_I^{-1}$ *and* m_N^{-1} *defined by (2.2) and (2.5), respectively. If* $\lim_{n\to\infty} \frac{h_r}{h_d} = c \in (0, \infty)$ *exists, then we have for every* $t \in (0, 1)$ *with* $m'(t) > 0$

$$\sqrt{nh_d}\Big(\hat{m}_I(t) - m_N(t) - \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{f}\Big)(t)\Big) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, s^2(t)),$$

where the asymptotic variance is given by

$$s^2(t) = \frac{\sigma^2(t)m'(t)}{f(t)} \int\int\int K_d(w + cm'(t)(v-u))K_d(w)K_r(u)K_r(v)dwdudv. \tag{3.8}$$

If $\lim_{n\to\infty} \frac{h_r}{h_d} = \infty$ *it follows for every* $t \in (0, 1)$ *with* $m'(t) > 0$

$$\sqrt{nh_r}\Big(\hat{m}_I(t) - m_N(t) - \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{f}\Big)(t)\Big) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \tilde{s}^2(t)), \tag{3.9}$$

where the asymptotic variance is given by

$$\tilde{s}^2(t) = \frac{\sigma^2(t)}{f(t)} \int K_r^2(u)du. \tag{3.10}$$

We conclude this section noting that the first assertion of Theorem 3.2 can be written as

$$\sqrt{nh_r}\Big(\hat{m}_I(t) - m_N(t) - \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{f}\Big)(t)\Big) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}\Big(0, cs^2(t)\Big)$$

and that a simple calculation shows

$$\lim_{c\to\infty} cs^2(t) = \tilde{s}^2(t). \tag{3.11}$$

Thus formally the second part of Theorem 3.2 could be identified from the first part using the relation (3.11). Moreover, observing Lemma 2.2 and the second part of Theorem 3.2 it follows that for a bandwidth h_d satisfying $h_d = o(h_r)$

$$\sqrt{nh_r}\Big(\hat{m}_I(t) - m(t) - \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{f}\Big)(t)\Big) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \tilde{s}^2(t)),$$

where $\tilde{s}^2(t)$ is defined in (3.10). In other words, if $h_d = o(h_r)$ the monotone estimator $\hat{m}_I$ exhibits the same first order asymptotic behaviour as the unconstrained estimate $\hat{m}$. A similar property was observed by Mammen (1991) for the L^2-projection of the Nadaraya-Watson estimate onto the space of all increasing functions. The choice of the appropriate bandwidths for the estimator $\hat{m}_I$ is discussed in more detail in the following paragraph.

3.2 Bandwidth selection

The choice of the two bandwidths is essential for the performance of the new smoothing procedure. While the bandwidth h_r for the regression estimate $\hat{m}$ can be chosen by standard methods [see e.g. Stone (1974), Härdle, Hall, Marron (1988) or Gasser, Kneip and Köhler (1991)], the choice of the bandwidth h_d in the second step of the density estimate is less clear. In the following we will discuss an efficient choice from an asymptotic point of view. For this recall that by Lemma 2.1 the approximation of the inverse function m^{-1} by m_N^{-1} is of order $O(h_d^2)$. This implies that the centering constant in Theorem 3.1 is given by

$$m^{-1}(t) + \Gamma(h_d, h_r) + o(h_r^2) + o(h_d^2), \tag{3.12}$$

where the function Γ is defined by

$$\Gamma(h_d, h_r) = \kappa_2(K_d)(m^{-1})''(t)h_d^2 - \kappa_2(K_r)\Big(\frac{m''f + 2m'f'}{fm'}\Big)(m^{-1}(t))h_r^2\ . \tag{3.13}$$

Note that in the case of a uniform density this term simplifies to

$$\Gamma(h_d, h_r) = \kappa_2(K_d)(m^{-1})''(t)h_d^2 - \kappa_2(K_r)\Big(\frac{m''}{m'}\Big)(m^{-1}(t))h_r^2 \tag{3.14}$$

and the same result is obtained if a local linear estimate is used for the estimation of the regression function [see Fan and Gijbels (1996)]. Similarly, the leading term of the bias of the estimate $\hat{m}_I$ is given by

$$\Gamma_I(h_d, h_r) = \kappa_2(K_d)\frac{m''(t)}{(m'(t))^2}h_d^2 + \kappa_2(K_r)m''(t)h_r^2, \tag{3.15}$$

if the design is uniform or a local linear estimate is used for the regression step. In the following we restrict ourselves to the case, where the Nadaraya-Watson estimate with a uniform design or a local linear estimator is used in the regression step. Other cases can be discussed exactly in the same way with an additional amount of notation. We choose the bandwidth

$$h_d = \gamma\, m'(m^{-1}(t))h_r, \tag{3.16}$$

for the estimate $\hat{m}_I^{-1}$ and

$$h_d = \gamma\, m'(t)h_r \tag{3.17}$$

for the estimate $\hat{m}_I$, for some constant $\gamma > 0$. The relevant information regarding the asymptotic normality of the monotone estimators with these bandwidths is summarized in the following corollary. The proof is a direct consequence of Theorem 3.1 and 3.2 and therefore omitted.

Corollary 3.3. *Assume that the assumptions of Theorem 3.1 are satisfied and that f is either the uniform density or the local linear estimate is used for the estimator $\hat{m}$ in (2.2).*

(1) If $t \in (m(0), m(1))$ satisfies $m'(m^{-1}(t)) > 0$ and the bandwidth h_d in the isotonic estimate (2.2) is chosen according to the rule (3.16), then

$$\sqrt{nh_r}\Big(\hat{m}_I^{-1}(t) + (\kappa_2(K_r) + \gamma^2\kappa_2(K_d))\Big(\frac{m''}{m'}\Big)(m^{-1}(t))h_r^2 - m^{-1}(t)\Big) \stackrel{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \frac{\sigma^2(m^{-1}(t))}{f(m^{-1}(t))\{m'(m^{-1}(t))\}^2}\,\mu_K^2(\gamma))$$

where the asymptotic variance is given by

$$\mu_K^2(\gamma) = \int\Big(\int K_r(u)K_r(u+\gamma w)du\Big)\Big(\int K_d(v)K_d(v+w)dv\Big)dw. \tag{3.18}$$

(2) If $t \in (0,1)$ satisfies $m'(t) > 0$ and the bandwidth h_d in the isotonic estimate (2.2) satisfies (3.17), then

$$\sqrt{nh_r}\Big(\hat{m}_I(t) - (\kappa_2(K_r) + \gamma^2\kappa_2(K_d))m''(t)h_r^2 - m(t)\Big) \stackrel{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \frac{\sigma^2(t)}{f(t)}\mu_K^2(\gamma))$$

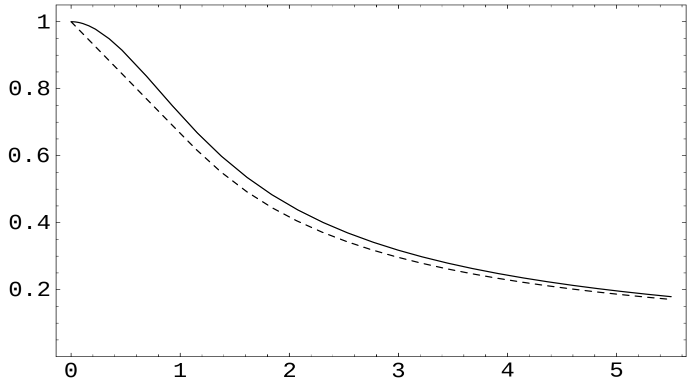

Figure 3.1: *The function μ_K^2 defined in (3.18) for the rectangular (dashed line) and Epanechnikov kernel (solid line), where $K_d = K_r$*

This result has some consequences from an asymptotic point of view. First note that it is easy to see that Corollary 3.3 also includes the case $\gamma = 0$ (corresponding to the second part of Theorem 3.1 and 3.2), if this is interpreted as

$$\lim_{h_r, h_d \to 0} h_r/h_d = \infty.$$

In this case the constant is given by $\mu_K^2(0) = \int K_r^2(u)du$ and coincides with the corresponding term in the unconstrained regression estimate. Similarly, the bias of the isotonic estimate $\hat{m}_I(t)$ is obtained as $\kappa_2(K_r)m''(t)h_r^2$. Thus with the choice $\gamma = 0$ (with the interpretation mentioned above) the monotone estimate exhibits the same (first) order asymptotic behaviour as the unconstrained regression estimate. Secondly, we consider the situation $\gamma > 0$, for which a simple application of Cauchy's, Jensen's inequality and Fubini's theorem shows that

$$\begin{aligned}
\mu_K^2(\gamma) &\le \Big\{\int\Big(\int K_r(u)K_r(u+\gamma w)du\Big)^2 dw\Big\}^{1/2}\Big\{\int\Big(\int K_d(v)K_d(v+w)dv\Big)^2 dw\Big\}^{1/2} \\
&\le \Big\{\int\Big(\int K_r(u)K_r^2(u+\gamma w)du\Big)dw\Big\}^{1/2}\Big\{\int\Big(\int K_d(v)K_d^2(v+w)dv\Big)dw\Big\}^{1/2} \\
&= \Big\{\frac{1}{\gamma}\int K_r^2(u)du\Big\}^{1/2}\Big\{\int K_d^2(v)dv\Big\}^{1/2} .
\end{aligned}$$

Consequently, observing the identitiy $\mu_K^2(0) = \int K_r^2(u)du$ and the first part of Corollary 3.3 it follows, that the estimator $\hat{m}_I$ with the choice (3.17), $\gamma \ge 1$ and $K_r = K_d$ has asymptotically a smaller variance but a larger bias than the monotone estimator obtained for the choice $\gamma = 0$, which is first order asymptotically equivalent to the unconstrained estimator. A proof of the inequality $\mu_K^2(\gamma) \le \mu_K^2(0)$ in the general case might be difficult. However, numerical results show that for the commonly used kernels the function μ_K^2 is decreasing with γ. More precisely, we did not find a kernel for which this function is not decreasing. A typical example is presented in Figure 3.1 for the case where K_d and K_r are chosen from the beta family.
In general the appropriate choice of the smoothing parameters is a more sophisticated task, even from an asymptotic point of view. The above results show that the variance is decreasing with large values of γ, while the converse holds true for the bias. Heuristically, the choice $\gamma = 0$ may have particular advantages if the standard error is small compared to the bias, while values as $\gamma = 0.5$ or $\gamma = 1$ may be appropriate for a small bias and larger standard errors. We will investigate these effects by means of a simulation study in the following section.
In the remaining part of this section we study the effect of the choice of γ in the rule (3.17) if the local optimal bandwidth

$$h_r = \Big(\frac{b(K_r)\sigma^2(t)}{4f(t)(m''(t))^2\kappa_2^2(K_r)n}\Big)^{\frac{1}{5}} \tag{3.19}$$

is used for the estimation of the regression function, where $b(K_r) = \int K_r^2(u)du$. A standard calculation shows that for this choice the first order approximation of the mean squared error is given by

$$h(\gamma) = \Big(\frac{b(K_r)\sigma^2(t)}{4nf(t)}\Big)^{4/5}(m''(t)\kappa_2(K_r))^{2/5}\Big\{\Big(1+\gamma^2\frac{\kappa_2(K_d)}{\kappa_2(K_r)}\Big)^2 + \frac{4}{b(K_r)}\mu_K^2(\gamma)\Big\} \tag{3.20}$$

where $\mu_K^2(\gamma)$ is defined in (3.18). The corresponding mean squared error for the unconstrained estimate is given by $h(0)$, which gives for the efficiency

$$e(\gamma) = \frac{h(\gamma)}{h(0)} = \frac{\Big(1+\gamma^2\frac{\kappa_2(K_d)}{\kappa_2(K_r)}\Big)^2 + \frac{4}{b(K_r)}\mu_K^2(\gamma)}{1+\frac{4}{b(K_r)}\mu_K^2(0)}. \tag{3.21}$$

In Figure 3.2 we display the function e for the cases, where $K_r = K_d$ is the Epanechnikov and rectangular kernel. We see that for the Epanechnikov kernel the optimal choice (minimizing $e(\gamma)$

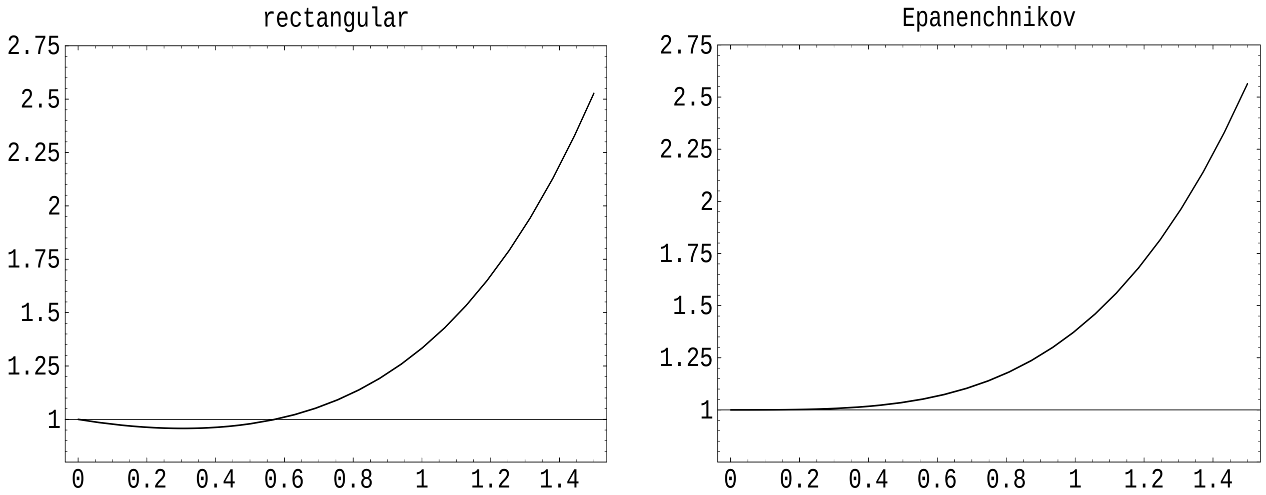

Figure 3.2: *The function e defined in (3.21) for the rectangular (left panel) and Epanechnikov kernel (right panel), where $K_d = K_r$.*

with respect to the parameter γ) is $\gamma = 0$, which corresponds to the case $\lim_{h_r, h_d \to 0} h_r/h_d = \infty$, while for the rectangular kernel the choice $\gamma \approx 0.3$ yields the smallest efficiency. In this case the bandwidths h_d and h_r should be chosen of the same order according to the rules (3.19) and (3.17). We investigated several other kernels (including the beta-family) and conclude that the situation displayed in the left panel of Figure 3.2 for the Epanechnikov kernel is rather typical. Except for the rectangular kernel, all kernels yield the same picture for the efficiency as displayed in Figure 3.2 for the Epanechnikov kernel. The situation is also similar in the case, where K_d not necessarily equals K_r. The efficiency is strictly increasing with respect to γ, except in the case, where a rectangular kernel is used for the preliminary estimation of the regression function. In this case the minimal efficiency is attained for some $\gamma > 0$ but the choice $\gamma = 0$ does not yield a loss of efficiency of more than 5%. These results indicate that from an asymptotic point of view the bandwidth h_d and h_r should not be of the same order, but bandwidths satisfying

$$\lim_{h_d, h_r \to 0} \frac{h_r}{h_d} = \infty$$

should be preferred for the monotone estimator. In general some care is necessary with these asymptotic arguments and we will illustrate the performance of the estimators for realistic sample sizes in the following section by means of a simulation study.

4 Finite sample properties

In this section we illustrate the behaviour of the new monotone estimator for finite sample sizes. We consider the nonparametric regression model (2.1) with a uniform design and the regression functions

$$m(x) = \frac{1}{2}(2x-1)^3 + \frac{1}{2}, \tag{4.1}$$

$$(4.2) \qquad m(x) = \sin\Big(\frac{\pi}{2}x\Big),$$

where the standard deviation of the errors is constant and given by $\sigma = 0.1$. For the regression estimate we use a local linear estimator with Epanechnikov kernel, where the bandwidth h_r is chosen as

$$(4.3) \qquad h_r = \Big(\frac{\hat{\sigma}^2}{n}\Big)^{1/5},$$

and $\hat{\sigma}^2$ denotes the estimator of Rice (1984) that is

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \Big(Y_{[i+1]} - Y_{[i]}\Big)^2 ,$$

where $Y_{[1]}, \dots, Y_{[n]}$ denote the observations ordered with respect to their corresponding X-values. For the density estimate we also use the Epanechnikov kernel with $N = 100$ design points, while the bandwidth is chosen as $h_d = h_r^3$ and $n = 100$ observation are used for the estimation of the regression function.

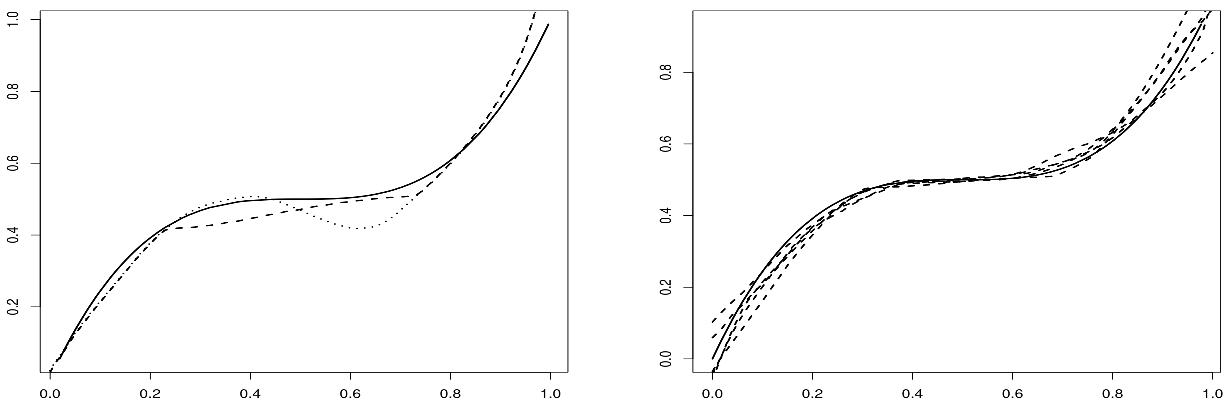

Figure 4.1: *The function* $m(x) = \frac{1}{2}(2x-1)^3 + \frac{1}{2}$ *(solid line) and its monotone estimates for* $n = 100$ *observations and* $\sigma = 0.1$. *Left panel: the monotone estimate (dashed line) and the local polynomial fit (dotted line) for a typical situation; right panel: five monotone estimates obtained from different simulations (dashed curves).*

In Figure 4.1 and 4.2 we show the estimate $\hat{m}_I$ based on a local linear estimator $\hat{m}$ in the first step of the estimation procedure for the model (4.1) and (4.2), respectively. The left parts of the figures show the "true" function, the local linear fit $\hat{m}$ and the corresponding monotone estimate $\hat{m}_I$ for a typical situation, while the right parts of the figures show the true function and five monotone estimates obtained from different simulations.

In the second part of our simulation study we investigate the choice of the bandwidth h_d for finite sample sizes in more detail. The bandwidth for the local linear estimate is given by (4.3) while

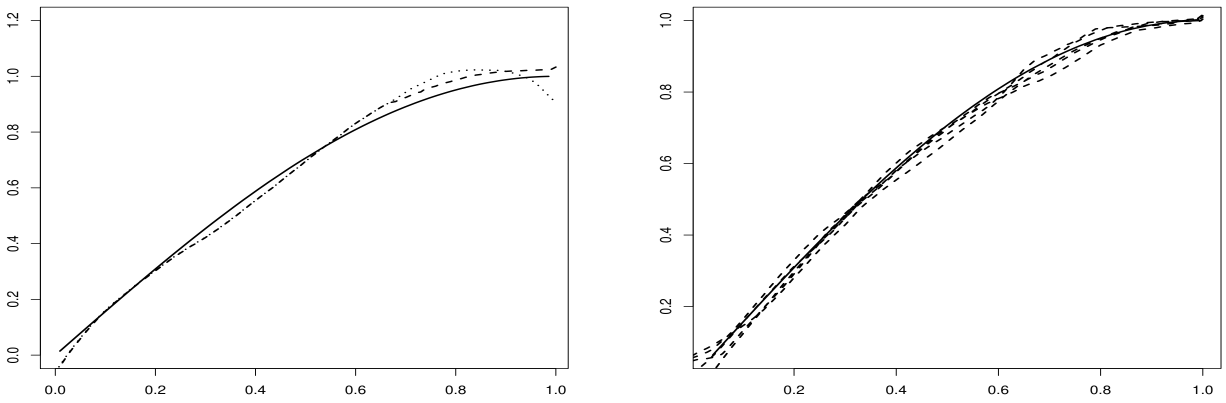

Figure 4.2: *The function* $m(x) = \sin(\frac{\pi}{2}x)$ *(solid line) and its monotone estimates for* $n = 100$ *observations and* $\sigma = 0.1$*. Left panel: the monotone estimate (dashed line) and the local polynomial fit (dotted line) for a typical situation; right panel: five monotone estimates obtained from different simulations (dashed curves).*

two choices of h_d are under consideration, that is $h_d = \frac{1}{2}h_r$ and $h_d = h_r^3$. In Table 4.1 and Table 4.2 we show the simulated mean squared error (MSE), bias and variance of the monotone estimates for various regression functions and design points, where the standard deviation is $\sigma = 0.1$ (Table 4.1) and $\sigma = 1$ (Table 4.2). Four cases for the regression function are considered in our study, namely

$$\begin{aligned}
(1) \quad m(x) &= x \\
(2) \quad m(x) &= x^2 \\
(3) \quad m(x) &= \sin(\frac{\pi}{2}x) \\
(4) \quad m(x) &= \frac{1}{2} + \frac{1}{2}(2x-1)^3 \ .
\end{aligned}$$

The values of the unconstrained local linear estimate are given in brackets. In all cases we observe a smaller variance and a larger bias of the constrained estimate $\hat{m}_I$ as expected from the asymptotic theory. The effect of the choice of the bandwith on the MSE depends on the size of the variance and the size of $m''(t)$. In the case $\sigma = 0.1$ the improvement with respect to the variance by choosing $\gamma = 1/2$ in the constant $\mu_K^2(\frac{1}{2})$ is partially compensated by the factor $\sigma^2 = 0.01$. As a consequence the MSE obtained for the choice $h_d = h_r^3$ is smaller than the MSE obtained by the choice $h_d = \frac{1}{2}h_r$ in cases where $|m''(t)|$ is large. For regression functions with a small value of $|m''(t)|$ this effect is not visible any more [see for example the case $m(x) = x$]. Consider for example the function $m(x) = \frac{1}{2}(2x-1)^3 + \frac{1}{2}$. At the point $x = \frac{1}{2}$ we have $m'(\frac{1}{2}) = m''(\frac{1}{2}) = 0$ and the larger bandwidths $h_d = \frac{1}{2}h_r$ for the density estimate yields a smaller MSE than the choice $h_d = h_r^3$. On the other hand, if $x = 0.2$ or $x = 0.8$ we have $|m''(0.2)| = 36.5$ and the effect of the bias is visible such that a smaller bandwidth h_d in the density estimate is appropriate. For larger variances, say $\sigma = 1$, the choice $h_d = \frac{1}{2}h_r$ usually yields a better MSE, but the advantages are not substantial.

However, the differences become more visible for larger standard deviations (these results are not presented here for the sake of brevity). Finally, we note that in our simulation study the monotone estimates have a smaller mean squared error than the unconstrained estimates for large noise [see Table 4.2], while no general pattern can be observed for small variances ($\sigma = 0.1$) [see Table 4.1]. Based on our numerical results we recommend the choice $h_d = h_r^3$ or $h_d = h_r^2$ for the bandwidth in the density estimation step, where the particular alternative depends on the desired smoothness of the monotone estimate. This choice has the additional advantage that the regions, where boundary effects affect the density estimate are very small and that the first order asymptotic behaviour of the monotone estimate coincides with that of the local linear estimate.

		$h_d = \frac{1}{2} \cdot h_r$			$h_d = h_r^3$		
$m(x)$	x	MSE	Bias	Varianz	MSE	Bias	Varianz
(1)	0.2	$3.88198 \cdot 10^{-4}$	$-3.45152 \cdot 10^{-3}$	$3.76285 \cdot 10^{-4}$	$3.98194 \cdot 10^{-4}$	$-1.28041 \cdot 10^{-3}$	$3.96555 \cdot 10^{-4}$
		$(4.16619 \cdot 10^{-4})$	$(-5.02454 \cdot 10^{-4})$	$(4.16366 \cdot 10^{-4})$	$(4.00358 \cdot 10^{-4})$	$(3.74106 \cdot 10^{-4})$	$(4.00218 \cdot 10^{-4})$
	0.5	$3.50221 \cdot 10^{-4}$	$-5.73912 \cdot 10^{-4}$	$3.49892 \cdot 10^{-4}$	$4.23894 \cdot 10^{-4}$	$-3.04373 \cdot 10^{-4}$	$4.23801 \cdot 10^{-4}$
		$(3.93074 \cdot 10^{-4})$	$(-5.59361 \cdot 10^{-4})$	$(3.92761 \cdot 10^{-4})$	$(4.23665 \cdot 10^{-4})$	$(-3.03034 \cdot 10^{-4})$	$(4.23573 \cdot 10^{-4})$
	0.8	$3.39777 \cdot 10^{-4}$	$2.14544 \cdot 10^{-3}$	$3.35174 \cdot 10^{-4}$	$4.03153 \cdot 10^{-4}$	$1.84656 \cdot 10^{-3}$	$3.99743 \cdot 10^{-4}$
		$(3.82281 \cdot 10^{-4})$	$(-7.17278 \cdot 10^{-4})$	$(3.81766 \cdot 10^{-4})$	$(3.99410 \cdot 10^{-4})$	$(1.28901 \cdot 10^{-4})$	$(3.99393 \cdot 10^{-4})$
(2)	0.2	$4.50312 \cdot 10^{-4}$	$1.21897 \cdot 10^{-2}$	$3.01722 \cdot 10^{-4}$	$4.55540 \cdot 10^{-4}$	$7.74593 \cdot 10^{-3}$	$3.95541 \cdot 10^{-4}$
		$(4.17314 \cdot 10^{-4})$	$(3.42046 \cdot 10^{-3})$	$(4.05614 \cdot 10^{-4})$	$(4.61647 \cdot 10^{-4})$	$(5.63278 \cdot 10^{-3})$	$(4.29919 \cdot 10^{-4})$
	0.5	$4.06086 \cdot 10^{-4}$	$6.70773 \cdot 10^{-3}$	$3.61092 \cdot 10^{-4}$	$4.37362 \cdot 10^{-4}$	$5.82293 \cdot 10^{-3}$	$4.03456 \cdot 10^{-4}$
		$(4.32560 \cdot 10^{-4})$	$(5.41455 \cdot 10^{-3})$	$(4.03243 \cdot 10^{-4})$	$(4.38001 \cdot 10^{-4})$	$(5.82658 \cdot 10^{-3})$	$(4.04052 \cdot 10^{-4})$
	0.8	$4.74827 \cdot 10^{-4}$	$1.07701 \cdot 10^{-2}$	$3.58832 \cdot 10^{-4}$	$4.62387 \cdot 10^{-4}$	$6.67026 \cdot 10^{-3}$	$4.17894 \cdot 10^{-4}$
		$(4.06161 \cdot 10^{-4})$	$(5.40935 \cdot 10^{-3})$	$(3.76900 \cdot 10^{-4})$	$(4.42168 \cdot 10^{-4})$	$(4.94562 \cdot 10^{-3})$	$(4.17709 \cdot 10^{-4})$
(3)	0.2	$4.14335 \cdot 10^{-4}$	$-6.47495 \cdot 10^{-3}$	$3.72410 \cdot 10^{-4}$	$4.26854 \cdot 10^{-4}$	$-3.46118 \cdot 10^{-3}$	$4.14875 \cdot 10^{-4}$
		$(3.98837 \cdot 10^{-4})$	$(-3.03924 \cdot 10^{-3})$	$(3.95595 \cdot 10^{-4})$	$(4.16850 \cdot 10^{-4})$	$(-1.71396 \cdot 10^{-3})$	$(4.13912 \cdot 10^{-4})$
	0.5	$4.05091 \cdot 10^{-4}$	$-4.07137 \cdot 10^{-3}$	$3.88515 \cdot 10^{-4}$	$4.30843 \cdot 10^{-4}$	$-4.22549 \cdot 10^{-3}$	$4.12988 \cdot 10^{-4}$
		$(4.38181 \cdot 10^{-4})$	$(-1.80049 \cdot 10^{-3})$	$(4.28944 \cdot 10^{-4})$	$(4.30739 \cdot 10^{-4})$	$(-4.22316 \cdot 10^{-3})$	$(4.12903 \cdot 10^{-4})$
	0.8	$4.77313 \cdot 10^{-4}$	$-1.34917 \cdot 10^{-2}$	$2.95285 \cdot 10^{-4}$	$3.97258 \cdot 10^{-4}$	$-6.89710 \cdot 10^{-3}$	$3.49688 \cdot 10^{-4}$
		$(4.64601 \cdot 10^{-4})$	$(-6.42359 \cdot 10^{-3})$	$(4.23338 \cdot 10^{-4})$	$(4.16981 \cdot 10^{-4})$	$(-5.67047 \cdot 10^{-3})$	$(3.84827 \cdot 10^{-4})$
(4)	0.2	$1.05505 \cdot 10^{-3}$	$-2.58045 \cdot 10^{-2}$	$3.89175 \cdot 10^{-4}$	$8.36376 \cdot 10^{-4}$	$-1.99779 \cdot 10^{-2}$	$4.37257 \cdot 10^{-4}$
		$(7.83679 \cdot 10^{-4})$	$(-1.83794 \cdot 10^{-2})$	$(4.45876 \cdot 10^{-4})$	$(7.68900 \cdot 10^{-4})$	$(-1.82236 \cdot 10^{-2})$	$(4.36800 \cdot 10^{-4})$
	0.5	$2.03017 \cdot 10^{-4}$	$-7.09120 \cdot 10^{-4}$	$2.02514 \cdot 10^{-4}$	$2.79063 \cdot 10^{-4}$	$2.20373 \cdot 10^{-4}$	$2.79015 \cdot 10^{-4}$
		$(4.20886 \cdot 10^{-4})$	$(-4.29921 \cdot 10^{-4})$	$(4.20701 \cdot 10^{-4})$	$(3.66362 \cdot 10^{-4})$	$(-2.70634 \cdot 10^{-4})$	$(3.66289 \cdot 10^{-4})$
	0.8	$9.49052 \cdot 10^{-4}$	$2.44436 \cdot 10^{-2}$	$3.51562 \cdot 10^{-4}$	$8.30827 \cdot 10^{-4}$	$2.01935 \cdot 10^{-2}$	$4.23048 \cdot 10^{-4}$
		$(6.98030 \cdot 10^{-4})$	$(1.71293 \cdot 10^{-2})$	$(4.04616 \cdot 10^{-4})$	$(7.62256 \cdot 10^{-4})$	$(1.84477 \cdot 10^{-2})$	$(4.21936 \cdot 10^{-4})$

Table 4.1: *Simulated mean squared error, bias and variance of the monoton estimator $\hat{m}_I$ and the local linear estimator $\hat{m}$ (values in brackets) for various regression functions and a uniform design: (1) $m(x) = x$, (2) $m(x) = x^2$, (3) $m(x) = \sin(\frac{\pi}{2}x)$ and (4) $m(x) = \frac{1}{2}(2x-1)^3 + \frac{1}{2}$. The bandwidth of the regression estimate is given by (4.3) and the estimates are calculated from $n = 100$ observations with standard deviation $\sigma = 0.1$.*

		$h_d = \frac{1}{2} \cdot h_r$			$h_d = h_r^3$		
	x	MSE	Bias	Varianz	MSE	Bias	Varianz
(1)	0.2	$2.20820 \cdot 10^{-2}$	$3.83279 \cdot 10^{-3}$	$2.20673 \cdot 10^{-2}$	$2.13866 \cdot 10^{-2}$	$-2.27475 \cdot 10^{-3}$	$2.13814 \cdot 10^{-2}$
		$(2.43494 \cdot 10^{-2})$	$(6.88163 \cdot 10^{-3})$	$(2.43021 \cdot 10^{-2})$	$(2.16107 \cdot 10^{-2})$	$(-3.91854 \cdot 10^{-3})$	$(2.15954 \cdot 10^{-2})$
	0.5	$1.20904 \cdot 10^{-2}$	$3.44610 \cdot 10^{-3}$	$1.20785 \cdot 10^{-2}$	$1.45233 \cdot 10^{-2}$	$-6.19947 \cdot 10^{-3}$	$1.44848 \cdot 10^{-2}$
		$(1.49110 \cdot 10^{-2})$	$(3.88867 \cdot 10^{-3})$	$(1.48959 \cdot 10^{-2})$	$(1.60610 \cdot 10^{-2})$	$(-5.43948 \cdot 10^{-3})$	$(1.60314 \cdot 10^{-2})$
	0.8	$2.06413 \cdot 10^{-2}$	$6.14231 \cdot 10^{-3}$	$2.06035 \cdot 10^{-2}$	$2.50356 \cdot 10^{-2}$	$-4.72176 \cdot 10^{-3}$	$2.50133 \cdot 10^{-2}$
		$(2.29607 \cdot 10^{-2})$	$(3.34988 \cdot 10^{-3})$	$(2.29495 \cdot 10^{-2})$	$(2.56745 \cdot 10^{-2})$	$(-3.64397 \cdot 10^{-3})$	$(2.56612 \cdot 10^{-2})$
(2)	0.2	$2.00228 \cdot 10^{-2}$	$1.29049 \cdot 10^{-2}$	$1.98562 \cdot 10^{-2}$	$2.10975 \cdot 10^{-2}$	$1.74607 \cdot 10^{-2}$	$2.07926 \cdot 10^{-2}$
		$(2.34644 \cdot 10^{-2})$	$(1.61745 \cdot 10^{-2})$	$(2.32028 \cdot 10^{-2})$	$(2.36844 \cdot 10^{-2})$	$(1.56773 \cdot 10^{-2})$	$(2.34386 \cdot 10^{-2})$
	0.5	$1.60288 \cdot 10^{-2}$	$5.70680 \cdot 10^{-2}$	$1.27721 \cdot 10^{-2}$	$1.52649 \cdot 10^{-2}$	$4.44132 \cdot 10^{-2}$	$1.32924 \cdot 10^{-2}$
		$(1.71533 \cdot 10^{-2})$	$(3.12126 \cdot 10^{-2})$	$(1.61791 \cdot 10^{-2})$	$(1.60654 \cdot 10^{-2})$	$(2.44530 \cdot 10^{-2})$	$(1.54674 \cdot 10^{-2})$
	0.8	$2.30870 \cdot 10^{-2}$	$1.92527 \cdot 10^{-2}$	$2.27164 \cdot 10^{-2}$	$2.39332 \cdot 10^{-2}$	$1.74988 \cdot 10^{-2}$	$2.36270 \cdot 10^{-2}$
		$(2.43430 \cdot 10^{-2})$	$(1.48117 \cdot 10^{-2})$	$(2.41236 \cdot 10^{-2})$	$(2.40616 \cdot 10^{-2})$	$(1.43056 \cdot 10^{-2})$	$(2.38570 \cdot 10^{-2})$
(3)	0.2	$2.33988 \cdot 10^{-2}$	$-1.47132 \cdot 10^{-2}$	$2.31823 \cdot 10^{-2}$	$2.38061 \cdot 10^{-2}$	$-1.45759 \cdot 10^{-2}$	$2.35936 \cdot 10^{-2}$
		$(2.43304 \cdot 10^{-2})$	$(-1.13764 \cdot 10^{-2})$	$(2.42010 \cdot 10^{-2})$	$(2.36457 \cdot 10^{-2})$	$(-1.39270 \cdot 10^{-2})$	$(2.34518 \cdot 10^{-2})$
	0.5	$1.44971 \cdot 10^{-2}$	$-4.78662 \cdot 10^{-2}$	$1.22059 \cdot 10^{-2}$	$1.43279 \cdot 10^{-2}$	$-3.89659 \cdot 10^{-2}$	$1.28096 \cdot 10^{-2}$
		$(1.60098 \cdot 10^{-2})$	$(-2.73210 \cdot 10^{-2})$	$(1.52634 \cdot 10^{-2})$	$(1.54996 \cdot 10^{-2})$	$(-2.53982 \cdot 10^{-2})$	$(1.48545 \cdot 10^{-2})$
	0.8	$2.15224 \cdot 10^{-2}$	$-2.40621 \cdot 10^{-2}$	$2.09434 \cdot 10^{-2}$	$2.19968 \cdot 10^{-2}$	$-2.24630 \cdot 10^{-2}$	$2.14922 \cdot 10^{-2}$
		$(2.46888 \cdot 10^{-2})$	$(-2.33662 \cdot 10^{-2})$	$(2.41428 \cdot 10^{-2})$	$(2.30151 \cdot 10^{-2})$	$(-1.79270 \cdot 10^{-2})$	$(2.26937 \cdot 10^{-2})$
(4)	0.2	$2.64675 \cdot 10^{-2}$	$-9.52509 \cdot 10^{-2}$	$1.73947 \cdot 10^{-2}$	$2.62166 \cdot 10^{-2}$	$-8.01244 \cdot 10^{-2}$	$1.97967 \cdot 10^{-2}$
		$(2.70579 \cdot 10^{-2})$	$(-5.91196 \cdot 10^{-2})$	$(2.35628 \cdot 10^{-2})$	$(2.80176 \cdot 10^{-2})$	$(-6.21125 \cdot 10^{-2})$	$(2.41596 \cdot 10^{-2})$
	0.5	$1.10375 \cdot 10^{-2}$	$3.90471 \cdot 10^{-3}$	$1.11022 \cdot 10^{-2}$	$1.26469 \cdot 10^{-2}$	$2.36610 \cdot 10^{-3}$	$1.26469 \cdot 10^{-2}$
		$(1.48730 \cdot 10^{-2})$	$(2.77092 \cdot 10^{-3})$	$(1.48654 \cdot 10^{-2})$	$(1.59408 \cdot 10^{-2})$	$(-3.69709 \cdot 10^{-3})$	$(1.59271 \cdot 10^{-2})$
	0.8	$2.56056 \cdot 10^{-2}$	$9.80171 \cdot 10^{-2}$	$1.59983 \cdot 10^{-2}$	$2.54270 \cdot 10^{-2}$	$7.57326 \cdot 10^{-2}$	$1.96916 \cdot 10^{-2}$
		$(2.64244 \cdot 10^{-2})$	$(6.44489 \cdot 10^{-2})$	$(2.22707 \cdot 10^{-2})$	$(2.73103 \cdot 10^{-2})$	$(5.56479 \cdot 10^{-2})$	$(2.42136 \cdot 10^{-2})$

Table 4.2: *Simulated mean squared error, bias and variance of the monoton estimator $\hat{m}_I$ and the local linear estimator $\hat{m}$ (values in brackets) for various regression functions and a uniform design: (1) $m(x) = x$, (2) $m(x) = x^2$, (3) $m(x) = \sin(\frac{\pi}{2}x)$ and (4) $m(x) = \frac{1}{2}(2x-1)^3 + \frac{1}{2}$. The bandwidth of the regression estimate is given by (4.3) and the estimates are calculated from $n = 100$ observations with standard deviation $\sigma = 1$.*

5 Appendix: proofs

Throughout this section we assume without loss of generality that the function m has a positive derivative on the interval $[0,1]$. The general case can easily be obtained by considering a subinterval, for which this property is satisfied (note that m' is continuous). Moreover, we assume for the sake of a transparent notation that the number of design points N in the estimate $\hat{m}_I$ equals the sample size n and write m_n instead of m_N.

Proof of Lemma 2.1. Obviously, we have

$$m_n^{-1}(t) = \int_0^1 \int_{-\infty}^t K_d\Big(\frac{m(x)-u}{h_d}\Big)\frac{1}{h_d}dudx \cdot (1 + O(\frac{1}{nh_d}))$$

and observing that the support of the kernel K_d is given by the interval $[-1,1]$ the leading term on the right hand side is estimated as follows

$$\text{(A.1)} \quad A(h_d) \; = \int_0^1 \int_{-\infty}^t K_d\Big(\frac{m(x)-u}{h_d}\Big)\frac{du}{h_d}dx$$

$$
\begin{aligned}
&= \int_0^{m^{-1}(t+h_d)} \int_{m(x)-h_d}^{t} K_d\Big(\frac{m(x)-u}{h_d}\Big)\frac{du}{h_d}dx \\
&= m^{-1}(t-h_d) \\
&\quad + \int_0^1 I\{m^{-1}(t-h_d) \le x \le m^{-1}(t+h_d)\} \int_{m(x)-h_d}^{t} K_d\Big(\frac{m(x)-u}{h_d}\Big)\frac{du}{h_d}dx \\
&= m^{-1}(t-h_d) + h_d \int_{\frac{m(0)-t}{h_d}}^{\frac{m(1)-t}{h_d}} I\{-1 \le z \le 1\}(m^{-1})'(t+zh_d)\int_z^1 K_d(v)dvdz.
\end{aligned}
$$

If $t \in (m(0), m(1))$ is fixed, we obtain from the identity

$$
\int_{-1}^{1}\int_z^1 K_d(v)dvdz = 1
$$

(note that K_d is symmetric and has compact support $[-1,1]$) and a Taylor expansion

$$
\begin{aligned}
A(h_d) &= m^{-1}(t-h_d) + h_d \int_{-1}^{1} (m^{-1})'(t+zh_d)\int_z^1 K_d(v)dvdz \\
&= m^{-1}(t) + h_d^2(m^{-1})''(t)\Big\{\frac{1}{2} + \int_{-1}^{1} z\int_z^1 K_d(v)dvdz\Big\} + o(h_d^2) \\
&= m^{-1}(t) + \kappa_2(K_d)h_d^2(m^{-1})''(t) + o(h_d^2)
\end{aligned}
$$

as $h_d \to 0$, where the last identity follows from the representation

$$
\int_{-1}^{1} z\int_z^1 K_d(v)dvdz = \frac{1}{2}\int_{-1}^{1} v^2K_d(v)dv - \frac{1}{2}.
$$

□

For a proof of Lemma 2.2 and Theorem 3.2 it is necessary to understand the operator, which maps a non-decreasing function m to its "quantile" $m^{-1}(t)$. We need a result on the functional delta method, which we could not find explicitly in the literature. For related results considering quantile processes see Fernholz (1983), Gill (1989) or Van der Vaart (1998). Consider a fixed $t \in \mathbb{R}$, and let $\mathcal{M}$ denote the set of all functions $H \in C^2[0,1]$ with positive derivative on the interval $[0,1]$, which contain t in the interior of their image, i.e. $t \in \operatorname{int} H([0,1])$. Consider the functional

$$
\Phi : \begin{cases} \mathcal{M} \to [0,1] \\ H \to H^{-1}(t) \end{cases}
$$

and define for $H_1, H_2 \in \mathcal{M}$ the function

$$
Q : \begin{cases} [0,1] \to \mathbb{R} \\ \lambda \to \Phi(H_1 + \lambda(H_2 - H_1)). \end{cases} \qquad \text{(A.2)}
$$

Note that in the case of existence $Q'(0)$ is the Gatéaux derivative of the functional Φ at H_1 in the direction of $H_2 - H_1$. The following result shows that this derivative exists and also gives the second derivative.

Lemma A.1. *The mapping* $Q : [0,1] \to \mathbb{R}$ *defined by (A.2) is two times continuously differentiable with*

$$
\begin{aligned}
\text{(A.3)} \qquad Q'(\lambda) &= -\frac{(H_2 - H_1)}{h_1 + \lambda(h_2 - h_1)} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(t) \\
\text{(A.4)} \qquad Q''(\lambda) &= Q'(\lambda)\Big\{\frac{-2(h_2 - h_1)}{h_1 + \lambda(h_2 - h_1)} + \frac{(H_2 - H_1)(h_1' + \lambda(h_2' - h_1'))}{\{h_1 + \lambda(h_2 - h_1)\}^2}\Big\} \circ Q(\lambda)
\end{aligned}
$$

where h_1, h_2 *denote the derivatives of* H_1, H_2*, respectively.*

Proof of Lemma A.1. Let

$$F(x, y) = (H_1 + x(H_2 - H_1))(y) - t,$$

then $Q(\lambda)$ is determined by the equation

$$F(\lambda, Q(\lambda)) = 0.$$

It is easy to see that the domain of the function Q can be extended in a neighbourhood of the interval $[0,1]$ and by the implicit function theorem it follows that Q is differentiable with derivative

$$Q'(\lambda) = -\frac{(H_2 - H_1) \circ Q(\lambda)}{h_1 \circ Q(\lambda) + \lambda(h_2 - h_1) \circ Q(\lambda)},$$

which proves (A.3). The calculation of the second derivative now follows by a straightforward application of the chain rule, which gives

$$
\begin{aligned}
Q''(\lambda) &= \frac{(H_2 - H_1)(h_2 - h_1)}{\{h_1 + \lambda(h_2 - h_1)\}^2} \circ Q(\lambda) \\
&- Q'(\lambda) \cdot \frac{(h_2 - h_1)(h_1 + \lambda(h_2 - h_1)) - (H_2 - H_1)(h_1' + \lambda(h_2' - h_1'))}{\{h_1 + \lambda(h_2 - h_1)\}^2} \circ Q(\lambda),
\end{aligned}
$$

and an application of (A.3) yields the representation (A.4).

□

Proof of Lemma 2.2. By a Taylor expansion we have from Lemma A.1 (with $H_1 = m^{-1}, H_2 = m_n^{-1}$)

$$m_n(t) - m(t) = \Phi(m_n^{-1}) - \Phi(m^{-1}) = Q(1) - Q(0) = Q'(\lambda^*)$$

for some $\lambda^* \in [0,1]$ (see Serfling (1980)), where

$$\text{(A.5)} \qquad Q'(\lambda^*) = -\frac{(m_n^{-1} - m^{-1})}{(m^{-1} + \lambda^*(m_n^{-1} - m^{-1}))'} \circ (m^{-1} + \lambda^*(m_n^{-1} - m^{-1}))^{-1}(t).$$

Note that

$$(m^{-1} + \lambda^*(m_n^{-1} - m^{-1})) \to m^{-1}$$

by Lemma 2.1 and introduce the notation $t_n = (m^{-1}+\lambda^*(m_n^{-1}-m^{-1}))^{-1}(t)$ (note that $t_n \to m(t)$). For the numerator in (A.5) we obtain

$$(m_n^{-1} - m^{-1})(t_n) - (m_n^{-1} - m^{-1})(m(t)) = (m_n^{-1} - m^{-1})'(\eta_n) \cdot (t_n - m(t)) \tag{A.6}$$

for some η_n with $|\eta_n - m(t)| \le |t_n - m(t)|$. For the first factor in (A.6) we have by a standard argument

$$\begin{aligned}(m_n^{-1} - m^{-1})'(\eta_n) &= \int_0^1 K_d\Big(\frac{m(x)-\eta_n}{h_d}\Big)\frac{dx}{h_d} - (m^{-1})'(\eta_n) + O\Big(\frac{1}{nh_d}\Big) \\ &= O(h_d^2) + O\Big(\frac{1}{nh_d}\Big) ,\end{aligned} \tag{A.7}$$

and as a consequence it follows from (A.5) and (A.6) that

$$Q'(\lambda^*) = -\frac{(m_n^{-1} - m^{-1}) \circ m(t)}{(m^{-1})'(m(t))} + o(h_d^2) + o\Big(\frac{1}{nh_d}\Big). \tag{A.8}$$

The assertion of Lemma 2.2 is now obtained from Lemma 2.1 and (A.5) [note that $(m^{-1})''(m(t)) = -m''(t)/\{m'(t)\}^3$].

□

Proof of Theorem 3.1. We only prove the first part of the theorem, the second assertion follows by similar arguments. We use the decomposition

$$\hat{m}_I^{-1}(t) = \frac{1}{nh_d}\int_{-\infty}^t \sum_{i=1}^n K_d\Big(\frac{\hat{m}(\frac{i}{n})-u}{h_d}\Big)du = m_n^{-1}(t) + \Delta_n(t), \tag{A.9}$$

where m_n^{-1} is defined in (2.5) and Δ_n is given by

$$\Delta_n(t) = \frac{1}{nh_d}\sum_{i=1}^n \int_{-\infty}^t \Big\{K_d\Big(\frac{\hat{m}(\frac{i}{n})-u}{h_d}\Big) - K_d\Big(\frac{m(\frac{i}{n})-u}{h_d}\Big)\Big\}du. \tag{A.10}$$

For the latter term it follows that

$$\Delta_n(t) = \Delta_n^{(1)}(t) + \frac{1}{2}\Delta_n^{(2)}(t), \tag{A.11}$$

where

$$\Delta_n^{(1)}(t) = \frac{1}{nh_d^2}\sum_{i=1}^n \int_{-\infty}^t K_d'\Big(\frac{m(\frac{i}{n})-u}{h_d}\Big)\Big\{\hat{m}(\frac{i}{n}) - m(\frac{i}{n})\Big\}du, \tag{A.12}$$

$$\Delta_n^{(2)}(t) = \frac{1}{nh_d^3}\sum_{i=1}^n \int_{-\infty}^t K_d''\Big(\frac{\xi_i-u}{h_d}\Big)\Big\{\hat{m}(\frac{i}{n}) - m(\frac{i}{n})\Big\}^2 du, \tag{A.13}$$

with $|\xi_i - m(\frac{i}{n})| < |\hat{m}(\frac{i}{n}) - m(\frac{i}{n})|$ $(i = 1, \dots, n)$. A straightforward calculation shows that

$$\begin{aligned}|\Delta_n^{(2)}(t)| &= \frac{1}{h_d^2}\Big|\frac{1}{n}\sum_{i=1}^n K_d'\Big(\frac{\xi_i - t}{h_d}\Big)\{\hat{m}(\frac{i}{n}) - m(\frac{i}{n})\}^2\Big| \\ &= \frac{1}{h_d^2}\Big|\int_0^1 K_d'\Big(\frac{m(x)-t}{h_d}\Big)\{\hat{m}(x) - m(x)\}^2 dx\Big| \cdot (1 + o_p(1)).\end{aligned}$$

If we assume that the kernel K_r has been appropriately modified near the boundaries [see Müller (1984)] it follows that the expectation of the first term in the last expression is of order

$$O\Big(\frac{1}{h_d}\Big\{\frac{1}{nh_r}+h_r^4\Big\}\Big).$$

This implies

$$\sqrt{nh_d}\Delta_n^{(2)}(t)=o_p(1)\ , \tag{A.14}$$

and a combination of (A.9), (A.11) and (A.14) shows that the assertion of Theorem 3.1 can be proved establishing the weak convergence

$$\sqrt{nh_d}\Big(\Delta_n^{(1)}(t)+\kappa_2(K_r)h_r^2\Big(\frac{m''f+2m'f'}{fm'}\Big)(m^{-1}(t))\Big)\overset{\mathcal{D}}{\Rightarrow}\mathcal{N}(0,g^2(t)). \tag{A.15}$$

For this we use the decomposition

$$\Delta_n^{(1)}(t)=\Big(\Delta_n^{(1.1)}(t)+\Delta_n^{(1.2)}(t)\Big)(1+o_p(1)) \tag{A.16}$$

with

$$\Delta_n^{(1.1)}(t) = \frac{-1}{n^2h_dh_r}\sum_{i,j=1}^n K_d\Big(\frac{m(\frac{i}{n})-t}{h_d}\Big)K_r\Big(\frac{X_j-\frac{i}{n}}{h_r}\Big)\frac{m(X_j)-m(\frac{i}{n})}{f(\frac{i}{n})} \tag{A.17}$$

$$\Delta_n^{(1.2)}(t) = \frac{-1}{n^2h_dh_r}\sum_{i,j=1}^n K_d\Big(\frac{m(\frac{i}{n})-t}{h_d}\Big)K_r\Big(\frac{X_j-\frac{i}{n}}{h_r}\Big)\sigma(X_j)\frac{\varepsilon_j}{f(\frac{i}{n})}. \tag{A.18}$$

For the first term we obtain

$$\begin{aligned}
E\Big[\Delta_n^{(1.1)}(t)\Big] &= -\frac{(1+o(1))}{h_rh_d}\int_0^1\int_0^1 K_d\Big(\frac{m(x)-t}{h_d}\Big)K_r\Big(\frac{y-x}{h_r}\Big)f(y)\frac{m(y)-m(x)}{f(x)}dydx\\
&= -h_r^2\kappa_2(K_r)\int_0^1\frac{1}{h_d}K_d\Big(\frac{m(x)-t}{h_d}\Big)\Big\{m''(x)+\frac{2m'(x)f'(x)}{f(x)}\Big\}dx\cdot(1+o(1))\\
&= -h_r^2\kappa_2(K_r)\Big(\frac{m''f+2m'f'}{fm'}\Big)(m^{-1}(t))\cdot(1+o(1)),
\end{aligned} \tag{A.19}$$

while the variance of $\Delta_n^{(1.1)}(t)$ is given by

$$\begin{aligned}
\mathrm{Var}\Big(\Delta_n^{(1.1)}(t)\Big) &= \frac{1}{n^3h_d^2h_r^2}\mathrm{Var}\Big(\sum_{i=1}^n K_d\Big(\frac{m(\frac{i}{n})-t}{h_d}\Big)K_r\Big(\frac{X_j-\frac{i}{n}}{h_r}\Big)\frac{m(X_j)-m(\frac{i}{n})}{f(\frac{i}{n})}\Big)\\
&\le \frac{1}{nh_d^2h_r^2}E\Big[\Big(\int_0^1 K_d\Big(\frac{m(x)-t}{h_d}\Big)K_r\Big(\frac{X_j-x}{h_r}\Big)\frac{m(X_j)-m(x)}{f(x)}dx\Big)^2\Big](1+o(1))\\
&= o\Big(\frac{1}{nh_d}\Big).
\end{aligned} \tag{A.20}$$

This implies [using assumption (3.1)]

$$\Delta_n^{(1.1)}(t)+h_r^2\kappa_2(K_r)\Big(\frac{m''f+2m'f'}{fm'}\Big)(m^{-1}(t))=o_p\Big(\frac{1}{\sqrt{nh_d}}\Big)\ , \tag{A.21}$$

and consequently the assertion (A.15) follows from

$$\sqrt{nh_d}\Delta_n^{(1.2)}(t) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, g^2(t)). \tag{A.22}$$

For a proof of this relation we note that $E[\Delta_n^{(1.2)}(t)] = 0$ and calculate the variance

$$\begin{aligned}
\operatorname{Var}(\sqrt{nh_d}\Delta_n^{(1.2)}(t)) &= \frac{1}{n^3 h_d h_r^2} \sum_{j=1}^n \operatorname{Var}\Big(\sum_{i=1}^n \frac{\sigma(X_j)\varepsilon_j}{f(\frac{i}{n})} K_d\Big(\frac{m(\frac{i}{n}) - t}{h_d}\Big) K_r\Big(\frac{X_j - \frac{i}{n}}{h_r}\Big)\Big) \\
&= \frac{1}{h_d h_r^2} \int_0^1 \sigma^2(x) \Big[\int_0^1 K_d\Big(\frac{m(y) - t}{h_d}\Big) K_r\Big(\frac{x - y}{h_r}\Big) \frac{dy}{f(y)}\Big]^2 f(x)dx \cdot (1 + o(1)) \\
&= \frac{1}{h_d h_r^2} \int_0^1 K_d\Big(\frac{m(z) - t}{h_d}\Big) \int_0^1 K_d\Big(\frac{m(y) - t}{h_d}\Big) \frac{1}{f(y)f(z)} \\
&\quad \times \int_0^1 \sigma^2(x) K_r\Big(\frac{x - y}{h_r}\Big) K_r\Big(\frac{x - z}{h_r}\Big) f(x)dxdydz \cdot (1 + o(1)) \\
&= \frac{\sigma^2(m^{-1}(t))}{(m'(m^{-1}(t))^2 f(m^{-1}(t))} \frac{h_d}{h_r} \int\int\int K_d(w)K_d(v)K_r(u) \\
&\quad \times K_r\Big(\frac{m^{-1}(t + h_d v) - m^{-1}(t + h_d w)}{h_r} + u\Big) dudvdw \cdot (1 + o(1)) \\
&= \frac{\sigma^2(m^{-1}(t))}{m'(m^{-1}(t)) f(m^{-1}(t))} \int\int\int K_d\Big(w + \frac{h_r}{h_d} m'(m^{-1}(t))(v - u)\Big) \\
&\quad \times K_d(w)K_r(u)K_r(v)dwdudv \cdot (1 + o(1)),
\end{aligned} \tag{A.23}$$

where we applied the substitution $v \to \{m(m^{-1}(t+h_d w)+h_r(v-u))-t\}/h_d$ and the last identity uses the relation

$$\lim_{\substack{h_r \to 0, h_d \to 0 \\ h_r/h_d \to c}} K_d\Big(\frac{m(m^{-1}(t + h_d w) + h_r(v - u)) - t}{h_d}\Big) = K_d(w + cm'(m^{-1}(t))(v - u)). \tag{A.24}$$

This proves the representation of the asymptotic variance in (3.4). For a proof of the asymptotic normality we calculate by similar arguments

$$\begin{aligned}
&\sum_{j=1}^n E\Big[\Big\{\frac{\sigma(X_j)}{n^{3/2} h_d^{1/2} h_r} \varepsilon_j \sum_{i=1}^n K_d\Big(\frac{m(\frac{i}{n}) - t}{h_d}\Big) K_r\Big(\frac{X_j - \frac{i}{n}}{h_r}\Big) \frac{1}{f(\frac{i}{n})}\Big\}^4\Big] \\
&= \frac{E[\varepsilon_1^4]}{nh_d^2 h_r^4} \int \Big\{\prod_{j=1}^4 \int K_d\Big(\frac{m(x_j) - t}{h_d}\Big) K_r\Big(\frac{x - x_j}{h_r}\Big) \frac{dx_j}{f(x_j)}\Big\} \sigma^4(x)dx \cdot (1 + o(1)) \\
&= \frac{\sigma^4(m^{-1}(t))E[\varepsilon_1^4]}{nh_d} \frac{(m^{-1})'(t)}{\{f(m^{-1}(t))\}^4} \int\int \Big\{\prod_{j=2}^4 \int K_d\Big(\tilde{x} + \frac{h_r}{h_d} m'(m^{-1}(t))(y_j - y_1)\Big) K_r(y_j)dy_j\Big\} \\
&\qquad \times K_d(\tilde{x})K_r(y_1)dy_1 d\tilde{x} \cdot (1 + o(1)) \\
&= O\Big(\frac{1}{nh_d}\Big) = o(1)
\end{aligned}$$

and the asymptotic normality in (A.22) follows from the central limit theorem of Ljapunoff.

□

Proof of Theorem 3.2. We only prove the first part of the theorem, the second assertion follows by exactly the same arguments. From Lemma A.1 we obtain the Taylor expansion

$$H_2^{-1}(t) - H_1^{-1}(t) = Q(1) - Q(0) = Q'(0) + \frac{1}{2}Q''(\lambda^*) \tag{A.25}$$

for some $\lambda^* \in [0,1]$ [see Serfling (1980)], which will now be applied for the functions $H_2 = \hat{m}_I^{-1}, H_1 = m_n^{-1}$. This gives for the estimator $\hat{m}_I$ and the quantity m_n at the point t the representation

$$\hat{m}_I(t) - m_n(t) = A_n + \frac{1}{2}B_n \ , \tag{A.26}$$

where

$$\begin{aligned} A_n &= -\frac{(\hat{m}_I^{-1} - m_n^{-1})}{(m_n^{-1})'} \circ m_n(t) \\ B_n &= \frac{2(\hat{m}_I^{-1} - m_n^{-1})(\hat{m}_I^{-1} - m_n^{-1})'}{\{(m_n^{-1} + \lambda^*(\hat{m}_I^{-1} - m_n^{-1}))'\}^2} \circ \Big(m_n^{-1} + \lambda^*(\hat{m}_I^{-1} - m_n^{-1})\Big)^{-1}(t) \\ &- \frac{(\hat{m}_I^{-1} - m_n^{-1})^2(m_n^{-1} + \lambda^*(\hat{m}_I^{-1} - m_n^{-1}))''}{\{(\hat{m}_I^{-1} + \lambda^*(\hat{m}_I^{-1} - m_n^{-1}))'\}^3} \circ (m_n^{-1} + \lambda^*(\hat{m}_I^{-1} - m_n^{-1}))^{-1}(t). \end{aligned}$$

At the end of this proof we will show the estimates

$$A_n = -\frac{(\hat{m}_I^{-1} - m_n^{-1})}{(m^{-1})'} \circ m(t) + o_p\Big(\frac{1}{\sqrt{nh_d}}\Big) \ , \tag{A.27}$$

$$B_n = o_p\Big(\frac{1}{\sqrt{nh_d}}\Big), \tag{A.28}$$

then the first assertion of Theorem 3.2 can be obtained as follows. From (A.27), (A.28) and (A.26) we have

$$\begin{aligned} &\sqrt{nh_d}\Big(\hat{m}_I(t) - m_n(t) - \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{f}\Big)(t)\Big) \\ &= -\sqrt{nh_d}\frac{(\hat{m}_I^{-1} - m_n^{-1}) \circ m(t) + \kappa_2(K_r)h_r^2\Big(\frac{m''f+2m'f'}{f}\Big)(t) \cdot (m^{-1})' \circ m(t)}{(m^{-1})' \circ m(t)} + o_p(1) \\ &= -m'(t)\sqrt{nh_d}\Big\{(\hat{m}_I^{-1} - m_n^{-1}) \circ m(t) + \kappa_2(K_r)h_r^2\Big(\frac{m''f + 2m'f'}{m'f}\Big)(t)\Big\} + o_p(1) \\ &\xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(t)), \end{aligned}$$

where $s^2(t)$ is defined in (3.8) and we used the first part of Theorem 3.1 in the last step.

For a proof of the estimate (A.27) we consider the difference

$$\begin{aligned} D_n &= (\hat{m}_I^{-1} - m_n^{-1}) \circ m_n(t) - (\hat{m}_I^{-1} - m_n^{-1}) \circ m(t) \\ &= (\hat{m}_I^{-1} - m_n^{-1})'(\xi_n)(m_n(t) - m(t)) \end{aligned} \tag{A.29}$$

where $|\xi_n - m(t)| \le |m_n(t) - m(t)|$. The first factor can be estimated as follows (recall the definition of Δ_n in (A.10))

$$
\begin{aligned}
\text{(A.30)} \quad \Delta_n'(\xi_n) &= (\hat m_I^{-1} - m_n^{-1})'(\xi_n) = \frac{1}{nh_d} \sum_{i=1}^{n} \Big\{ K_d\Big(\frac{\hat m(\frac{i}{n}) - \xi_n}{h_d}\Big) - K_d\Big(\frac{m(\frac{i}{n}) - \xi_n}{h_d}\Big) \Big\} \\
&= \frac{1}{nh_d^2} \sum_{i=1}^{n} K_d'\Big(\frac{\eta_{i,n} - \xi_n}{h_d}\Big) \Big\{ \hat m(\frac{i}{n}) - m(\frac{i}{n}) \Big\},
\end{aligned}
$$

where

$$
|\eta_{i,n} - m(\frac{i}{n})| \le |\hat m(\frac{i}{n}) - m(\frac{i}{n})| = O(R_n) \quad \text{a.s.}
$$

with

$$
R_n = \Big(\frac{\log h_r^{-1}}{nh_r}\Big)^{1/2}
$$

[see Mack and Silverman (1982), Theorem B]. This yields

$$
\begin{aligned}
\text{(A.31)} \quad \Delta_n'(\xi_n) &= \frac{1}{nh_d^2} \sum_{i=1}^{n} K_d'\Big(\frac{m(\frac{i}{n}) - \xi_n}{h_d}\Big) \Big\{ \hat m(\frac{i}{n}) - m(\frac{i}{n}) \Big\} + O\Big(\frac{R_n^2}{h_d^3}\Big) \quad \text{a.s.} \\
&= \frac{1}{h_d^2} \int K_d'\Big(\frac{m(x) - m(t)}{h_d}\Big) \{\hat m(x) - m(x)\} dx + O\Big(R_n + \frac{R_n^2}{h_d^3} + \frac{1}{nh_d}\Big) \quad \text{a.s.} \\
&= O\Big(\frac{R_n}{h_d} + \frac{R_n^2}{h_d^3} + \frac{1}{nh_d}\Big) \qquad \text{a.s.}
\end{aligned}
$$

As a consequence we obtain from (A.29) and Lemma 2.2

$$
D_n = O\Big(R_n h_d + \frac{R_n^2}{h_d} + \frac{h_d}{n}\Big) = o\Big(\frac{1}{\sqrt{nh_d}}\Big) \quad \text{a.s.}
$$

The estimate (A.27) now follows from the fact that $(m_n^{-1})'(t) = (m^{-1})'(t) + o(1)$ [see the proof of Lemma 2.1].

The second estimate (A.28) is proved similarly and we only indicate the main steps. First we decompose $B_n = 2B_{n1} - B_{n2}$ where

$$
\begin{aligned}
\text{(A.32)} \qquad B_{n1} &= \frac{(\hat m_I^{-1} - m_n^{-1})(\hat m_I^{-1} - m_n^{-1})'(t_n)}{\{(m_n^{-1} + \lambda^*(\hat m_I^{-1} - m_n^{-1})\}^2(t_n)} \\
\text{(A.33)} \qquad B_{n2} &= \frac{(\hat m_I^{-1} - m_n^{-1})^2(m_n^{-1} + \lambda^*(\hat m_I^{-1} - m_n^{-1}))''(t_n)}{\{\hat m_I^{-1} + \lambda^*(\hat m_I^{-1} - m_n^{-1})\}^3(t_n)}
\end{aligned}
$$

and $t_n = (m_n^{-1} + \lambda^*(\hat m_I^{-1} - m_n^{-1}))^{-1}(t)$. Note that

$$
t_n \xrightarrow{P} m(t), \quad (m_n^{-1} + \lambda^*(\hat m_I^{-1} - m_n^{-1})) \xrightarrow{P} m^{-1}.
$$

Observing (A.31) and Theorem 3.1 we therefore obtain

$$
\begin{aligned}
B_{n1} &= O_p\Big(\frac{1}{\sqrt{nh_d}} \cdot \Big(\frac{R_n}{h_d} + \frac{R_n^2}{h_d^3}\Big)\Big) = o_p\Big(\frac{1}{\sqrt{nh_d}}\Big), \\
B_{n2} &= O_p\Big(\frac{1}{nh_d}\Big),
\end{aligned}
$$

which proves (A.28). □

Acknowledgements. The authors are grateful to Isolde Gottschlich who typed numerous versions of this paper with considerable technical expertise and to E. Mammen and W. Polonik for useful discussions and some help with the references. The work of the authors was supported by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen.

References

H.D. Brunk (1955). Maximum likelihood estimates of monotone parameters. Ann. Math. Statist. 26, 607-616.

K.F. Cheng, P.E. Lin (1981). Nonparametric estimation of a regression function. Z. Wahrsch. Verw. Gebiete 57, 223-233.

J. Fan, I. Gijbels (1996). Local polynomial modelling and its applications. Chapman and Hall, London.

L.T. Fernholz (1983). Von Mises Calculus for Statistical Functionals. Lecture Notes in Statistics 19, Springer, N.Y.

J. Friedman, R. Tibshirani (1984). The monotone smoothing of scatterplots. Technometrics 26, 243-250.

T. Gasser, A. Kneip, W. Köhler (1991). A flexible and fast method for automatic smoothing. J. Amer. Statist. Assoc. 86, 643-652.

R.D. Gill (1989). Non- and semiparametric maximum likelihood estimators and the von Mises method. Scand. J. Statist. 16, 97-128.

W. Härdle, P. Hall, S. Marron (1988). How far are automatically chosen regression smoothing parameters from their optimum? (with discussion). J. Amer. Statist. Assoc. 86, 643-652.

P. Hall, L.-S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. Ann. Statist. 29, 624-647.

C. Kelly, J. Rice (1990). Monotone smoothing with application to dose response curves and the assessment of synergism. Biometrics 46, 1071-1085.

Y.P. Mack, B.W. Silverman (1982). Weak and strong uniform consistency of kernel regression estimates. Z. Wahrsch. Verw. Gebiete 61, 405-415.

E. Mammen (1991). Estimating a smooth monotone regression function. Ann. Statist. 19, 724-740.

E. Mammen, J.S. Marron, B.A. Turlach, M.P. Wand (2001). A general projection framework for constrained smoothing. Statist. Sci. 16, 232-248.

E. Mammen, C. Thomas-Agnam (1999). Smoothing splines and shape restrictions. Scand. J. Statist. 26, 239-252.

H.G. Müller (1985). Kernel estimators of zeros and of location and size of extrema of regression functions. Scand. J. Statist. 12, 221-232.

H.G. Müller, T. Gasser (1979). Kernel estimates of regression functions. In: Smoothing Techniques for Curve Estimation. Lecture Notes in Mathematics. Springer-Verlag, New York.

H. Mukerjee (1988). Monotone nonparametric regression. Ann. Statist. 16, 741-750.

E.A. Nadaraya (1964). On estimating regression. Theory of Probability and its Applications 15, 134- 137.

J.O. Ramsay (1988). Monotone regression splines in action (with comments). Statist. Sci. 3, 425-461.

J. Rice (1984). Bandwidth choice for nonparametric regression. Ann. Statist. 12, 1215-1230.

R.J. Serfling (1980). Approximation Theorems of Mathematical Statistics. Wiley, N.Y.

M. Stone (1974). Cross validatory choice and assessment of statistical predictions (with discussion). J. Roy. Statist. Soc., Ser. B, 36, 111-147.

A. Van der Vaart (1998). Asymptotic Statistics. Cambridge, University Press.

M.P. Wand, M.G. Jones (1995). Kernel smoothing. Chapman and Hall, London.

G.S. Watson (1964). Smooth regression analysis. Sankya, Ser. A, 26, 359-372.

F.T. Wright (1981). The asymptotic behavior of monotone regression estimates. Ann. Statist. 9, 443-448.