Estimating a convex function in nonparametric regression

Melanie Birke

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: melanie.birke@ruhr-uni-bochum.de

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@ruhr-uni-bochum.de

FAX: +49 2 34 32 14 559

July 1, 2005

Abstract

A new nonparametric estimate of a convex regression function is proposed and its stochastic properties are studied. The method starts with an unconstrained estimate of the derivative of the regression function, which is firstly isotonized and then integrated. We prove asymptotic normality of the new estimate and show that it is first order asymptotically equivalent to the initial unconstrained estimate if the regression function is in fact convex. If convexity is not present the method estimates a convex function whose derivative has the same L^p-norm as the derivative of the (non-convex) underlying regression function. The finite sample properties of the new estimate are investigated by means of a simulation study and the application of the new method is demonstrated in two data examples.

AMS Subject classification: 62G05, 62G07

Keywords and Phrases: nonparametric regression, order restricted inference, convexity, Nadaraya-Watson estimate, nondecreasing rearrangement

1 Introduction

The estimation of functions under shape restrictions is an important problem in the analysis of the relationship among variables. In many settings experimenters are in the position of having strong presumptions that these relationships satisfy certain qualitative restrictions such as monotonicity, convexity or concavity. Typical examples appear in economics (indirect utility, production or cost functions), medicine (dose response experiments) or biology (growth curves). Much effort has been spent to the problem of estimating a monotone regression function using the least squares approach [see e.g. Brunk (1955), Mukerjee (1988) among others and Barlow, Bartholomew, Bremner and

Brunk (1972) or Robertsen, Wright and Dykstra (1988) for a summary of this work]. In the case of concave (convex) regression a concave (convex) least squares estimate was first proposed by Hildreth (1954) and its consistency was proved by Hanson and Pledger (1976). Some algorithms for the calculation of the concave (convex) least squares estimate can be found in Wu (1982) and Fraser and Massam (1989). Mammen (1991) derived the rate of convergence of a least squares estimator of a convex or concave regression function and its derivative at a fixed point, while recently Groeneboom, Jongblood and Wellner (2001) proved consistency and derived the asymptotic distribution of the estimator at a fixed point of positive curvature.

In the present paper we propose an alternative estimate of a convex or concave regression function, which is not based on the least squares technique and particularly attractive to users of conventional smoothing methods. We will restrict ourselves to the case of convexity, but the case of concavity will be obvious from our discussion (see also the second example in Section 5). On the one hand our work is motivated by the fact that the commonly used least squares technique and also other projection based techniques produce a rather unsmooth convex estimate, even if the underlying regression function "is known" to be smooth. On the other hand we are interested in a simple and computationally efficient convex estimate, because convex estimates of the regression function are usually calculated by successive projections [see e.g. Dykstra (1983) or Han (1988)] and these iterative procedures can be very slow in some cases.

The new estimate is obtained in several steps. We first estimate the derivative of the regression function, which is isotonized in a second step to obtain a strictly isotone and smooth estimate of the derivative of the regression function. In a final step this estimate is integrated to obtain a strictly convex regression estimate, which is (at least) two times continuously differentiable. We prove consistency and asymptotic normality of the new estimate with the common rates of convergence in nonparametric regression. In particular first order asymptotic equivalence of the new convex estimator to the unconstrained estimate is established, if the underlying regression is in fact convex. Otherwise it is shown that the procedure proposed in this paper produces a convex curve, whose derivative has for any $p > 0$ the same L^p-norm as the derivative of the given regression function. Because of its simplicity the new method is particularly appealing to users of conventional kernel methods if some prior knowledge regarding the smoothness of the regression function is available. Our approach is carefully explained in Section 2, where we also study its main properties from an asymptotic point of view. Some further properties of the new estimate are discussed in Section 3, while the finite sample performance of the new method is investigated in Section 4 by means of a simulation study. In Section 5 two data examples are investigated and the application of the new procedure is illustrated. Finally, all proofs and technical details are deferred to an Appendix. It is also remarkable that the procedure proposed in this paper is universally applicable to other problems of convex or concave estimation (for example convex density or hazard rate estimation or the convex estimation of a parametric curve). Whenever an unconstrained estimate of a convex (or concave) curve in a particular estimation problem is available, and the method is applied to this preliminary estimator, the resulting curve estimator is convex (or concave) and first order asymptotic equivalent to the unconstrained preliminary estimate. However, for the sake of brevity and definiteness this paper is restricted to the problem of estimating a nonparametric convex regression curve.

2 Convex estimation

Consider the common nonparametric regression model

$$(2.1) \qquad Y_i = m\left(X_i\right) + \sigma\left(X_i\right)\varepsilon_i \quad i = 1, \ldots, n,$$

where $X_1, \ldots, X_n$ are i.i.d. random variables with density $f : [0,1] \to \mathbb{R}$ bounded away from zero and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. independent of $X_1, \ldots, X_n$ with mean 0, variance 1 and existing fourth moment. We assume that the variance function $\sigma^2 : [0,1] \to \mathbb{R}^+$ is continuous and that the density and regression function are three times continuously differentiable. We note that m is strictly convex if and only if its derivative m' is strictly increasing. Therefore we first construct a strictly isotone estimate of the derivative of the regression function which will be integrated in a second step. For this purpose any unconstrained nonparametric estimate (kernel type, local polynomial, series or spline estimator) of the regression function could be used. For the sake of definiteness we consider the Nadaraya-Watson estimate and some remarks regarding a local linear estimate can be found in Remark 2.3. To be precise let K_r denote a kernel of order three [see Gasser, Müller and Mammitzsch (1985)] with compact support, say $[-1,1]$, which is three times continuously differentiable with Lipschitz continuous third derivative and define the kernel estimator of the derivative of order $p \in \mathbb{N}$ of the regression function as $\hat{m}^{(p)}(x) = \frac{\partial^p}{\partial x^p}\hat{m}(x)$ where $\hat{m}$ denotes the Nadaraya-Watson estimate given by

$$(2.2) \qquad \hat{m}(x) = \frac{\sum_{i=1}^{n} K_r\left(\frac{x-X_i}{h_r}\right)Y_i}{\sum_{i=1}^{n} K_r\left(\frac{x-X_i}{h_r}\right)}$$

[see Nadaraya (1964) or Watson (1964)]. Here h_r denotes a bandwidth converging to 0 with increasing sample size and the kernel K_r has been appropriately modified in order to address for boundary effects [see e.g. Müller (1985)]. Following Dette, Neumeyer and Pilz (2005) we consider a further kernel, say K_d, with compact support which is two times continuously differentiable with second derivative bounded away from zero and a bandwidth h_d and define

$$(2.3) \qquad \hat{\psi}_{h_d}\left(t\right) = \frac{1}{h_d}\int_0^1 \int_{-\infty}^t K_d\left(\frac{\hat{m}'\left(v\right) - u}{h_d}\right) du\, dv$$

as an estimate of the inverse function of m', say $(m')^{-1}$ at the point t. Note that the function in (2.3) is strictly increasing if h_d is sufficiently small and consequently its inverse is a strictly isotone and smooth estimate of the derivative of the regression function. Finally, the strictly convex estimate of the regression function m is defined by

$$(2.4) \qquad \hat{m}_C(x, u_0) = \hat{m}(u_0) + \int_{u_0}^x \hat{\psi}_{h_d}^{-1}(z)dz,$$

where $\hat{\psi}_{h_d}^{-1}$ denotes the inverse of the function $\hat{\psi}_{h_d}$ and (at the moment) $u_0 \in (0,1)$ is an arbitrary point. The choice of u_0 will be carefully discussed in Section 3. Note that the estimate $\hat{m}_C$ is obviously strictly convex because its derivative $\hat{\psi}_{h_d}^{-1}$ is strictly increasing by our construction.

Before we study the main asymptotic properties of our method we will present an alternative interpretation of our approach. The main idea is similar to the concept of nondecreasing rearrangements

3

[see Ryff (1965, 1970) or Bennett and Sharpely (1988)]. To be precise let g be a continous function with domain $[0,1]$ and range $R(g) \subseteq \mathbb{R}$. We define an operator ψ on $C([0,1])$, which maps the function g to the function

$$(2.5) \qquad \psi(g) : \begin{cases} R(g) & \to [0,1] \\ t & \to \psi(g)(t) = \int_{[0,1]} I\{g(x) \le t\} dx \ . \end{cases}$$

Obviously, we have for a strictly isotone function g

$$(2.6) \qquad \psi(g) = g^{-1} \quad \text{on } R(g),$$

On the other hand, the function $\psi(g)$ is always isotone independently if the function g possesses this property. In the following we denote by $\psi^{-1}(g)$ the inverse of the function $\psi(g)$, then it follows by similar arguments as given in Birke and Dette (2005) that for any continuous function g the identity

$$(2.7) \qquad \int_{[0,1]} |g(x)|^p dx = \int_{[0,1]} |\psi^{-1}(g)(x)|^p dx$$

holds for all $p \in (0, \infty]$. In other words: the function $\psi^{-1}(g)$ is increasing on the domain of g, has the same L^p-norm as the function g and coincides with g if this function is itself strictly increasing. In this sense the function $\psi^{-1}(g)$ can be considered as a monotone approximation of the function g. Note that the function $\psi^{-1}(g)$ is not necessarily differentiable. However, smoothing can easily be accomplished by replacing ψ in (2.5) by the operator $\psi_{h_d}(g) : \mathbb{R} \to \mathbb{R}$ defined by

$$(2.8) \qquad \psi_{h_d}(g)(t) = \frac{1}{h_d} \int_{[0,1]} \int_{-\infty}^{t} K_d\Big(\frac{g(x) - u}{h_d}\Big) du dx,$$

where h_d denotes a further bandwidth, which converges to 0 sufficiently fast. The function $\psi_{h_d}(g)$ is obviously differentiable and a standard calculation shows $\psi_{h_d}(g) = \psi(g) + o(1)$. In the present context we apply this concept to the function m' and denote by $\psi_{h_d}^{-1}(g)$ the inverse of the function $\psi_{h_d}(g)$. Note that $\psi_{h_d}(\hat{m}') = \hat{\psi}_{h_d}$ corresponds to the estimate (2.3) and that heuristically (if m' is strictly increasing)

$$(2.9) \qquad \hat{\psi}_{h_d} - (m')^{-1} = \psi_{h_d}(\hat{m}') - \psi(m') \approx \psi_{h_d}(\hat{m}') - \psi_{h_d}(m')$$
$$\approx \frac{\partial}{\partial \lambda} \psi_{h_d}(m' + \lambda(\hat{m}' - m'))|_{\lambda=0} \approx -\Big(\frac{\hat{m}' - m'}{m''}\Big) \circ (m')^{-1}$$

Applying this heuristic argument a second time [basically replacing $\hat{\psi}_{h_d}$ by $\hat{\psi}_{h_d}^{-1}$ and m' by $(m')^{-1}$] yields

$$\hat{\psi}_{h_d}^{-1} - m' \approx -\Big(\frac{\hat{\psi}_{h_d} - (m')^{-1}}{((m')^{-1})'}\Big) \circ m' \approx \hat{m}' - m',$$

where we used the representation (2.9) for the last approximation. Finally, we obtain from (2.10) for the convex estimate in (2.4)

$$\hat{m}_C(x, u_0) - m(x) = \hat{m}(u_0) - m(u_0) + \int_{u_0}^{x} \Big(\hat{\psi}_{h_d}^{-1}(z) - m'(z)\Big) dz$$
$$\approx \hat{m}(x) - m(x).$$

4

Therefore, if the regression function m is strictly convex, we expect that the new convex estimate $\hat{m}_C$ exhibits the same asymptotic behaviour as the unconstrained Nadaraya-Watson estimate of the regression function.

A detailed proof of such a statement is substantially more complicated and presented in the Appendix. For this purpose we require the following basic assumptions regarding the smoothing parameters in the estimates. The bandwidths h_d and h_r in the estimates (2.2) and (2.3) have to satisfy

(2.11.a) $$h_d,\ h_r \rightarrow 0$$

(2.11.b) $$nh_d,\ nh_r \rightarrow \infty$$

(2.11.c) $$h_d/h_r^{3/2} \rightarrow 0$$

(2.11.d) $$nh_r^7 = O(1)$$

(2.11.e) $$(\log h_r^{-1})^{3/2}/nh_r^4 h_d^2 = o(1)$$

if the sample size n converges to infinity. Our main results make the heuristic arguments of the previous paragraph precise.

Theorem 2.1. *Assume that the assumptions stated at the beginning of Section 2 and in (2.11.a)-(2.11.e) are satisfied. If the regression function m in (2.1) is strictly convex, then we have for any $x \in (0,1)$ with $m''(x) > 0$ and any $u_0 \in (0,1)$*

(2.12) $$\hat{m}_C(x, u_0) - m(x) = \hat{m}(x) - m(x) + o_p\left(\frac{1}{\sqrt{nh_r}}\right)$$

Corollary 2.2. *If the assumptions of Theorem 2.1 are satisfied, then*

(2.13) $$\sqrt{nh_r}\left(\hat{m}_C(x, u_0) - m(x) - b_n(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma(x))$$

for any $x, u_0 \in (0,1)$, where the asymptotic bias and variance are given by

$$b_n(x) = h_r^3 \kappa_3(K_r)\frac{(mf)^{(3)} - mf^{(3)}}{f}(x) + o\left(h_r^3\right),$$

$$\gamma(x) = \int_{-1}^{1} K_r^2(y)\,dy\left(\frac{\sigma^2}{f}\right)(x),$$

respectively and $\kappa_3(K_r) = 1/3!\ \int_{-1}^{1} u^3 K_r(u)du$.

Remark 2.3. Note that Theorem 2.1 and Corollary 2.2 show that the convex estimate $\hat{m}_C(x, u_0)$ is first order asymptotic equivalent to the preliminary unconstrained Nadaraya-Watson estimate $\hat{m}$ [see e.g. Härdle (1991)].

Moreover, a careful inspection of the proofs in the appendix shows that similar results can be derived, if alternative smoothing procedures are used as preliminary estimate. For example if $\hat{m}'$ is the derivative of a local polynomial estimate [see Fan and Gijbels (1996)] the asymptotic statement (2.13) is still true, where the bias has to be replaced by

$$\tilde{b}_n(x) = h_r^3 \kappa_3(K_r)m^{(3)}(x) + o(h_r^3)$$

3 Further discussion

In this section we briefly discuss the choice of the initial point $u_0 \in (0,1)$ which is crucial for the performance of the proposed procedure. For the sake of a transparent presentation we begin with the deterministic case, where the estimators in (2.4) have been replaced by their deterministic counterparts. The transfer to the estimates considered in Section 2 will be obvious from our discussion and is briefly mentioned at the end of this section.

Recall the definition of the function ψ and ψ_{h_d} in (2.5) and (2.8), respectively, and note that any function of the form

$$(3.1) \qquad m_C(x, u_0) = m(u_0) + \int_{u_0}^{x} \psi_{h_d}^{-1}(m')(t) dt$$

is strictly convex. Moreover, if m' is strictly increasing (in other words m is strictly convex) and $h_d \to 0$ we have $\psi_{h_d}^{-1}(m') \to m'$ and as a consequence

$$(3.2) \qquad m_C(x, u_0) \xrightarrow[h_d \to 0]{} m(x) \quad \text{for all } x, u_0 \in (0,1).$$

In this case the choice of the point u_0 is irrelevant, provided that h_d is sufficiently small. On the other hand, it is also important to discuss the impact of the choice of u_0 in the case, where m is not convex, because in applications the preliminary estimate $\hat{m}$ will rarely be convex (or its derivative $\hat{m}'$ isotone). If m is not convex each function $m_C(\cdot, u_0)$ with $u_0 \in (0,1)$ defines a "convex rearrangement" of the function $m(\cdot)$ whose derivative has approximately (if $h_d \to 0$) the same L^p-norm as the initial unconstrained function m. We will discuss further properties of this rearrangement in the following.

Example 3.1. In order to fix ideas we have illustrated the situation for the function

$$(3.3) \qquad m(x) = \frac{1}{2} + \frac{1}{2}(2x-1)^3 \quad x \in [0,1]$$

in Figure 1, which is obviously neither convex nor concave. In this case we calculate the limit of (3.1) as $h_d \to 0$, that is

$$(3.4) \qquad \bar{m}_C(x, u_0) = m(u_0) + \int_{u_0}^{x} \psi^{-1}(m')(t) dt.$$

We have

$$\psi(m')(t) = \sqrt{\frac{t}{3}}; \quad t \in [0,3],$$

which gives $\psi^{-1}(m')(t) = 3t^2$ and

$$(3.5) \qquad \begin{aligned} \bar{m}_C(x, u_0) &= m(u_0) + \int_{u_0}^{x} \psi^{-1}(m')(t) dt \\ &= \frac{1}{2} + \frac{1}{2}(2u_0 - 1)^3 - u_0^3 + x^3 \end{aligned}$$

as convex rearrangement of the function m. In Figure 1 we display several of these functions and our basic function m defined by (3.3). Note that the derivatives of the functions $\bar{m}_C(x, u_0)$ and
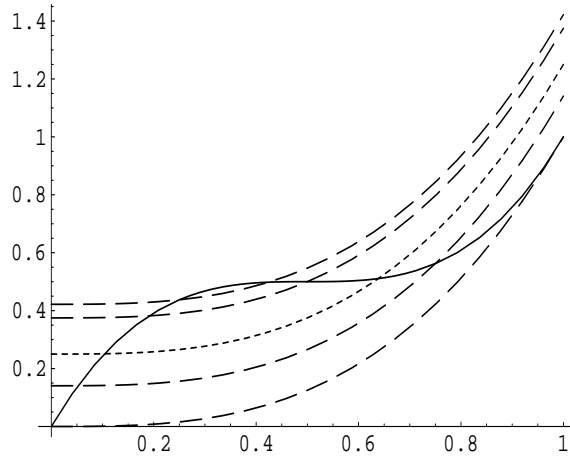
6

Figure 1: *The function m defined in (3.3) (solid line) and several convex rearrangements $\bar{m}_C(\cdot, u_0)$ defined by (3.4) with $u_0 = 0, 0.25, 0.5$ and 0.75 (dashed lines). The dotted line corresponds to the best L^2-approximation of the function m by the class $\{\bar{m}_C(x, u_0) | u_0 \in [0,1]\}$ defined in (3.5).*

$m(x)$ are given by $3x^2$ and $3(2x-1)^2$, respectively, and that both functions have for all $p > 0$ the same L^p-norm on the interval $[0,1]$.

If m is not strictly convex, an appropriate convex approximation of the function m by elements of the form (3.1) is now determined by minimizing the L^2-norm

$$(3.6) \qquad \Phi(u_0) = \int_0^1 (m(x) - m_C(x, u_0))^2 dx$$

over the interval $(0, 1)$. Our main result of this section characterizes the solution of this approximation problem.

Theorem 3.2. *Let Ψ_{h_d} denote any function with $(\Psi_{h_d})' = \psi_{h_d}^{-1}(m')$. A point $u_0^* \in (0, 1)$ minimizes the function Φ defined in (3.6) if and only if*

$$(3.7) \qquad m(u_0^*) - \Psi_{h_d}(u_0^*) = \int_0^1 (m(u) - \Psi_{h_d}(u)) du.$$

Moreover, we have for any u_0^* minimizing (3.6)

$$(3.8) \qquad m_C(x, u_0^*) = \int_0^1 m_C(x, u) du.$$

In other words: the function on the right hand side of (3.8) is of the form (3.1) and minimizes the L^2-distance in (3.6).

7

Example 3.3. In the situation considered in Example 3.1 we have $\Psi_{h_d} \to \Psi$ if $h_d \to 0$, where $\Psi' = \psi^{-1}(m')$, that is

$$m(u) - \Psi(u) = \frac{1}{2} + \frac{1}{2}(2u - 1)^3 - u^3,$$

$$\int_0^1 (m(u) - \Psi(u))du = \frac{1}{4}.$$

The solutions of the analogue of equation (3.7) in the interval $(0, 1)$ are given by

$$u_{0,1}^* \approx 0.103741; \; u_{0,2}^* \approx 0.638824,$$

and both points correspond to the best L^2-approximation of the function (3.3) by elements of the form (3.5); i.e.

$$\bar{m}_C(x, u_{0,j}^*) = \int_0^1 \bar{m}_C(x, u)du = \frac{1}{4} + x^3.$$

The application of Theorem 3.2 for the construction of a convex estimator of the form (2.4) with minimal L^2-distance to the initial unconstrained estimate $\hat{m}$ is now straightforward. Note that it is not necessary to calculate the solution of the sample analogue of (3.7). The representation (3.8) can be applied directly and

$$(3.9) \qquad\qquad \hat{m}_C(x) := \int_0^1 \hat{m}_C(x, u_0)du_0$$

yields the convex estimate, which has minimal L^2-distance to the unconstrained regression estimate $\hat{m}$ in the class

$$\{\hat{m}_C(x, u_0) \mid u_0 \in (0, 1)\}$$

defined by (2.4). We finally remark that the computational effort for the numerical evaluation of the integral in (3.9) can be decreased observing the relation

$$\hat{m}_C(x, u_0) = \hat{m}(x) + \hat{m}(u_0) - \hat{m}_C(u_0, x),$$

which is obvious from the representation (2.4).

4   Finite sample properties

In this section we briefly illustrate the finite sample properties of the convex estimate of the regression function by means of a simulation study. For this purpose we consider three examples

$$(4.1) \qquad\qquad m(x) = e^{3(x-1)},$$

$$(4.2) \qquad\qquad m(x) = \frac{16}{9}(x - \frac{1}{4})^2,$$

$$(4.3) \qquad\qquad m(x) = \begin{cases} -4x + 1 & \text{if } \; 0 \le x \le \frac{1}{4} \\ 0 & \text{if } \; \frac{1}{4} \le x \le \frac{3}{4} \\ 4x - 3 & \text{if } \; \frac{3}{4} \le x \le 1. \end{cases}$$
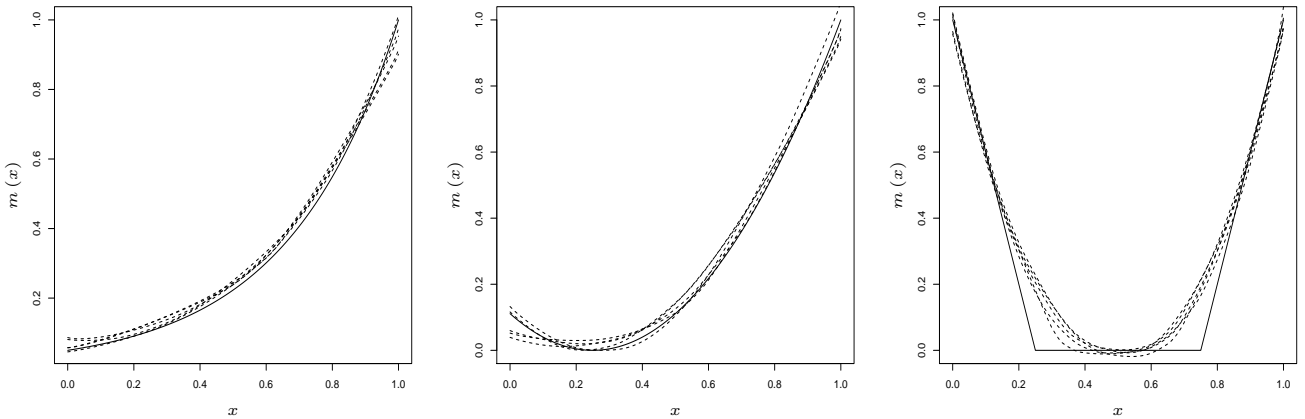
8

Figure 2: *The regression functions (4.1) (left part), (4.2) (middle part) and (4.3) (right part) and five convex estimates from different simulation runs. The sample size is $n = 100$ and the standard deviation is $\sigma = 0.1$.*

Note that all functions are convex but in contrast to the functions in (4.1) and (4.2) the function in (4.3) does not satisfy the assumptions required for the asymptotic theory in Section 2. The errors in the regression model (2.1) are assumed to be normal and homoscedastic with standard deviation $\sigma = 0.1$, while a uniform design was simulated for the explanatory variables. We used a sample of $n = 100$ observations to estimate the regression function. In the following $\hat{\sigma}^2$ denotes the estimator of the integrated variance proposed by Rice (1984). For the preliminary (unconstrained) estimate of m' the derivative of the local linear estimate was used in order to address for boundary effects. The bandwidth h_r in this estimate was chosen by

$$(4.4) \qquad\qquad h_r = \Big(\frac{\hat{\sigma}^2}{n}\Big)^{1/7}$$

which is proportional to the asymptotical optimal bandwidth for a three times continuously differentiable regression and density function while the bandwidth h_d for the density step is given by

$$(4.5) \qquad\qquad h_d = h_r^{8/5}.$$

The kernels K_d and K_r are both chosen as Epanechnikov kernel. For the selection of the point u_0 we used the best L^2-approximation proposed in Section 3, where the integral in (3.9) is approximated by a Riemann sum. In Figure 2 we display for each regression function five typical estimates obtained from different simulation runs. We observe a reasonable performance of the estimates for the regression functions (4.1) and (4.2). In example (4.3) the estimate does not precisely reproduce the true function at points where the regression function is not differentiable, but in the regions where the derivative is two times differentiable the estimates are again very reliable.

In the second part of this simulation study we investigate the mean squared error, bias and variance of the convex estimator proposed in this paper. For this we consider again the three regression functions in (4.1) - (4.3) and calculate by 2000 simulation runs the curves for the mean squared
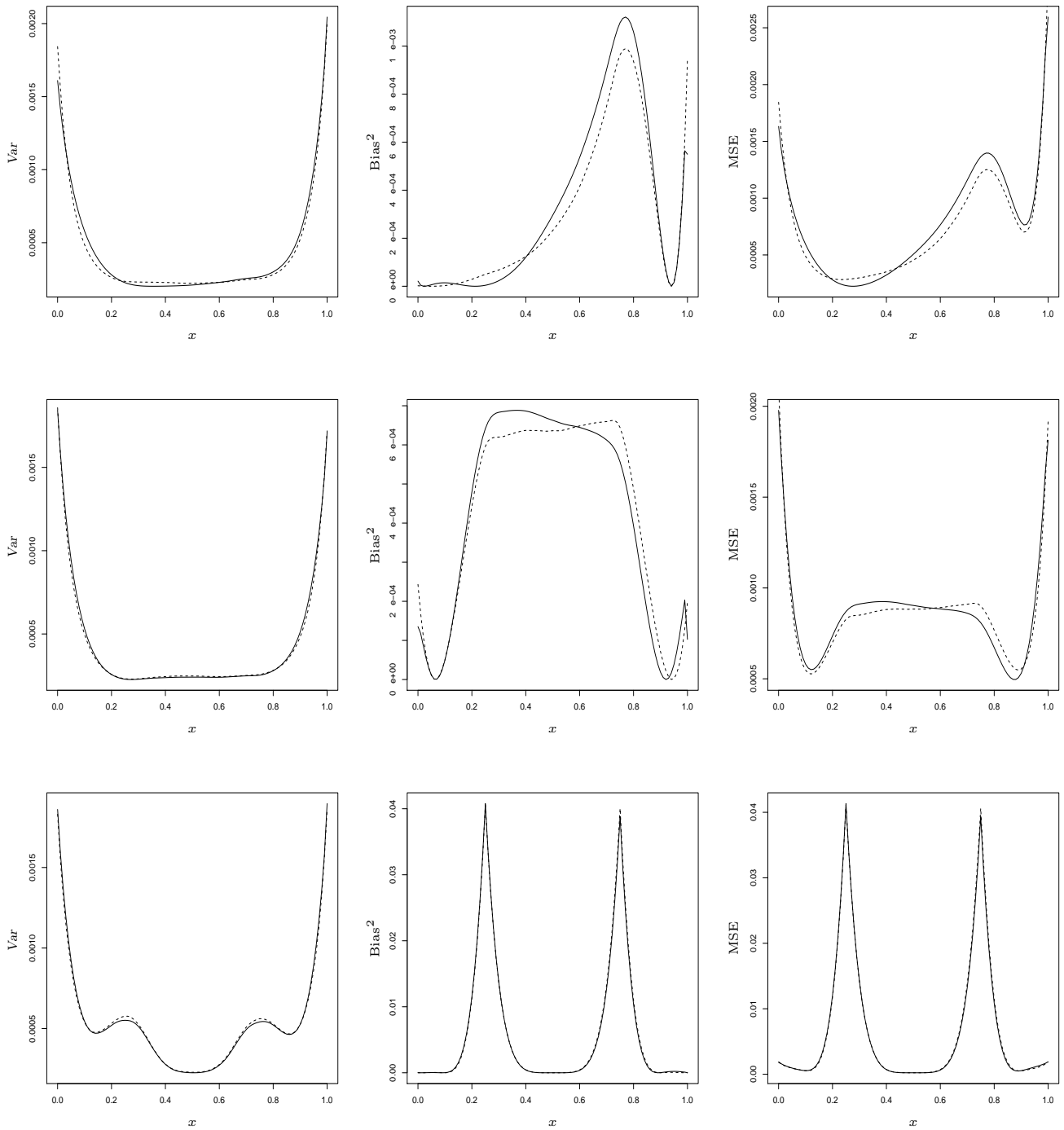
9

Figure 3: *Simulated variance, squared bias and mean squared error of the convex estimator $\hat{m}_C$ (solid line) and the local linear estimate $\hat{m}$ (dashed line). The sample size is $n = 100$, the standard deviation is $\sigma = 0.1$, while the regression functions are given by (4.1) (upper panel), (4.2) (middle panel) and (4.3) (lower panel), respectively.*

error, squared bias and variance. The results are depicted in Figure 3 for all three cases under consideration. In this figure the local linear (unconstrained) estimate is represented by the dashed line, while the new convex estimate $\hat{m}_C$ is depicted by the solid line. As predicted by the asymptotic theory only small differences between the unconstrained and convex estimates are observed. In the situation of example (4.1) we see that the squared bias of the local linear estimate $\hat{m}$ is smaller if $x \in [0.4, 0.9]$ while it is larger if $x \in [0.1, 0.4]$. Both estimates show a similar behaviour with respect to the variance criterion, where there are slight advantages for the convex estimate in the interior of the design space. The mean squared error is dominated by the bias and smaller for the local linear estimate in most parts of the interval $[0, 1]$ [see the first line in Figure 3].

For the regression function (4.2) we observe from the middle line of Figure 3 that the local linear estimate yields a smaller (squared) bias than the convex estimate $\hat{m}_C$ if $x < 0.6$, while the converse is true if $x \in [0.6, 0.9]$. The variance of the convex estimate is slightly smaller than the variance of the local linear estimate. Because the variance of both estimates is very similar, a comparison of the mean squared error curves shows some advantages for the constrained estimate if $x > 0.6$ and a smaller mean squared error of the local linear estimate if $x < 0.6$. Again the bias dominates the mean squared error. For the regression function (4.3) the situation is slightly different. Here both estimates have a similar variance, but there are some (small) advantages for the convex estimate in a neighbourhood of points, where the regression function is not differentiable [see the left lower panel in Figure 3]. The squared bias curves of both estimates are very similar. Again the squared bias of the local linear estimate is larger at points where the derivative of the regression function is discontinuous. In these regions the mean squared error is dominated by the squared bias, which is reflected in the right panel of the third line of Figure 3.

Note that in these examples the bias dominates the variance. A partial explanation of this observation stems from the fact that the variance of the errors of the simulated observations is rather small, namely $\sigma^2 = 0.01$. In order to compare the unconstrained and convex estimate in the case of a larger variance we present in Figure 4 the simulated variance, squared bias and mean squared error curves for the functions (4.1) - (4.3), where the simulated observations have variance $\sigma^2 = 1$. In example (4.3) the squared bias and variance are of the same size and there are no substantial visual differences compared to the case considered in Figure 3 [see the last lines in Figure 3 and 4]. However, in example (4.1) and (4.2) the situation is different and the variance dominates the squared bias. In these two examples the variance curves of both estimates show a similar behaviour with small advantages for the convex estimate [see the left panel in the first and second line of Figure 4], while the squared bias of the local linear estimate is larger than that of the constrained estimate [see the middle panels]. As a consequence the mean squared error of the new convex estimate is usually smaller than the mean squared error of the local linear estimate [see the right panel in the first and second line of Figure 4].
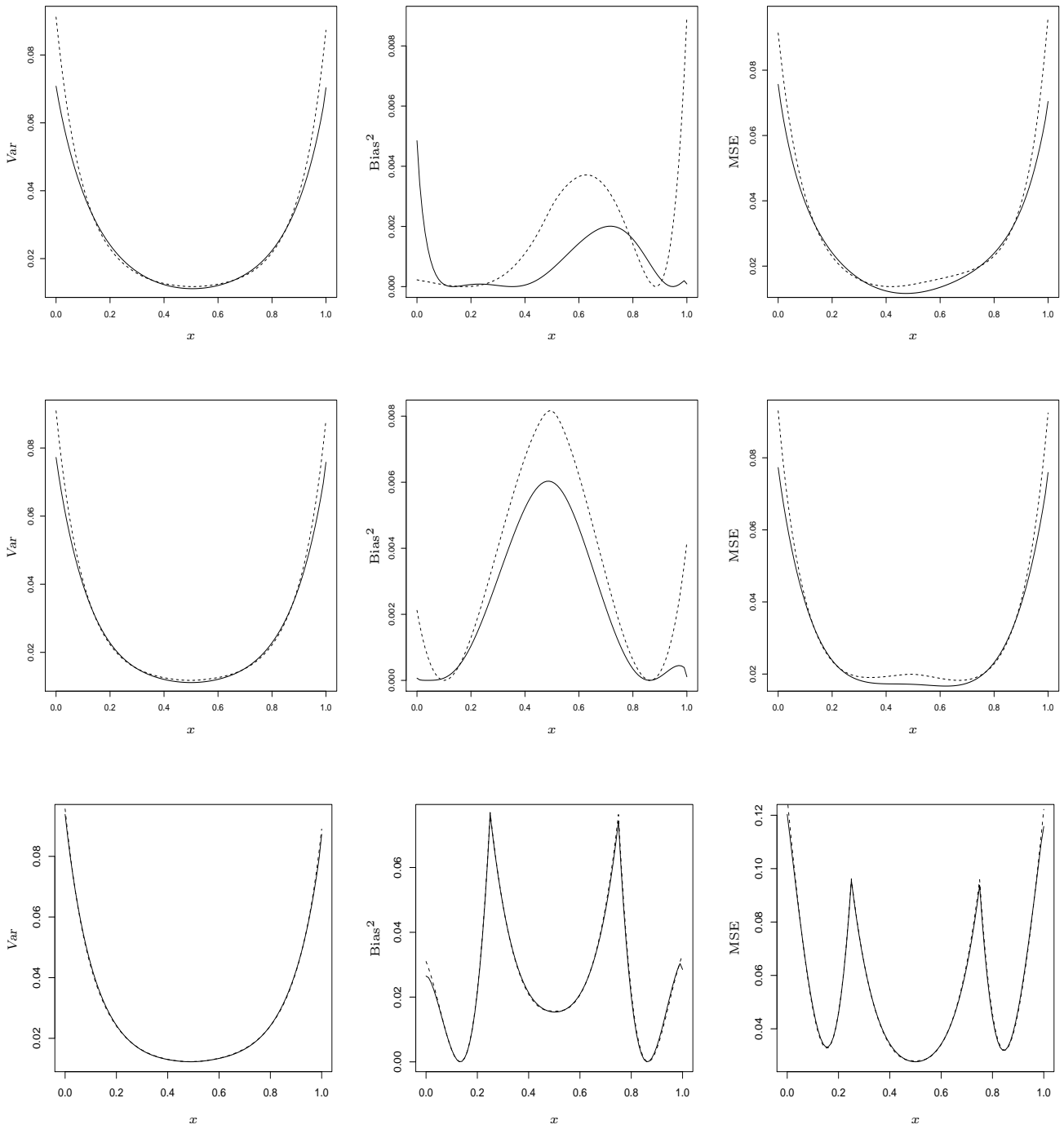
Figure 4: *Simulated variance, squared bias and mean squared error of the convex estimator $\hat{m}_C$ (solid line) and the local linear estimate $\hat{m}$ (dashed line). The sample size is $n = 100$, the standard deviation is $\sigma = 1$, while the regression functions are given by (4.1) (upper panel), (4.2) (middle panel) and (4.3) (lower panel), respectively.*
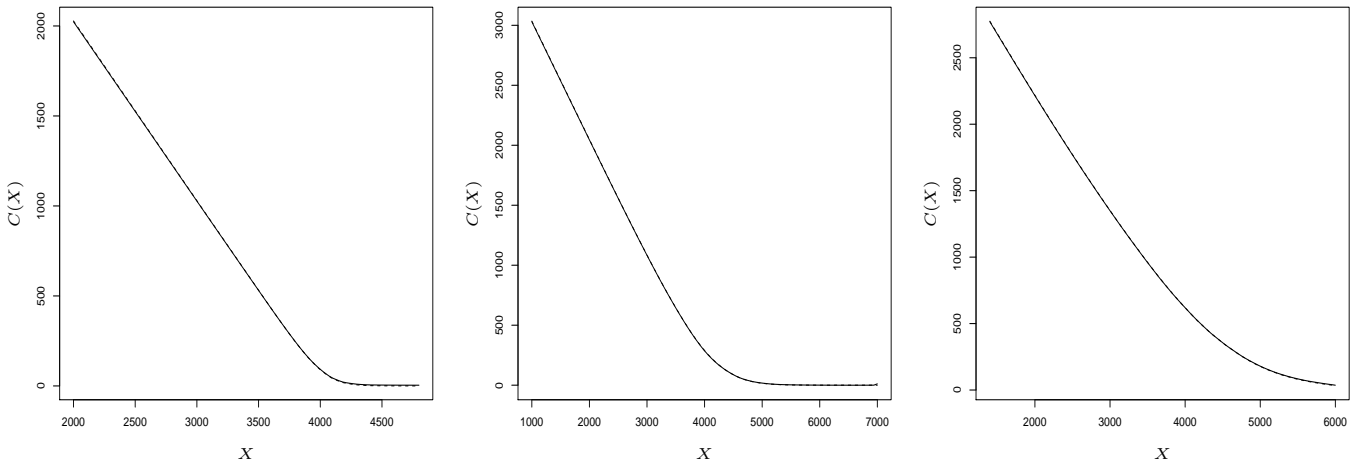
Figure 5: *The local linear estimate (dotted line) and new convex estimate (solid line) of the call pricing function of three DAX options considered as a function of the strike price. Left panel: February 9, 2004, expiration time 494 days; middle panel: January 2, 2004, expiration time 14 days; right panel: January 2, 2004; expiration time 168 days.*

5   Examples

5.1   Estimation of a convex option pricing function

The call pricing function of a European option, say C, at time t usually depends on the underlying asset price at date t, say S_t, the strike price X, the time to expiration τ, the deterministic risk free interest rate $r_{t,\tau}$ and the correspondend dividend yield $\delta_{t,\tau}$ of the asset. More precisely, this function is given by

$$C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = e^{-r_{t,\tau}} \int_0^\infty \max\{S_T - X, 0\} p^*(S_T \mid S_t, \tau, r_{t,\tau}, \delta_{t,\tau}) dS_T,$$

where $T = \tau + t$ is the expiration date and p^* denotes the state-price density, which is also called risk neutral density [see e.g. Black and Scholes (1973), or Cox, Ingersoll and Ross (1985)]. By taking derivatives of C with respect to X it was pointed out by Aït-Sahalia and Duarte (2003) that the function C must be decreasing and convex in X in order to rule out arbitrage opportunities. In particular, any local non-convexity of the call pricing function implies negative state prices, violating the no-arbitrage principle.

We will now illustrate a possible application of the proposed procedure and construct a convex estimate of the function C considered as a function of the strike price X. For this we analyze data from a DAX-option, which was kindly provided by K. Pilz (Sal. Oppenheim Bank). The data shows for a particular day the strike price of the option, expiration time, volatility, call price and closing price of the DAX. We construct a convex estimate of the call pricing function as a function of X. As preliminary estimate $\hat{m}'$ we use the derivative of a local linear estimator with bandwidth h_r chosen by least squares cross validation. The bandwidth h_d is chosen as $h_d = h_r^{8/5}$ and its influence is less
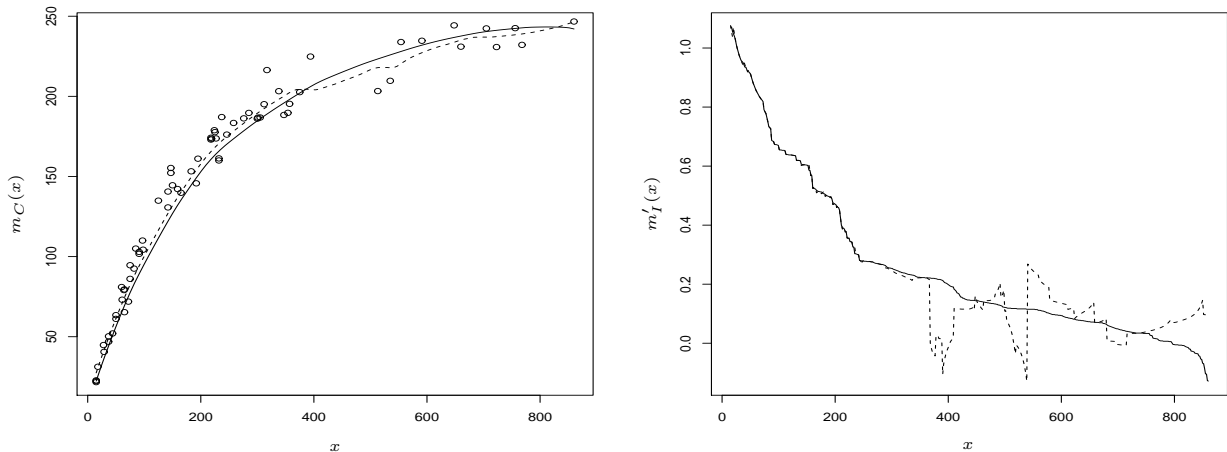
13

Figure 6: *Left panel: The local linear estimate $\hat{m}$ (dotted line) and convex estimate $\hat{m}_C$ (solid line) of the dry weight (in milligrams) of the eye lense as a function of the age (in days) estimated from 71 free-living wild rabbits. Right panel: the derivative $\hat{m}'$ of the local linear estimator and its antitone estimate $\tilde{\psi}_{h_d}^{-1}$ defined by (5.1).*

crucial. In Figure 5 we show the local linear and the new convex estimate of the call pricing function of a DAX option as a function of the strike price for 3 representative cases. The left part of Figure 5 corresponds to the price at February 9, 2004, where the expiration time is 494 days. Similarly, the middle and right panel of Figure 5 correspond to call option prices at January 2, 2004, where expiration time is 14 and 168 days, respectively. We observe that in all cases a similar picture. The unconstrained estimate $\hat{m}$ of the pricing function is convex and also decreasing as a function of the strike price X. For this reason the convex estimate $\hat{m}_C$ essentially coincides with the local linear estimate, which is definitely a desirable property of the new procedure. If the resulting convex estimate $\hat{m}_C$ would not be decreasing it could be easily monotonized by the method described in Dette, Neumeyer and Pilz (2005).

5.2 Rabbits data

In the previous example the local linear estimate was already convex and the new convex estimate basically reproduced the unconstrained estimate. We now consider a situation, where the local linear estimate does not satisfy the constraints of convexity or concavity. For this we use an example considered by Dudzinsky and Mykytowycz (1961), who analyzed the relation between age and eye lens weight for rabbits in Australia. In this study, the dry weight of the eye lens was measured (in milligrams) for 71 free-living wild rabbits of known age (measured in days). A detailed description of the experiment and the data can be found in http://www.statsci.org/data/oz/rabbit.html. The data was analyzed by Ratkowski (1983) using the nonlinear growth model

$$m(x) = a\, e^{-b/(x+c)}, \ x > 0,$$

14

which is obviously concave with respect to the variable x. In the left panel of Figure 6 we display a local linear estimate and its "concave" rearrangement. A concave estimate is obtained by replacing the statistic $\hat{\psi}_{h_d}^{-1}$ in (2.4) by $\tilde{\psi}_{h_d}^{-1}$, where

$$(5.1) \qquad \tilde{\psi}_{h_d}(t) = \frac{1}{h_d} \int_0^1 \int_t^\infty K_d\left(\frac{\hat{m}'(v) - u}{h_d}\right) du\, dv$$

(note that $\tilde{\psi}_{h_d}^{-1}$ is an antitone estimate of m'). The bandwidth for the unconstrained estimate is again chosen by least square cross-validation while the bandwidth h_d in the monotonizing procedure is given by $h_d = h_r^{8/5}$. We observe that the local linear estimate of the regression function is not concave if the age of the rabbits is larger than 400 days. Therefore a monotonization of the derivative is performed. The derivative of the local linear estimate and its antitone rearrangement $\tilde{\psi}_{h_d}$ are depicted in the right panel of Figure 6. The new method yields a concave estimate, which fits the data adequately and preserves the features of the unconstrained local linear estimate in regions where this is already concave.

6   Appendix: Proofs

Note that the estimate $\hat{\psi}_{h_d}^{-1}$ in the integral of (2.4) is obtained as the inverse of the estimate $\hat{\psi}_{h_d}$ defined in (2.3). It was shown in Dette, Neumeyer and Pilz (2005) that the operator, which maps a function onto its quantile (at a certain point) is two times Gatéaux differentiable. In order to make the present paper self-contained we recall this result at the beginning of the proof.

Consider a fixed point $t \in \mathbb{R}$, 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. $t \in \text{int}\, H([0,1])$. 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

$$(6.1) \qquad Q : \begin{cases} [0,1] & \to & \mathbb{R} \\ \lambda & \to & \phi(H_1 + \lambda(H_2 - H_1)). \end{cases}$$

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 (6.1) is two times continuously differentiable with*

$$(6.2) \qquad Q'(\lambda) = -\frac{(H_2 - H_1)}{h_1 + \lambda(h_2 - h_1)} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(t),$$

$$(6.3) \qquad Q''(\lambda) = Q'(\lambda)\left\{\frac{-2(h_2 - h_1)}{h_1 + \lambda(h_2 - h_1)} + \frac{(H_2 - H_1)(h_1' + \lambda(h_2' - h_1'))}{\{h_1 + \lambda(h_2 - h_1)\}^2}\right\} \circ Q(\lambda),$$

15

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

6.1 Proof of Theorem 2.1.

By an application of Lemma A.1 we obtain for some $\lambda^* \in [0,1]$

$$(6.4) \qquad \hat{\psi}_{h_d}^{-1}(t) - m'(t) = A_n(t) + \frac{1}{2}B_{n,\lambda^*}(t),$$

where

$$(6.5) \qquad A_n(t) = -\frac{\hat{\psi}_{h_d} - m'^{-1}}{(m'^{-1})'} \circ m'(t),$$

m'^{-1} denotes the inverse of the (strictly increasing) function m' and

$$(6.6) \qquad B_{n,\lambda^*}(t) = \frac{2(\hat{\psi}_{h_d} - m'^{-1})(\hat{\psi}_{h_d} - m'^{-1})'}{\left\{(m'^{-1} + \lambda^*(\hat{\psi}_{h_d} - m'^{-1}))'\right\}^2} \circ (m'^{-1} + \lambda^*(\hat{\psi}_{h_d} - m'^{-1})^{-1}(t)$$

$$- \frac{(\hat{\psi}_{h_d} - m'^{-1})^2(m'^{-1} + \lambda^*(\hat{\psi}_{h_d} - m'^{-1}))''}{\left\{(m'^{-1} + \lambda^*(\hat{\psi}_{h_d} - m'^{-1}))'\right\}^3} \circ (m'^{-1} + \lambda^*(\hat{\psi}_{h_d} - m'^{-1}))^{-1}(t).$$

This yields for the estimate $\hat{m}_C$ in (2.4) the representation

$$(6.7) \qquad \hat{m}_C(x, u_0) - m(x) = \int_{u_0}^{x} \left(\hat{\psi}_{h_d}^{-1}(z) - m'(z)\right) dz + (\hat{m}(u_0) - m(u_0))$$

$$= \int_{u_0}^{x} A_n(z) \, dz + \frac{1}{2} \int_{u_0}^{x} B_{n,\lambda^*}(z) \, dz + (\hat{m}(u_0) - m(u_0))$$

for some $\lambda^* \in [0,1]$. We now treat the two terms in this expansion separately.

Lemma A.2. *If the assumptions of Theorem 2.1 are satisfied, we have*

$$\int_{u_0}^{x} A_n(z) dz + \hat{m}(u_0) - m(u_0) = (\hat{m} - m)(x) + o_p(\frac{1}{\sqrt{nh_r}})$$

Lemma A.3. *If the assumptions of Theorem 2.1 are satisfied we have*

$$\int_{u_0}^{x} B_{n,\lambda^*}(z) \, dz = o_p(\frac{1}{\sqrt{nh_r}})$$

The proof of Lemma A.2 and A.3 is complicated and given below. The assertion of Theorem 2.1 now follows from (6.7), Lemma A.2 and A.3. □

6.2 Proof of Lemma A.2.

Without loss of generality we assume $0 < u_0 < x < 1$; the opposite case is treated by exactly the same arguments. Recalling the definition of $A_n(t)$ in (6.5) we have

$$(6.8) \qquad \int_{u_0}^x A_n(z)\,dz = A_{n,1}(x) + A_{n,2}(x),$$

where

$$A_{n,1}(x) = -\int_{m'(u_0)}^{m'(x)} (\hat{\psi}_{h_d}(t) - \psi_{h_d}(t))dt,$$

$$(6.9)$$

$$A_{n,2}(x) = -\int_{m'(u_0)}^{m'(x)} (\psi_{h_d}(t) - m'^{-1}(t))dt,$$

and the non-random quantity ψ_{h_d} is given by

$$\psi_{h_d}(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{m'(v) - u}{h_d}\right) du\,dv$$

Using the definition of $\hat{\psi}_{h_d}$ in (2.3) we derive for the first term the decomposition

$$A_{n,1}(x) = \Big\{ -\int_{m'(u_0)}^{m'(x)} \Big[\frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\Big(\frac{\hat{m}'(v) - u}{h_d}\Big) du\,dv$$

$$-\frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\Big(\frac{m'(v) - u}{h_d}\Big) du\,dv \Big] dt \Big\}$$

$$(6.10) \qquad\qquad = \Delta_n^{(1)}(x) + \frac{1}{2}\Delta_n^{(2)}(x),$$

where

$$(6.11) \qquad \Delta_n^{(1)}(x) = -\frac{1}{h_d^2} \int_{m'(u_0)}^{m'(x)} \int_0^1 \int_{-\infty}^t K_d'\Big(\frac{m'(v) - u}{h_d}\Big) du\,(\hat{m}'(v) - m'(v))\,dv\,dt$$

$$= \frac{1}{h_d} \int_{m'(u_0)}^{m'(x)} \int_0^1 K_d\Big(\frac{m'(v) - t}{h_d}\Big)\,(\hat{m}'(v) - m'(v))\,dv\,dt,$$

$$(6.12) \qquad \Delta_n^{(2)}(x) = -\frac{1}{h_d^3} \int_{m'(u_0)}^{m'(x)} \int_0^1 \int_{-\infty}^t K_d''\Big(\frac{\xi(u,v) - u}{h_d}\Big) du\,(\hat{m}'(v) - m'(v))^2\,dv\,dt,$$

and $|\xi(u,v) - m'(v)| < |\hat{m}'(v) - m'(v)|$. For the first term (6.11) we have (observing that the inequality $0 < u_0 < x < 1$ implies $m'(u_0) < m'(x)$)

$$\Delta_n^{(1)}(x) = \frac{1}{h_d} \int_0^1 \int_{m'(u_0)}^{m'(x)} K_d\Big(\frac{m'(v) - t}{h_d}\Big)\,dt\,(\hat{m}(v) - m(v))'\,dv$$

17

$$= \int_0^1 \int_{\frac{m'(v)-m'(x)}{h_d}}^{\frac{m'(v)-m'(u_0)}{h_d}} K_d(t)\, dt\, (\hat{m}(v) - m(v))'\, dv$$

$$= \int_{m'^{-1}(m'(u_0)+h_d)}^{m'^{-1}(m'(x)-h_d)} \int_{-1}^1 K_d(t)\, dt\, (\hat{m}(v) - m(v))'\, dv$$

$$+ \int_0^{m'^{-1}(m'(u_0)+h_d)} \int_{-1}^{\frac{m'(v)-m'(u_0)}{h_d}} K_d(t)\, dt\, (\hat{m}(v) - m(v))'\, dv$$

$$+ \int_{m'^{-1}(m'(x)-h_d)}^1 \int_{\frac{m'(v)-m'(x)}{h_d}}^1 K_d(t)\, dt\, (\hat{m}(v) - m(v))'\, dv$$

$$(6.13) \qquad = \Delta_n^{(1.1)}(x) + \Delta_n^{(1.2)}(x) + \Delta_n^{(1.3)}(x),$$

where the last line defines the terms $\Delta_n^{(1.j)}$ $(j = 1, 2, 3)$, $\hat{m}(x)$ denotes the Nadaraya-Watson estimate of the regression function defined in (2.2) and we have used the fact that our construction is based on $\hat{m}'(x) = \frac{\partial}{\partial x}\hat{m}(x)$ as an estimate for the derivative of the regression function. A tedious calculation yields

$$(6.14)\quad \Delta_n^{(1.1)}(x) = (\hat{m} - m)(m'^{-1}(m'(x) - h_d)) - (\hat{m} - m)(m'^{-1}(m'(u_0) + h_d))$$

$$= (\hat{m} - m)(x) - h_d(m'^{-1})'(m'(x))(\hat{m} - m)'(x) + \frac{h_d^2}{2}((\hat{m} - m) \circ m'^{-1})''(\zeta_x)$$

$$- (\hat{m} - m)(u_0) - h_d(m'^{-1})'(m'(u_0))(\hat{m} - m)'(u_0) - \frac{h_d^2}{2}((\hat{m} - m) \circ m'^{-1})''(\zeta_u),$$

where $|\zeta_z - m'(z)| \le h_d$ if $z = x$ or $z = u_0$. Similarly, we obtain for sufficiently small h_d

$$(6.15)\quad \Delta_n^{(1.2)}(x) = h_d \int_{-1}^1 (m'^{-1})'(m'(u_0) + h_d v) \int_{-1}^v K_d(t)\, dt\, (\hat{m} - m)'(m'^{-1}(m'(u_0) + h_d v))\, dv$$

$$= h_d \int_{-1}^1 K_d(t) \int_t^1 (m'^{-1})'(m'(u_0) + h_d v)(\hat{m} - m)'(m'^{-1}(m'(u_0) + h_d v))\, dv\, dt$$

$$= \int_{-1}^1 K_d(t) \int_{m'^{-1}(m'(u_0)+h_d t)}^{m'^{-1}(m'(u_0)+h_d)} (\hat{m} - m)'(v)\, dv\, dt$$

$$= \int_{-1}^1 K_d(t) \left[(\hat{m} - m)(m'^{-1}(m'(u_0) + h_d)) - (\hat{m} - m)(m'^{-1}(m'(u_0) + h_d t)) \right] dt$$

$$= (\hat{m} - m)(u_0) + h_d(m'^{-1})'(m'(u_0))(\hat{m} - m)'(u_0) + \frac{h_d^2}{2}((\hat{m} - m) \circ m'^{-1})''(\zeta_u)$$

$$- \int_{-1}^1 K_d(t) \Big\{ (\hat{m} - m)(u_0) + h_d t(m'^{-1})'(m'(u_0))(\hat{m} - m)'(u_0)$$

$$+ \frac{h_d^2 t^2}{2}((\hat{m} - m) \circ m'^{-1})''(\zeta_{u,t}) \Big\} dt$$

$$= h_d(m'^{-1})'(m'(u_0))(\hat{m} - m)'(u_0) + \frac{h_d^2}{2}((\hat{m} - m) \circ m'^{-1})''(\zeta_u)$$

18

$$+h_d^2 \int_{-1}^{1} K_d(t) t^2 ((\hat{m}-m) \circ m'^{-1})'' (\zeta_{u,t}) \, dt,$$

where $|\zeta_u - m'(u_0)| \le h_d$, $|\zeta_{u,t} - m'(u_0)| \le h_d |t| \le h_d$. The remaining term in (6.13) is estimated similarly, that is

$$(6.16) \quad \Delta_n^{(1.3)}(x) = \frac{h_d^2}{2} \int_{-1}^{1} K_d(t) t^2 ((\hat{m}-m) \circ m'^{-1})'' (\zeta_{x,t}) + h_d(m'^{-1})'(m'(x))(\hat{m}-m)'(x)$$
$$-\frac{h_d^2}{2} ((\hat{m}-m) \circ m'^{-1})'' (\zeta_x)$$

where $|\zeta_x - m'(x)| \le h_d$, $|\zeta_{x,t} - m'(x)| \le h_d |t| \le h_d$.

Observing (6.13), (6.14), (6.15) and (6.16) we therefore obtain

$$(6.17) \qquad \Delta_n^{(1)}(x) = (\hat{m}-m)(x) - (\hat{m}-m)(u_0)$$
$$+\frac{h_d^2}{2} \int_{-1}^{1} K_d(t) t^2 ((\hat{m}-m) \circ m'^{-1})'' (\zeta_{u,t}) \, dt$$
$$-\frac{h_d^2}{2} \int_{-1}^{1} K_d(t) t^2 ((\hat{m}-m) \circ m'^{-1})'' (\zeta_{x,t}) \, dt$$
$$= -(\hat{m}-m)(u_0) + (\hat{m}-m)(x) + o\left(\frac{1}{\sqrt{nh_r}}\right),$$

where the second equality follows by the estimate

$$(6.18) \qquad \sup_{u} \left| \hat{m}^{(k)}(u) - m^{(k)}(u) \right| = O\left(\left(\frac{\log h_r^{-1}}{nh_r^{2k+1}} \right)^{1/2} \right), \quad k = 1, 2;$$

which can be shown by simmilar methods as in Mack and Silverman (1982) and assumption (2.11.e). We now derive an estimate for the second term $\Delta_n^{(2)}(x)$ in (6.10). For this we note that a tedious calculation yields

$$(6.19) \qquad \Delta_n^{(2)}(x) = \Delta_n^{(2.1)}(x)(1 + o_p(1)),$$

where

$$(6.20) \quad \Delta_n^{(2.1)}(x) = \frac{1}{h_d^2} \int_{m'(u_0)}^{m'(x)} \int_{0}^{1} K_d'\left(\frac{m'(v)-t}{h_d}\right)(\hat{m}'(v)-m'(v))^2 \, dv dt$$
$$= -\frac{1}{h_d} \int_{0}^{1} \int_{\frac{m'(v)-m'(u_0)}{h_d}}^{\frac{m'(v)-m'(x)}{h_d}} K_d'(t) \, dt \, (\hat{m}'(v)-m'(v))^2 \, dv$$
$$= -\frac{1}{h_d} \int_{0}^{1} \left[K_d\left(\frac{m'(v)-m'(x)}{h_d}\right) - K_d\left(\frac{m'(v)-m'(u_0)}{h_d}\right) \right] (\hat{m}'(v)-m'(v))^2 \, dv$$
$$= -\int_{-1}^{1} K_d(v) (m'^{-1})'(m'(x)+h_dv)(\hat{m}'-m')^2 (m'^{-1}(m'(x)+h_dv)) dv$$
$$+\int_{-1}^{1} K_d(v) (m'^{-1})'(m'(u_0)+h_dv)(\hat{m}'-m')^2 (m'^{-1}(m'(u_0)+h_dv)) dv$$

19

Observing that with the assumptions (2.11.a), (2.11.b) and (2.11.d) for the bandwidth h_r the mean squared error between the derivative of the Nadaraya-Watson estimate and the derivative of the regression function is of order $O\left(\frac{1}{nh_r^3}\right)$ we therefore obtain

$$(6.21) \qquad\qquad \mathrm{E}\left|\Delta_n^{(2.1)}(x)\right| \;=\; O\left(\frac{1}{nh_r^3}\right).$$

Finally, for the term $A_{n,2}(x)$ in (6.9) it follows that

$$A_{n,2}(x) = -h_d^2 \left\{ \kappa_2\left(K_d\right)\left(\frac{1}{m''(x)} - \frac{1}{m''(u_0)}\right) + o\left(1\right) \right\},$$

which implies

$$(6.22) \qquad\qquad \sqrt{nh_r}\,A_{n,2}(x) = -\frac{h_d^2}{h_r^3}\sqrt{nh_r^7}O\left(1\right) = o\left(1\right)$$

Now a combination of (6.8), (6.10), (6.13), (6.17), (6.19), (6.21) and (6.22) yields the assertion of Lemma A.1.

$$\square$$

6.3 Proof of Lemma A.3.

Recalling the definition of B_{n,λ^*} in (6.6) we obtain

$$(6.23) \qquad \int_{u_0}^x B_{n,\lambda^*}\left(z\right)dz = \int_{u_0}^x B_{n,0}(z)dz(1+o_p(1)) = \left\{2B_n^{(1)}\left(x\right) - B_n^{(2)}\left(x\right)\right\}(1+o_p(1)),$$

where

$$(6.24) \qquad\qquad B_n^{(1)}(x) \;=\; \int_{m'(u_0)}^{m'(x)} \frac{(\hat\psi_{h_d} - m'^{-1})(\hat\psi_{h_d} - m'^{-1})'}{(m'^{-1})'}(t)dt,$$

$$(6.25) \qquad\qquad B_n^{(2)}(x) \;=\; \int_{m'(u_0)}^{m'(x)} \frac{(\hat\psi_{h_d} - m'^{-1})^2(m'^{-1})''}{\{(m'^{-1})'\}^2}(t)dt.$$

We now prove that the expressions $\sqrt{nh_r}B_n^{(i)}$ $(i=1,2)$ converge to 0 in L^1. For this we exemplarily consider the term $B_n^{(1)}$, the second term is treated by exactly the same arguments. Note that

$$(6.26) \qquad\qquad B_n^{(1)}\left(x\right) \;=\; B_n^{(1.1)}\left(x\right) + B_n^{(1.2)}\left(x\right) + B_n^{(1.3)}\left(x\right) + B_n^{(1.4)}\left(x\right),$$

where

$$(6.27) \qquad\qquad B_n^{(1.1)}\left(x\right) \;=\; \int_{m'(u_0)}^{m'(x)} \frac{(\hat\psi_{h_d} - \psi_{h_d})(\hat\psi_{h_d} - \psi_{h_d})'}{(m'^{-1})'}(t)dt,$$

$$(6.28) \qquad\qquad B_n^{(1.2)}\left(x\right) \;=\; \int_{m'(u_0)}^{m'(x)} \frac{(\psi_{h_d} - m'^{-1})(\hat\psi_{h_d} - \psi_{h_d})'}{(m'^{-1})'}(t)dt,$$

20

$$(6.29) \qquad B_n^{(1.3)}(x) = \int_{m'(u_0)}^{m'(x)} \frac{(\hat{\psi}_{h_d} - \psi_{h_d})(\psi_{h_d} - m'^{-1})'}{(m'^{-1})'}(t)dt,$$

$$(6.30) \qquad B_n^{(1.4)}(x) = \int_{m'(u_0)}^{m'(x)} \frac{(\psi_{h_d} - m'^{-1})(\psi_{h_d} - m'^{-1})'}{(m'^{-1})'}(t)dt.$$

For the term $B_n^{(1.1)}(x)$ we derive the further decomposition

$$(6.31) \qquad B_n^{(1.1)}(x) = \left\{ B_n^{(1.1.1)}(x) - B_n^{(1.1.2)}(x) - B_n^{(1.1.3)}(x) + B_n^{(1.1.4)}(x) \right\}(1 + o_p(1))$$

with

$$(6.32) \qquad B_n^{(1.1.1)}(x) = \frac{1}{h_d^3} \int_{m'(u_0)}^{m'(x)} \frac{1}{(m'^{-1})'(t)} \left\{ \int_0^1 K_d\left(\frac{m'(v) - t}{h_d}\right)(\hat{m}' - m')(v)\, dv \right\}$$
$$\times \left\{ \int_0^1 K_d'\left(\frac{m'(w) - t}{h_d}\right)(\hat{m}' - m')(w)\, dw \right\} dt,$$

$$(6.33) \qquad B_n^{(1.1.2)}(x) = \frac{1}{2h_d^4} \int_{m'(u_0)}^{m'(x)} \frac{1}{(m'^{-1})'(t)} \left\{ \int_0^1 K_d\left(\frac{m'(v) - t}{h_d}\right)(\hat{m}' - m')(v)\, dv \right\}$$
$$\times \left\{ \int_0^1 K_d''\left(\frac{m'(w) - t}{h_d}\right)(\hat{m}' - m')^2(w)\, dw \right\} dt,$$

$$(6.34) \qquad B_n^{(1.1.3)}(x) = \frac{1}{2h_d^4} \int_{m'(u_0)}^{m'(x)} \frac{1}{(m'^{-1})'(t)} \left\{ \int_0^1 K_d'\left(\frac{m'(v) - t}{h_d}\right)(\hat{m}' - m')(v)\, dv \right\}$$
$$\times \left\{ \left(\int_0^1 K_d'\left(\frac{m'(w) - t}{h_d}\right)(\hat{m}' - m')^2(w)\, dw \right\} dt,$$

$$(6.35) \qquad B_n^{(1.1.4)}(x) = \frac{1}{4h_d^5} \int_{m'(u_0)}^{m'(x)} \frac{1}{(m'^{-1})'(t)} \left\{ \int_0^1 K_d'\left(\frac{m'(v) - t}{h_d}\right)(\hat{m}' - m')^2(v)\, dv \right\}$$
$$\times \left\{ \int_0^1 K_d''\left(\frac{m'(w) - t}{h_d}\right)(\hat{m}' - m')^2(w)\, dw \right\} dt.$$

A tedious calculation yields for sufficiently small h_d

$$B_n^{(1.1.1)}(x) = \frac{1}{h_d} \int_{m'(u_0)}^{m'(x)} \frac{1}{(m'^{-1})'(t)} \left\{ \int_{-1}^1 K_d(v)(m'^{-1})'(t + h_d v)(\hat{m}' - m')(m'^{-1}(t + h_d v))dv \right\}$$
$$\times \left\{ \int_{-1}^1 K_d'(w)(m'^{-1})'(t + h_d w)(\hat{m}' - m')(m'^{-1}(t + h_d w))dw \right\} dt$$
$$= \int_{m'(u_0)}^{m'(x)} \frac{1}{(m'^{-1})'(t)} \left(\int_{-1}^1 K_d(v) \left\{ (m'^{-1})'[(\hat{m}' - m') \circ m'^{-1}](t) \right. \right.$$
$$+ \left. h_d v \left[(m'^{-1})''[(\hat{m}' - m') \circ m'^{-1}] + (m'^{-1})'^2[(\hat{m}'' - m'') \circ m'^{-1}] \right](\xi_{v,t}) \right\} dv \right)$$
$$\times \left(\int_{-1}^1 w K_d'(w) \left[(m'^{-1})''[(\hat{m}' - m')] \circ m'^{-1} + (m'^{-1})'^2[(\hat{m}'' - m'')] \circ m'^{-1} \right](\xi_{w,t})\, dw \right) dt$$

21

$$= \int_{m'(u_0)}^{m'(x)} (\hat{m}' - m') \circ m'^{-1}(t) \int_{-1}^{1} wK'_d(w)(m'^{-1})''[(\hat{m}' - m') \circ m'^{-1}](\xi_{w,t}) \, dw dt$$

$$- \frac{1}{2m''} (\hat{m}' - m')^2(z) \Big|_{u_0}^{x} - \frac{1}{2} \int_{u_0}^{x} \frac{m'''}{m''^2} (\hat{m}' - m')^2(z) \, dz$$

$$+ h_d \lambda \int_{m'(u_0)}^{m'(x)} (\hat{m}' - m') \circ m'^{-1}(t) \int_{-1}^{1} w^2 K'_d(w)$$

$$\times \left[2(m'^{-1})''(m'^{-1})'[(\hat{m}'' - m'') \circ m'^{-1}] + (m'^{-1})'^{\,3}[(\hat{m}''' - m''') \circ m'^{-1}] \right] (\tilde{\xi}_{w,t}) dw dt$$

$$+ h_d \int_{m'(u_0)}^{m'(x)} \frac{1}{(m'^{-1})'(t)} \Big(\int_{-1}^{1} v K_d(v) \left[(m'^{-1})''[(\hat{m}' - m') \circ m'^{-1}] \right.$$

$$\left. + (m'^{-1})'^{\,2}[(\hat{m}'' - m'') \circ m'^{-1}] \right] (\xi_{v,t}) \, dv \Big)$$

$$\times \Big(\int_{-1}^{1} w K'_d(w) \left[(m'^{-1})''[(\hat{m}' - m') \circ m'^{-1}] + (m'^{-1})'^{\,2}[(\hat{m}'' - m'') \circ m'^{-1}] \right] (\xi_{w,t}) \, dw \Big) dt \; ,$$

where we used the fact that

$$\int_{-1}^{1} K'_d(u) du = 0 \; , \quad \int_{-1}^{1} u K'_d(u) du = -1$$

at several places. Using the estimate (6.18) this gives for the expectation

$$\mathrm{E} \left| B_n^{(1.1.1)}(x) \right| \leq O\left(\frac{\log h_r^{-1}}{n h_r^3} \right) + O\left(\frac{h_d \log h_r^{-1}}{n h_r^4} \right) + O\left(\frac{h_d \log h_r^{-1}}{n h_r^5} \right)$$

$$= O\left(\frac{\log h_r^{-1}}{n h_r^3} \right) + O\left(\frac{h_d \log h_r^{-1}}{n h_r^5} \right) = o\left(\frac{1}{\sqrt{n h_r}} \right),$$

where we have used assumptions (2.11.d) and (2.11.e) for the last estimate. The other terms in (6.26) are treated similarly, that is

$$\mathrm{E} \left| B_n^{(1.1.2)}(x) \right| \leq \frac{1}{h_d} \left[O\left(\frac{(\log h_r^{-1})^{3/2}}{n^{3/2} h_r^{9/2}} \right) + O\left(\frac{(\log h_r^{-1})^{3/2}}{n^{3/2} h_r^{11/2}} \right) \right] = o\left(\frac{1}{\sqrt{n h_r}} \right),$$

$$\mathrm{E} \left| B_n^{(1.1.3)}(x) \right| \leq \left[O\left(\left(\frac{\log h_r^{-1}}{n h_r^3} \right)^{1/2} \right) + O\left(\left(\frac{\log h_r^{-1}}{n h_r^5} \right)^{1/2} \right) \right] \left[O\left(\frac{\log h_r^{-1}}{n h_r^3} \right) + O\left(\frac{\log h_r^{-1}}{n h_r^4} \right) \right]$$

$$= O\left(\frac{(\log h_r^{-1})^{3/2}}{n h_r^6 \sqrt{n h_r}} \right) = o\left(\frac{1}{\sqrt{n h_r}} \right)$$

$$\mathrm{E} \left| B_n^{(1.1.4)}(x) \right| \leq \frac{1}{h_d^2} \left(O\left(\frac{\log h_r^{-1}}{n h_r^3} \right) + O\left(\frac{\log h_r^{-1}}{n h_r^4} \right) \right) O\left(\frac{\log h_r^{-1}}{n h_r^3} \right) = o\left(\frac{1}{\sqrt{n h_r}} \right).$$

Furthermore it can be shown by similar methods as used above that

$$B_n^{(1.i)}(x) = o\left(\frac{1}{\sqrt{n h_r}} \right)$$

for $i = 2, 3, 4$ and the assertion of Lemma A.3 follows.

$$\square$$

6.4   Proof of Theorem 3.2.

Note that the function
$$\Delta(x) := m(x) - \Psi_{hd}(x)$$
is continuous and consequently there exists at least one solution of equation (3.7). Moreover, observing the definition (3.1) and the fact $\Psi'_{h_d} = \psi_{h_d}^{-1}(m')$ we have
$$m_C(x, u_0) = m(u_0) - \Psi_{h_d}(u_0) + \Psi_{h_d}(x)$$
which gives

(6.36)
$$\min_{u_0 \in [0,1]} \int_0^1 (m_C(x, u_0) - m(x))^2 dx = \min_{u_0 \in [0,1]} \int_0^1 (\Delta(u_0) - \Delta(x))^2 dx,$$
$$= \min_{g \in G} \int_0^1 (g - \Delta(x))^2 dx,$$

where the set G is defined by $G = \Delta([0, 1])$. Now
$$\arg\min_{g \in \mathbb{R}} \int_0^1 (g - \Delta(x))^2 dx = \int_0^1 \Delta(x) dx$$

and by the first part of the proof there exists at least one point $u_0^* \in [0, 1]$ such that

(6.37)
$$\int_0^1 \Delta(x) dx = \Delta(u_0^*).$$

Consequently, we obtain from (6.36)

(6.38)
$$\arg\min_{u_0 \in [0,1]} \int_0^1 (m_C(x, u_0) - m(x))^2 dx = u_0^*$$

for any point satisfying (6.37), which proves the first part of the assertion. For a proof of the remaining part we note that we obtain for any u_0^* satisfying (6.38) [or equivalently (3.7) or (6.37)]

$$m_C(x, u_0^*) = m(u_0^*) + \int_{u_0}^x \psi_{h_d}^{-1}(m')(t) dt$$
$$= m(u_0^*) - \Psi_{h_d}(u_0^*) + \Psi_{h_d}(x)$$
$$= \int_0^1 [m(u) - \Psi_{h_d}(u) + \Psi_{h_d}(x)] du$$
$$= \int_0^1 m_C(x, u) du ,$$

which completes the proof of Theorem 3.2.   □

References

Y. Aït-Sahalia, J. Duarte (2003). Nonparametric Option Pricing under Shape Restrictions. Journal of Econometrics, 116, 9-47

R.E. Barlow, D.J. Bartholomew, J.M. Bremner, H.D. Brunk (1972). Statistical inference under order restrictions: The theory and application of isotonic regression. Wiley, New York.

C. Bennett, R. Sharpley (1988). Interpolation of Operators. Academic Press, N.Y.

M. Birke, H. Dette, (2005). A note on estimating a monotone regression function by combining kernel and density estimates. Technical report, Department of Mathematics.
http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm

F. Black, M. Scholes (1973). The pricing of options and corporate liabilities, J. Polit. Econ. 81, 637-654.

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

Cox, J.C. & Ingersoll, J.E. & Ross, S.A. (1985). A theory of the term structure of interest rates, Econometrica 53, 385-407.

H. Dette, N. Neumeyer, K.F. Pilz (2005). A simple nonparametric estimator of a monotone regression function. Technical report, Department of Mathematics.
http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm

M. L. Dudzinski, R. Mykytowycz (1961). The eye lens as an indicator of age in the wild rabbit in Australia. CSIRO Wildlife Research, 6, 156-159.

R.L. Dykstra (1983). An algorithm for restrices least squares regression. J. Amer. Statist. Assoc. 78, 837-842.

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

D.A.S. Fraser, H. Massam (1989). A mixed primal-dual bases algorithm for regression under inequality constraints. Application to concave regression. Scand. J. Statist. 16, 65-74.

T. Gasser, H.-G. Müller, V. Mammitzsch (1985). Kernels for nonparametric curve estimation. J. Roy. Statist. Soc. Ser. B 47, 238-252.

P. Groeneboom, G. Jongbloed, J.A. Wellner (2001). Estimation of convex functions: characterizations and asymptotic theory. Ann. Statist. 26, 1653-1698.

W. Härdle (1991). Applied nonparametric regression. Cambridge University Press, Cambridge.

S.P. Han (1988). A successive projection method. Math. Programming 40, 1-14.

P.L. Hanson, G. Pledger (1976). Consistency in concave regression. Ann. Statist. 4, 1038-1050.

C. Hildreth (1954). Point estimates of ordinates of concave functions. J. Amer. Statist. Assoc. 49, 598-619.

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). Nonparametric regression under qualitative smoothness assumptions. Ann. Statist. 19, 741-759.

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

E.A. Nadaraya (1964). On estimating regression. Theory Probab. Appl. 9, 141-142

D.A. Ratkowsky (1983). Nonlinear Regression Modeling. Marcel Dekker Inc., New York

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

T. Robertson, F.T. Wright, R.L. Dykstra (1988). Order restricted statistical inference. Wiley Series in Probability and Mathematical Statistics. Chichester (UK)

J.V. Ryff (1965). Orbits of L_1 under doubly stochastic transformations. Trans. Am. Math. Soc. 117, 92-100.

J.V. Ryff (1970). Measure preserving transformations and rearrangements. J. Math. Anal. Appl. 31, 449-458.

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

C.F. Wu (1982). Some algorithms for concave and isotonic regression. Stud. Management. Sci. 19, 105-116.