A note on estimating a smooth monotone regression by combining kernel and density estimates

Melanie Birke
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: melanie.birke@rub.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de
FAX: +49 234 3214 559

July 3, 2008

Abstract

In a recent paper Dette, Neumeyer and Pilz (2006) proposed a new nonparametric estimate of a smooth monotone regression function. This method is based on a non-decreasing rearrangement of an arbitrary unconstrained nonparametric estimator. Under the assumption of a twice continuously differentiable regression function the estimate is first order asymptotic equivalent to the unconstrained estimate and other type of smooth monotone estimates. In this note we provide a more refined asymptotic analysis of the monotone regression estimate. It is shown that, compared to the unconstrained estimate, the new monotone estimate asymptotically reduces the L^p-error, if the "true" regression function is isotone for any $p \geq 1$. Moreover, in the case, where the regression function is increasing but only once continuously differentiable we prove asymptotic normality of an appropriately standardized version of the estimate, where the asymptotic variance is of order $n^{-2/3-\varepsilon}$, the bias is of order $n^{-1/3+\varepsilon}$ and $\varepsilon > 0$ is small. Therefore the rate of convergence of the new estimate does not coincide with the rate of the estimate obtained from monotone least squares estimation, but the asymptotic distribution of the new estimate is simpler. Additionally, if the derivative of the regression function is Hölder continuous the rate of convergence of the new estimate is faster than $n^{-1/3}$.

AMS Subject Classification: 62G05, 62G20

Keywords and Phrases: order restricted inference, monotone estimation, greatest convex minorant, Nadaraya-Watson estimate, non-decreasing rearrangements

1 Introduction

One of the most important problems in applied statistics is the estimation of relationships among observable variables. Often a specific parametric form of a regression model cannot be postulated and nonparametric estimation methods have become increasingly popular in recent years. 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. Typical examples appear in economics where monotonicity applies to production, profit and cost function [see e.g. Matzkin (1994), Aït-Sahalia and Duarte (2003) among others] or in medicine where the probability of contracting a certain disease depends monotonically on certain factors. Since the early work of Brunk (1955) numerous authors have proposed monotone estimates of the regression function [see e.g. Cheng and Lin (1981), Wright (1982), Mukerjee (1988), Mammen (1991) and Friedman and Tibshirani (1984), Ramsay (1988), Kelly and Rice (1990), Mammen and Thomas-Agnan (1999), Mammen, Marron, Turlach and Wand (2001) and Hall and Huang (2001) among many others]. We refer the interested reader to the nice reviews of the literature by Delecroix and Thomas-Agnan (2000) and Gijbels (2005).

In a recent paper Dette, Neumeyer and Pilz (2006) introduced an alternative monotone estimate of a smooth and strictly increasing regression function, which is based on the concept of non-decreasing rearrangements [see Hardy, Littlewood and Pólya (1952) or Lorentz (1953)]. This method is called density regression estimate, because it combines a density and regression estimator. In a first step an estimate of the inverse of the monotone regression function is constructed using an integrated density estimator based on data from a preliminary regression estimate, while the final estimate is obtained by an inversion of the function obtained from the first step. If the regression function is twice continuously differentiable asymptotic normality of an appropriately standardized estimate with rate $n^{-2/5}$ can be proved, where n denotes the sample size. If the bandwidths are chosen appropriately it is also shown that the new estimate is first order asymptotic equivalent to a smoothed version of a monotone least squares estimate as considered by Mukerjee (1988) or Mammen (1991).

The present paper has two purposes. On the one hand we provide further insight in the statistical properties of the estimate of Dette et al. (2006). In particular we demonstrate that a smooth rearrangment of the unconstrained estimator reduces asymptotically the L^p-estimation error if the true regression function is increasing. On the other hand we investigate the properties of this estimate in the case where the regression function is only once continuously differentiable. We derive the asymptotic distribution of the density regression estimate under this assumption and show that it differs from the asymptotic distribution of the monotone least squares estimate (which is not smooth). For this estimate Brunk (1955) showed that an appropriately normalized version converges weakly with rate $n^{-1/3}$ to a random variable, which is defined as the slope at the point 0 of the greatest convex minorant of the process $W(t)+t^2$, where W is a two sided Wiener-Levy process [see also Robertson, Wright and Dykstra (1989), Theorem 9.2.4]. If additional smoothness is added, the density regression estimate is again asymptotically normal distributed with rate $n^{-2/5}$ [see e.g. Mammen (1991)]. By an appropriate choice of the smoothing parameters in the estimate of Dette et al. (2006) we show in the present paper that in the case of a once continuously differentiable regression function the density regression estimate is still asymptotically normal distributed, where the variance is of order $n^{-2/3-\varepsilon}$, the bias is of order $n^{-1/3+\varepsilon}$ and $\varepsilon > 0$ is small. The larger rate of the mean squared error can be considered as a price, which has to be paid to preserve asymptotic normality of the isotone estimate.

The paper is organized as follows. In Section 2 we briefly review the estimate of Dette et al. (2006). Section 3 contains our main results, and we establish asymptotic normality of the density regression estimate in the case of a once continuously differentiable regression function. We also prove uniform almost sure consistency of the estimate in this case and compare the asymptotic results for a once continuously differentiable regression function to the case where the regression function is at least twice continuously differentiable. In Section 4 we present a simulation study in order to illustrate the finite sample properties of the new estimate. We also compare the L^p-estimation error of the new procedure with the S+I and I+S estimator discussed in Mukerjee (1988) and Mammen (1991) and demonstrate that the density regression estimate has a smaller L^p-estimation error than these estimates. It is also shown that the estimator discussed here is less sensitive to undersmoothing than the unconstrained estimator. Finally, all proves are defered to an appendix.

2 Monotone smoothing by inversion

Consider the nonparametric regression model

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

where $\{(X_i, Y_i)\}_{i=1}^n$ is a bivariate sample of i.i.d. observations such that the random variables X_i are located in the interval $[0, 1]$ and have a continuous density f. The random variables ε_i are also assumed as i.i.d. with zero mean, existing fourth moment and independent of the $\{X_i\}_{i=1}^n$. The regression function m is assumed to be strictly monotone and further assumptions which are required for our main asymptotic statements will be presented in the following section (these are not needed for the definition of the monotone estimate). For the sake of transparency we will restrict ourselves to the problem of estimating a strictly increasing regression function, but the antitone case can be treated exactly in the same way. Following Dette et al. (2006) we consider a transformation of the regression function defined by

$$m_I^{-1}(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d \left(\frac{m(v) - u}{h_d} \right) du dv, \tag{2.2}$$

where K_d is a given density and h_d denotes a bandwidth converging to 0 with increasing sample size. Note that m_I^{-1} is isotone even if the function m is not isotone. Therefore we can calculate its inverse, which will be denoted by m_I throughout this paper. Because the regression function in (2.1) is unknown, we use in the following as approximation its nonparametric estimate $\hat{m}$. In principle any estimate could be used here, but for the sake of simplicity we restrict ourselves to the Nadaraya-Watson estimate

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

where K_r is a further kernel and h_r a second bandwidth. The estimate of m_I^{-1} is then obtained as

$$\hat{m}_I^{-1}(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^{t} K_d\left(\frac{\hat{m}(v) - u}{h_d}\right) du dv, \tag{2.4}$$

and the isotone estimate of the regression function is finally defined as the inverse of the function $\hat{m}_I^{-1}$ and denoted by $\hat{m}_I$. Note that for computational reasons Dette et al. (2006) replaced the integral with respect to dv in (2.4) by a discrete approximation of the Riemann integral, but in this paper we will work with the representation (2.4) for the sake of simplicity. It is easy to see that all results presented in this paper remain true, if the integral with respect to dv is replaced by its discrete approximation as considered in Dette et al. (2006).

Note that the derivative of the expression (2.2) with respect to the variable t corresponds to the expectation of a kernel density estimate of an i.i.d. sample of the random variable $m(U)$, where U denotes a random variable with a uniform distribution on the interval $[0, 1]$. This justifies our notation K_d and h_d in (2.2), where the index d corresponds to the phrase density. Similarly, the index r in (2.3) reveals the fact that $\hat{m}$ is an estimate of the regression function. For this reason we will also call $\hat{m}_I$ density regression estimate in the following discussion.

Remark 2.1. As pointed out by a referee the function $\tilde{m}_I$ might exhibit a non intuitive behaviour in the case where the function m is decreasing. For example, if $m(x) = 1 - x$, it follows $\tilde{m}_I(x) = x$, while the best isotonic L^p-approximation of m by non-decreasing function is given by $m^*(x) = \frac{1}{2}$. This indicates that the proposed estimator cannot be used to test whether the regression is really isotonic, as opposed to the best L^p-isotonic approximation. On the other hand, it can be used to test strict isotonicity [see Birke and Dette (2007)]. Moreover, the new estimate might have advantages for data with an irregular behaviour in the tails. In such cases, an initial isotonization (and eventually smoothing in a second step) will not fully correct this effect, but procedures smoothing and then isotonizing are likely to produce better results.

3 A refined asymptotic analysis

If m and f are twice continuously differentiable, Dette et al. (2006) proved the asymptotic normality of the density regression estimate and we refer the reader to this article for details. In particular these authors showed asymptotic normality of the random variable $\left(\hat{m}_I^{-1}(t) - \mathrm{E}\left[\hat{m}_I^{-1}(t)\right]\right)$. This result is then used to establish the asymptotic normality of $(\hat{m}_I(t) - \mathrm{E}[\hat{m}_I(t)])$. Their results imply, that in the case $h_d = o(h_r)$ the isotone estimate is first order asymptotic equivalent to the Nadaraya-Watson estimate. In the following we will demonstrate that in the case of a once continuously differentiable regression function a standardization of order $\sqrt{nh_d}$ is required and that the condition

$$\lim_{h_d \to 0, h_r \to 0} \frac{h_d}{h_r} = \infty \tag{3.1}$$

is sufficient (among other technical assumptions) to obtain asymptotic normality of the statistic

$$\sqrt{nh_d}\left(\hat{m}_I^{-1}(t) - \mathrm{E}\left[\hat{m}_I^{-1}(t)\right]\right). \tag{3.2}$$

We will then use this result and a result on the uniform convergence of the estimates $\hat{m}_I^{-1}$ and $(\hat{m}_I^{-1})'$ to obtain asymptotic normality of the monotone estimate $\hat{m}_I$. The derivation of our asymptotic results requires a substantially more refined analysis as given in Dette et al. (2006). In particular we need the following basic assumptions

(V1) The random variables $\{X_i\}_{i=1,\dots,n}$ are i.i.d. with positive density $f : [0,1] \to \mathbb{R}^+$, such that $f \in C^1([0,1])$.

(V2) The random variables $\{\varepsilon_i\}_{i=1,\dots,n}$ are i.i.d. with $\mathrm{E}[\varepsilon_i] = 0$, $\mathrm{E}[\varepsilon_i^2] = 1$ and $\mathrm{E}[\varepsilon_i^4] < \infty$. Moreover, the sequence of the ε_i is independent of the sequence of the X_i.

(V3) The regression function $m : [0,1] \to \mathbb{R}$ is strictly increasing and $m \in C^1([0,1])$.

(V4) The variance function $\sigma : [0,1] \to \mathbb{R}^+$ is continuous.

(W1) The kernel K_r has compact support given by the interval $[-1,1]$ and $K_r \in C^1([-1,1])$.

(W2) The kernel K_d is symmetric, twice continuously differentiable, of order 2 and has compact support given by the interval $[-1,1]$. Moreover $K_d(1) = K_d(-1) = 0$ and K_d'' is bounded away from zero.

(W3) The bandwidths h_r and h_d of the density regression estimate satisfy $h_r, h_d \to 0$, $nh_r, nh_d \to \infty$ as $n \to \infty$, and additionally we assume that the following relations hold

$$\begin{aligned} h_r &= o(h_d) \\ nh_d^{3/(1-3\varepsilon)} &= O(1), \quad \text{for some } 0 < \varepsilon < \frac{1}{12}, \\ nh_r^3 &= O(1), \\ \frac{\log h_r^{-1}}{nh_r^2 h_d} &= o(1) \end{aligned}$$

Our first result shows that the kernel density estimate reduces asymptotically the L^p-error.

Theorem 3.1. *If the assumptions (V1)-(V4), (W1)-(W3) are satisfied and the regression function m is $q \geq 1$ times continuously differentiable, then*

$$\Big(\int_0^1 |\hat{m}_I(x) - m(x)|^p dx\Big)^{1/p} \leq \Big(\int_0^1 |\hat{m}(x) - m(x)|^p dx\Big)^{1/p} + O(D_n^{1/p})$$

where

$$D_n = \alpha_1 \; C_n + \alpha_2 \frac{1}{h_d} C_n^2 + \alpha_3 \frac{1}{h_d^3} C_n^3 + \alpha_4 h_d^q$$

and $C_n = ||\hat{m} - m||_\infty$.

For the Nadaraya-Watson estimate we have $C_n = O(\log h_r^{-1}/nh_r)^{1/2}$ and under the specified assumptions on the bandwidth h_d the remainder D_n converges to 0 for $n \to \infty$. The asymptotic reduction of the estimation error for $\hat{m}_I$ can already be observed for moderate sample sizes (see the examples in Section 4).

Theorem 3.2. *If the assumptions (V1)-(V4), (W1)-(W3) are satisfied, then it follows for any $t \in (m(0), m(1))$ with $m'(m^{-1}(t)) > 0$*

$$\sqrt{nh_d}\left(\hat{m}_I^{-1}(t) - m^{-1}(t) + h_r a_{K_d,K_r}(t) - h_d \, b_{K_d}(t)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, g^2(t)\right),$$

where a_{K_d,K_r} and b_{K_d} are given by (A.8) and (A.6) in the Appendix, respectively, and

$$g^2(t) \quad = \quad \frac{\sigma^2(m^{-1}(t))}{f(m^{-1}(t))\, m'(m^{-1}(t))} \int_{-1}^1 K_d^2(y)\, dy \tag{3.3}$$

Note that the final monotone estimate of the regression function is obtained by an inversion of the function $\hat{m}_I^{-1}$. Dette et al. (2006) investigated the properties of the operator which maps a strictly increasing function onto a given quantile by a functional delta method assuming a twice continuously differentiable regression function. In the case where $m \in C^1([0,1])$ this argument is not applicable any more and we replace it by using the fact that the estimate $\hat{m}_I^{-1}$ converges uniformly to m^{-1} on proper subsets of the interval $(m(0), m(1))$. This statement is precisely formulated in the following theorem and of own interest.

Theorem 3.3. *Assume that $(V1)-(V4), (W1)-(W3)$ are satisfied and that $m'(m^{-1}(t)) > 0$ for all $t \in (m(0), m(1))$. Let $\delta > 0$, define $J := J(\delta) = [m(0) + \delta, m(1) - \delta]$, then*

$$\sup_{t \in J} \left|\hat{m}_I^{-1}(t) - m^{-1}(t)\right| = O\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2} + o(h_d) \quad a.s.$$

$$\sup_{t \in J} \left|\left(\hat{m}_I^{-1}\right)'(t) - \left(m^{-1}\right)'(t)\right| = O\left(\frac{\log h_r^{-1}}{nh_r h_d^2}\right)^{1/2} + o(1) \quad a.s.$$

For the statement of the final result of this section we define for any $\delta > 0$, $\eta > 0$ the set

$$I(\eta) := \left[m^{-1}(m(0)+\delta)+\eta, m^{-1}(m(1)-\delta)-\eta\right] \tag{3.4}$$

Theorem 3.4. *Assume that the assumptions (V1)-(V4), (W1)-(W3), $h_d/h_r \to \infty$ are satisfied, then it follows for any $x \in I(\eta)$ with $m'(x) > 0$*

$$\sqrt{nh_d}\left(\hat{m}_I(x) - m(x) - h_r a_{K_d,K_r}(m(x))\, m'(x) + h_d b_{K_d}(m(x))\, m'(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, s^2(x)\right) \ ,$$

where b_{K_d} and $a_{K_d,K_r}(t)$ are defined in equation (A.6) and (A.8) in the Appendix, respectively and

$$s^2(x) = \frac{\sigma^2(x)m'(x)}{f(x)} \int_{-1}^{1} K_d^2(y)dy.$$

Remark 3.5. Note that the result of Theorem 3.4 requires the condition $h_r = o(h_d)$, which is used at several steps in the proofs in the Appendix. This assumption differs from the choice proposed by Dette et al. (2006), who recommended $h_d = o(h_r)$ in the case of a twice continuously differentiable regression function. We were not able to derive an asymptotic law if $\lim_{h_d,h_r\to 0} h_d/h_r = c \in [0,\infty)$ and $m \in C^1([0,1])$, because a proof of the corresponding statements requires various contradicting conditions regarding the bandwidths h_d and h_r. Consequently, from an asymptotic point of view - in contrast to the case considered by Dette et al. (2006) - the parameter h_d cannot be considered as a secondary parameter. On the other hand simulation results show, that the impact of the smoothing parameter h_d on the performance of the estimate is much smaller compared to the smoothing parameter h_r of the initial unconstrained estimate. From our numerical experience we recommend

$$h_r = cn^{-1/3}, \tag{3.5}$$

where the constant c is determined from the data (e.g. by cross validation) and the bandwidth h_d should be chosen slightly larger.

Remark 3.6. In the remaining part of this paper we discuss the case, where the bandwidth h_r of the regression estimate is chosen by (3.5), which corresponds to the optimal rate (with respect to mean squared error) for the initial unconstrained estimate. In this case condition (W3) yield for the bandwidth in the density step $h_d = n^{-1/3}\alpha_n$, where the sequence α_n converges to infinity such that $\alpha_n = O(n^{\varepsilon})$ for some $0 < \varepsilon < \frac{1}{12}$, $\log n = o(\alpha_n)$. Now the statement of Theorem 3.4 simplifies to

$$n^{1/3}\sqrt{\alpha_n}\{\hat{m}_I(x) - m(x)\} - \alpha_n^{1/2} a_{K_d,K_r}(m(x))m'(x) + \alpha_n^{3/2} b_{K_d}(m(x))m'(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x)),$$

for any x with $m'(x) > 0$. In particular, if $\alpha_n = n^\varepsilon$ and $\varepsilon > 0$ is sufficiently small, this gives the order $O(n^{-2/3-\varepsilon})$ for the variance and $o(n^{-1/3+\varepsilon})$ for the bias. For the isotone least squares estimate the order of the mean squared error is $O(n^{-2/3})$. Therefore the slightly larger order of the mean squared error of $\hat{m}_I$ can be considered as a price which has to be paid to obtain an asymptotically normal distributed estimate.
Finally, if m' is Hölder continuous of order γ and the constant ε satisfies $\varepsilon < \frac{2\gamma}{9+6\gamma}$, then

$$n^{1/3}\alpha_n^{1/2}(\hat{m}_I(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x))$$

and the rate of the density regression estimate is faster than $n^{-1/3}$.

Remark 3.7. A careful inspection of the proofs in the Appendix shows that the assertion of Theorem 3.2 - 3.4 remain correct if the assumptions (V3) and (V4) are satisfied in a neighborhood of the point $x \in (0, 1)$.

4 A simulation study

In this section we present a small simulation study in order to illustrate the finite sample properties of the density regression estimate and to compare it with the unconstrained estimate and the monotone (S+I and I+S) estimates proposed by Mukerjee (1988) and Mammen (1991), which are most similar in spirit with the density regression estimate proposed in this paper. In order to avoid the domination by boundary effects we use the local linear estimate as initial estimate $\hat{m}$ in the density regression estimate and in the estimates S+I and I+S. In a first example we investigate the L^p-estimation error of the unconstrained local linear estimate $\hat{m}$, the isotone estimate $\hat{m}_I$ proposed in this paper and the S+I and I+S estimates $\hat{m}_{SI}$ and $\hat{m}_{IS}$ discussed in Mukerjee (1988) and Mammen (1991). For computing $\hat{m}_{SI}$, in a first step an unconstrained local linear estimator of m is calculated which is monotonized in a second step by the PAVA algorithm [see e.g. Barlow, Bartholomew, Bremner and Brunk, (1972)]. The estimate $\hat{m}_{IS}$ is obtained by applying the local linear estimate to the isotonized data obtained from the isotone least squares estimate [see Brunk (1955)]. For the estimator $\hat{m}_I$ we choose the second smoothing parameter $h_d = h_r^{1.01}$ and h_r proportional to $(\hat{\sigma}^2/n)^{1/3}$, where $\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1}(Y_{i+1} - Y_i)^2$ denotes the variance estimator proposed by Rice (1984). For the estimates $\hat{m}_{SI}$ and $\hat{m}_{IS}$ we use the same bandwidth h_r in the local linear smoothing procedure. We generate samples of size $n = 25$ from the homoscedastic nonparametric regression model (2.1) for the two once continuously differentiable regression functions

$$m_1(x) = \begin{cases} -8x^2 + 4x & x \in [0, \frac{1}{4}] \\ \frac{1}{2} & x \in (\frac{1}{4}, \frac{3}{4}] \\ 8x^2 - 12x + 5 & x \in (\frac{3}{4}, 1] \end{cases}$$

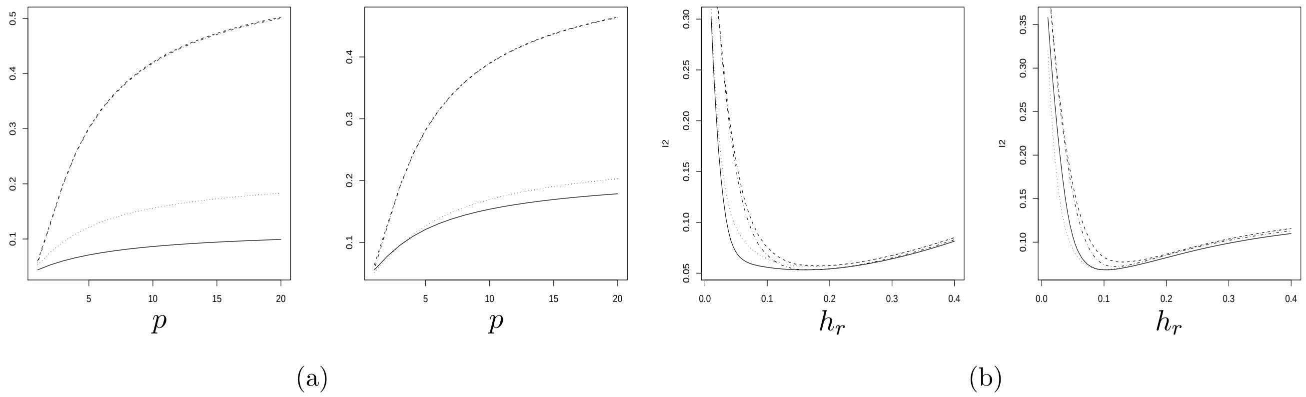


Figure 1: *(a) The L^p-errors as a function of $p \in [1,20]$ and (b) the L^2-error as a function of the bandwidth $h_r \in [0.01, 0.4]$ for the unconstrained (dashed line), the density regression (solid line) and the S+I (dotted line) and I+S (dot-dashed line) estimator. The regression functions are given by m_1 (left part of subfigures) and m_2 (right part of subfigures).*

and

$$m_2(x) = \begin{cases} \frac{12}{5}x^2 & x \in [0, \frac{1}{4}] \\ -\frac{2}{5}x^2 + \frac{7}{5}x - \frac{7}{40} & x \in (\frac{1}{4}, \frac{3}{4}] \\ \frac{12}{5}x^2 - \frac{14}{5}x + \frac{7}{5} & x \in (\frac{3}{4}, 1] \end{cases}$$

where the standard deviation is $\sigma = 0.05$. The explanatory variables are uniformly distributed on the interval $[0, 1]$, while the errors are standard normal distributed. The L^p-errors are estimated from 500 simulation runs and displayed for different values of p in Figure 1(a). The results clearly indicate that the density regression estimate reduces the L^p-estimation error of the unconstrained estimator substantially for all $p \in [1, \infty]$, which reflects the asymptotic statement in Theorem 3.1. We observe that in the two examples all monotone estimates reduce the L^p-error compared to the unconstrained estimate. The reduction is only minor for the I+S estimate, and the density regression estimate produces the smallest L^p-error. We have also investigated other cases, which are not displayed here for the sake of brevity, and found a very similar behaviour of the four estimates.

The L^p-error only gives information about the global performance of the estimates but does not reflect local properties of the different methods at points where the regression function is only once continuously differentiable. Therefore we show in Table 1 the simulated MSE of the different estimates at the points $x = 1/4$ and $x = 1/3$ (the behaviour at the point $x = 3/4$ is very similar and the point $x = 1/3$ is considered for a comparison of the once and twice differentiable case). For the regression function m_1 the density regression estimator behaves much better than the unconstrained, the S+I and the I+S estimator at both points. In this example the estimation error is not worse at points where the regression function is only once continuously differentiable. For the regression function m_2 the situation is slightly different. Here, the density regression estimate

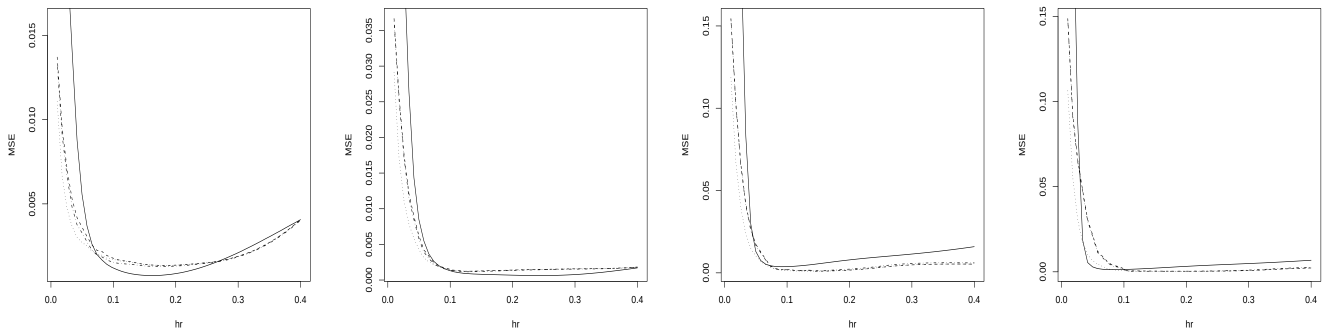
(a)

MSE	$x = 1/4$	$x = 1/3$
$\hat{m}_I$	0.000611	0.000608
$\hat{m}$	0.001242	0.001241
S+I	0.001240	0.001241
I+S	0.001169	0.001176

(b)

MSE	$x = 1/4$	$x = 1/3$
$\hat{m}_I$	0.010820	0.002556
$\hat{m}$	0.013683	0.015598
S+I	0.008047	0.004531
I+S	0.013256	0.014869

Table 1: *Simulated MSE of the different estimates at the points $x = 1/4$ and $x = 1/3$. The regression functions are m_1 (a) and m_2 (b).*



(a) (b)

Figure 2: *Simulated MSE as a function of the bandwidth h_r at $x = 1/4$ (left part of subfigures) and $x = 1/3$ (right part of subfigures)for the density regression (solid line), the unconstrained (dashed line), the S+I (dotted line) and the I+S (dot-dashed line). The subfigures correspond to the regression functions m_1 (a) and m_2 (b).*

has a better performance than the unconstrained and the I+S estimator at the point $x = 1/4$ and is worse than the S+I estimator. In this example the density regression estimate shows a better performance at the point $x = 1/3$, where the function is twice continuously differentiable.

In the second part of the simulation study we investigate the impact of the isotonization on the sensitivity of the estimates with respect to the choice of the smoothing parameter. Cuesta-Albertos, Domínguez-Menchero and Matrán (1995) proved that isotonization makes the S+I and I+S estimate less sensitive with respect to the selection of the smoothing parameter, and a similar behaviour can be observed for the density regression estimate. Figure 1(b) shows the L^2-estimation errors as functions of the bandwidth $h_r \in (0, 0.4)$. We observe that all estimates behave very similar for bandwidths near the optimal one for the unconstrained estimator. In the case of moderate undersmoothing, we see advantages of the density regression estimator $\hat{m}_I$ over the unconstrained as well as the S+I and, for m_2, also the I+S estimator. Combining all the information given in

Figure 1(b) we conclude that in the considered examples the density regression estimate $\hat{m}_I$ is more robust against moderate undersmoothing.
We finally show in Figure 2 the dependence of the MSE at the points $x = 1/4$ and $x = 1/3$ on the bandwidth. For the regression function m_1 the density regression estimate shows a better performance at both points than all other estimates over a large range of bandwidths. For the regression function m_2 the density regression estimate is better than the other estimates for small bandwidths but worse for larger bandwidths. The estimate is more sensitive with respect to the choice of the bandwidth at the point $x = 1/4$ than at $x = 1/3$. At the point $x = 1/3$ the minimal value of the MSE is comparable with that of the other estimates but attained for a smaller bandwidth while in $x = 1/4$ the minimal MSE is again attained for a smaller bandwidth than that for the other estimates but slightly larger than the best MSE obtained by the other estimates.

Acknowledgements. The authors are grateful to Isolde Gottschlich and Martina Stein who typed numerous versions of this paper with considerable technical expertise. The work of the authors was supported by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen. The authors would also like to thank two referees and the Associate Editor for their constructive comments on an earlier version of this paper.

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 and H.D. Brunk (1972). Statistical inference under order restrictions. Wiley, New York.

M. Birke, H. Dette (2007). Testing strict monotonicity in nonparametric regression. Math. Methods of Statistics 16, 1-13.

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.

V. Chernozhukov, I. Fernandez-Val, A. Galichon (2007). Improving Estimates of Monotone Functions by Rearrangement. arXiv:0704.3686. http://arxiv.org/abs/0704.3686v1

J.A. Cuesta-Albertos, J.S. Domínguez-Menchero, C. Matrán (1995). Consistency of the L^p-best monotone approximations. J. Statist. Plann. Inf. 47, 295-318.

M. Delecroix, C. Thomas-Agnan (2000). Spline and kernel regression under shape restrictions.

Smoothing and regression. Approaches, computation and application. Wiley Series in Probability and Statistics 109-133.

H. Dette, N. Neumeyer, K.F. Pilz (2006). A simple nonparametric estimator of a strictly monotone regression function. Bernoulli 12, 469-490.

H. Dette, K.F. Pilz (2006). A comparative study of monotone nonparametric kernel estimates. Journal of Statistical Computation and Simulation 76, 41-56.

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

I. Gijbels (2005). Monotone regression. In N. Balakrishnan, S. Kotz, C.B. Read and B. Vadakovic (eds), The Encyclopedia of Statistical Sciences, 2nd edition. Hoboken, NJ: Wiley.

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

G. H. Hardy, J. E. Littlewood, G. Pólya (1952). Inequalities. Cambridge University Press.

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

G. G. Lorentz (1953). An inequality for rearrangements. Amer. Math. Monthly 60, 176-179.

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-Agnan (1999). Smoothing splines and shape restrictions. Scand. J. Statist. 26, 239-252.

R.L. Matzkin (1994). Restrictions of econometric theory in nonparametric methods. Handbook of econometrics, 2523-2558.

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.

N. Neumeyer (2007). A note on uniform consistency of monotone function estimators. Statist. Probab. Lett. 77, 693–703.

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

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

F.T. Wright (1982). The asymptotic behavior of monotone regression estimates. Ann. Statist. 9, 443-448.

Appendix

Proof of Theorem 3.1. If $h_d \to 0$ it is easy to see that the function m_I^{-1} defined by (2.2) can be approximated as $m_I^{-1}(t) = \tilde{m}_I^{-1}(t) + o(1)$, where

$$\tilde{m}_I^{-1}(t) = \int_0^1 I\{m(x) \le t\}dx, \tag{A.1}$$

and the precise order of the error of this approximation depends on the smoothness of the regression function m. We define $\tilde{\hat{m}}_I$ as the function obtained by (A.1) for $m = \hat{m}$ and denote by $\tilde{m}_I$ and $\tilde{\hat{m}}_I$ the corresponding inverses. By a result of Chernozhukov, Fernndez-Val and Galichon (2007) it follows for all $1 \le p \le \infty$

$$\Big(\int_0^1 (\tilde{\hat{m}}_I(x) - m(x))^p\Big)^{1/p} \le \Big(\int_0^1 (\hat{m}(x) - m(x))^p\Big)^{1/p}, \tag{A.2}$$

where the inequality is strict, whenever $\hat{m}$ is strictly decreasing on a subset of positive Lebesgue measure. Theorem 3.1 can now be proven by similar arguments as in Neumeyer (2007), which give

$$|\ ||\hat{m}_I - m||_p - ||\tilde{\hat{m}}_I - m||_p\ | \le C(|m(1) - m(0)| + C_n)D_n^{1/p}$$

for some constant $C > 0$. The details are omitted for the sake of brevity. □

Proof of Theorem 3.2. The assertion follows by combining several lemmas which we prove in the following. We begin the asymptotic analysis with a Taylor expansion of the difference $\hat{m}_I^{-1}(t) - m^{-1}(t)$, that is

$$\hat{m}_I^{-1}(t) - m^{-1}(t) = m_I^{-1}(t) - m^{-1}(t) + \Delta_n^{(1)}(t) + \frac{1}{2}\Delta_n^{(2)}(t), \tag{A.3}$$

where the quantities $\hat{m}_I^{-1}$ and m_I^{-1} are defined in (2.4) and (2.2) respectively,

$$\Delta_n^{(1)}(t) = -\frac{1}{h_d}\int_0^1 K_d\left(\frac{m(v) - t}{h_d}\right)(\hat{m}(v) - m(v))\,dv \tag{A.4}$$

$$\Delta_n^{(2)}(t) = \frac{1}{h_d^3}\int_0^1 \int_{-\infty}^t K_d''\left(\frac{\xi(u,v) - u}{h_d}\right)du\,(\hat{m}(v) - m(v))^2\,dv, \tag{A.5}$$

and $|\xi(u,v)-m(v)| \leq |\hat{m}(v)-m(v)|$. We now investigate the three terms in this expansion separately.

Lemma A.1. *We have for any t with $m'(m^{-1}(t)) > 0$ and some $\lambda \in [0,1]$*

$$m_I^{-1}(t) - m^{-1}(t) = h_d \int_{-1}^{1} uK_d(u) \left(m^{-1}\right)'(t + h_d\lambda u)\, du =: b_{K_d}(t) \tag{A.6}$$

Proof. Using the same arguments as in Dette et al. (2006) yields for some $\lambda \in [0,1]$

$$\begin{aligned} D_n(t) &= m^{-1}(t-h_d) + h_d \int_{-1}^{1} \left(m^{-1}\right)'(t+h_d z) \int_z^1 K_d(v)\, dv dz - m^{-1}(t) \\ &= h_d \int_{-1}^{1} zK_d(z) \left(m^{-1}\right)'(t + h_d\lambda z)\, dz\,. \end{aligned}$$

$\square$

Lemma A.2. *We have for any t with $m'(m^{-1}(t)) > 0$*

$$\Delta_n^{(1)}(t) + h_r a_{K_d,K_r}(t) = \Delta_n^{(1.2)}(t) + o_p\Big(\frac{1}{\sqrt{nh}}\Big),$$

where $h = h_d$ or $h = h_r$,

$$\Delta_n^{(1.2)}(t) = -\frac{1}{nh_rh_d} \sum_{i=1}^{n} \int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right) K_r\left(\frac{v-X_i}{h_r}\right) \frac{\sigma(X_i)\,\varepsilon_i}{f(v)} dv, \tag{A.7}$$

$$a_{K_d,K_r}(t) = \int_{-1}^{1} K_d(v) \int_{-1}^{1} uK_r(u) \frac{m'\left(m^{-1}(t+h_dv) + h_r\mu u\right)}{m'\left(m^{-1}(t+h_dv)\right)} dudv. \tag{A.8}$$

Proof. We use the decomposition

$$\Delta_n^{(1)}(t) = \left(\Delta_n^{(1.1)}(t) + \Delta_n^{(1.2)}(t)\right)\left(1 + o_p(1)\right), \tag{A.9}$$

where $\Delta_n^{(1.2)}(t)$ is defined in (A.7) and

$$\Delta_n^{(1.1)}(t) = -\frac{1}{nh_rh_d} \sum_{i=1}^{n} \int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right) K_r\left(\frac{v-X_i}{h_r}\right) \frac{m(X_i)-m(v)}{f(v)} dv. \tag{A.10}$$

For the expectation of $\Delta_n^{(1.1)}(t)$ we obtain for some $\mu, \nu \in [0,1]$

$$\begin{aligned} \mathrm{E}\left[\Delta_n^{(1.1)}(t)\right] &= -\frac{1}{h_rh_d} \int_{m^{-1}(t-h_d)}^{m^{-1}(t+h_d)} \int_{v-h_r}^{v+h_r} K_d\left(\frac{m(v)-t}{h_d}\right) K_r\left(\frac{v-y}{h_r}\right) \frac{m(y)-m(v)}{f(v)} f(y)\, dydv \\ &= -h_r \int_{-1}^{1}\int_{-1}^{1} K_d(v)\, yK_r(y) \frac{m'\left(m^{-1}(t+h_dv) - h_r\mu y\right)}{m'\left(m^{-1}(t+h_dv)\right)} \\ &\qquad \times \left\{1 + h_r y \frac{f'\left(m^{-1}(t+h_dv) - h_r\nu y\right)}{f\left(m^{-1}(t+h_dv)\right)}\right\} dydv \\ &= h_r a_{K_d,K_r}(t) + o\left(\frac{1}{\sqrt{nh}}\right), \end{aligned}$$

where $h = h_d$ or $h = h_r$. On the other hand it was shown by Dette et al. (2006) that for $\lim_{h_d,h_r\to 0} h_r/h_d = c \in [0,\infty)$ $\mathrm{Var}(\Delta^{(1.1)}(t)) = o_p\Big(\frac{1}{\sqrt{nh_d}}\Big) = o_p\Big(\frac{1}{\sqrt{nh_r}}\Big)$ (note that the derivation of this statement in that paper only requires a regression function, which is once continuously differentiable). Finally, the expectation of $\Delta_n^{(1.2)}(t)$ is obviously 0, while the variance is obtained by a straightforward calculation as $\lim_{h_d,h_r\to 0} \mathrm{Var}(\sqrt{nh_d}\Delta_n^{(1.2)}(t)) = g^2(t)$. The assertion of the Lemma is now obvious from (A.9). □

Lemma A.3. *We have $\Delta_n^{(2)}(t) = \Delta_n^{(2.1)}(t)\,(1 + o_p(1))$, where the random variable $\Delta_n^{(2.1)}$ is defined by*

$$\Delta_n^{(2.1)}(t) = -\frac{1}{h_d^2}\int_0^1 K_d'\left(\frac{m(v)-t}{h_d}\right)(\hat{m}(v)-m(v))^2\,dv = O_p\Big(\frac{1}{\sqrt{nh_d}}\Big\{\frac{1}{\sqrt{nh_r^2h_d}} + \sqrt{\frac{nh_r^4}{h_d}}\Big\}\Big).$$

Proof. Recalling the definition of the term $\Delta_n^{(2)}(t)$ in (A.5) we obtain

$$\begin{aligned}
\Delta_n^{(2)}(t) &= \frac{1}{h_d^3}\int_0^1\int_{-\infty}^t K_d''\Big(\frac{m(v)-u}{h_d}\Big)(\hat{m}(v)-m(v))^2 \\
&\quad\times\Big[1+\Big(K_d''\Big(\frac{m(v)-u}{h_d}\Big)\Big)^{-1}\Big(K_d''\Big(\frac{\xi(u,v)-u}{h_d}\Big)-K_d''\Big(\frac{m(v)-u}{h_d}\Big)\Big)\Big]dudv \\
&= \frac{1}{h_d^2}\int_0^1\int_{\frac{m(v)-t}{h_d}}^\infty K_d''(u)(\hat{m}(v)-m(v))^2 \\
&\quad\times\Big[1+(K_d''(u))^{-1}\Big(K_d''\Big(u+\frac{\xi(m(v)-h_du,v)-m(v)}{h_d}\Big)-K_d''(u)\Big)\Big]dudv, \qquad \text{(A.11)}
\end{aligned}$$

where we used the substitution $u \to m(v) - h_d u$ in the second step. Using the estimate

$$\sup_u |\hat{m}(u) - m(u)| = O\Big(\Big(\frac{\log h_r^{-1}}{nh_r}\Big)^{1/2}\Big) \quad \text{a.s.} \qquad \text{(A.12)}$$

[see Mack and Silverman (1982)] we obtain

$$\frac{1}{h_d}\sup_{u,v}|\xi(m(v)-h_du,v)-m(v)| \;\le\; \frac{1}{h_d}\sup_v|\hat{m}(v)-m(v)| \;=\; O_P\Big(\frac{\log h_r^{-1}}{nh_rh_d^2}\Big)^{1/2} = o_P(1),$$

where we use assumption (W3) for the last estimate. By the continuity of K_d'' we have

$$\Big|K_d''\Big(u+\frac{\xi(m(v)-h_du,v)-m(v)}{h_d}\Big)-K_d''(u)\Big| = o_p(1) \qquad \text{(A.13)}$$

Therefore it follows from (A.11)

$$\begin{aligned}
|\Delta_n^{(2)}(t)| &\le \frac{1}{h_d^2}\Big|\int_0^1\int_{\frac{m(v)-t}{h_d}}^\infty K_d''(u)(\hat{m}(v)-m(v))^2dudv\Big|\;(1+o_p(1)) \\
&\le \frac{1}{h_d^2}\Big|\int_0^1 K_d'\Big(\frac{m(v)-t}{h_d}\Big)(\hat{m}(v)-m(v))^2dv\Big|\;(1+o_p(1)).
\end{aligned}$$

Observing $E[|\hat{m}(v) - m(v)|^2] = O(\frac{1}{nh_r} + h_r^2)$ we obtain $|\Delta_n^{(2)}(t)| = O_p(\frac{1}{nh_rh_d} + \frac{h_r^2}{h_d})$. The assertion of Lemma A.3 follows now from assumption (W3). $\square$

Observing the decomposition (A.3), Lemma A.1, A.2 and A.3 we obtain

$$\begin{aligned} & \sqrt{nh_d}\left(\hat{m}_I^{-1}(t) - m^{-1}(t) + h_r a_{K_d,K_r}(t) - h_d b_{K_d}(t)\right) \\ = & \sqrt{nh_d}\left(\Delta_n^{(1)}(t) + h_r a_{K_d,K_r}(t)\right)(1 + o_p(1)) + \sqrt{nh_d}\Delta_n^{(2)}(t). \\ = & \sqrt{nh_d}\Delta_n^{(1.2)}(t)\,(1 + o_p(1)) + o_p(1) \end{aligned}$$

By Ljapunoff's Theorem the remaining term is asymptotically normal with variance $g^2(t)$ defined in (3.3), which proves the assertion of Theorem 3.2. $\square$

Proof of Theorem 3.3. Let $s \in \{0, 1\}$, then it follows

$$\sup_{t\in J} |\left(\hat{m}_I^{-1}\right)^{(s)}(t) - \left(m^{-1}\right)^{(s)}(t)| \leq T_1^{(s)} + T_2^{(s)}, \tag{A.14}$$

where $T_1^{(s)} := \sup_{t\in J} |\left(\hat{m}_I^{-1}\right)^{(s)}(t) - \left(m_I^{-1}\right)^{(s)}(t)|$, $T_2^{(s)} := \sup_{t\in J} |\left(m_I^{-1}\right)^{(s)}(t) - (m^{-1})^{(s)}(t)|$. Observing the decomposition in (A.3) we therefore obtain $T_1^{(s)} \leq \Delta_n^{(1),(s)} + \frac{1}{2}\Delta_n^{(2),(s)}$, where we used the notation $\Delta_n^{(k),(s)} = \sup_{t\in J} |\,(\Delta_n^{(k)})^{(s)}(t)\,|$, $k = 1, 2;\ s = 0, 1$, the upper index (s) means differentiation with respect to the variable t (s times) and $\Delta_n^{(1)}(t)$ and $\Delta_n^{(2)}(t)$ are defined in (A.4) and (A.5), respectively. Assume that h_d is sufficiently small such that

$$\{t + h_d v \mid t \in J(\delta), |v| \leq 1\} \subset [m(0), m(1)], \tag{A.15}$$

then this term can be estimated as follows

$$\begin{aligned} \Delta_n^{(k),(s)} \leq & \frac{1}{h_d^{k+s-1}} \int_{-1}^{1} \left|K_d^{(k+s-1)}(v)\right| \sup_{t\in J} \left|(m^{-1})'(t + h_d v)\right| \\ & \times \left(\sup_{t\in J} \left|(\hat{m} - m) \circ m^{-1}(t + h_d v)\right|\right)^k dv \\ \leq & \frac{1}{h_d^{k+s-1}} \sup_z |\left(m^{-1}\right)'(z)| \sup_z |\hat{m}(z) - m(z)|^k \int_{-1}^{1} |K_d^{(k+s-1)}(v)|dv. \end{aligned} \tag{A.16}$$

Using (A.12) we obtain from (A.16) the estimate

$$\Delta_n^{(k),(s)} = O\Big(\frac{\log h_r^{-1}}{nh_r h_d^{2(k+s-1)/k}}\Big)^{k/2} \quad \text{a.s. } (k = 1, 2,\ s = 0, 1) \tag{A.17}$$

In the case $k = 2$ this gives $\Delta_n^{(2),(s)} = O\Big(\frac{\log h_r^{-1}}{nh_r h_d^{1+s}}\Big) = O\Big(\frac{\log h_r^{-1}}{nh_r h_d^{2s}}\Big)^{1/2}$ a.s. ($s = 0, 1$) by assumption (W3), while for the terms $\Delta_n^{(1),(s)}$ this estimate follows directly from (A.17). This yields for $s = 0, 1$

$$T_1^{(s)} = O\Big(\Big(\frac{\log h_r^{-1}}{nh_r h_d^{2s}}\Big)^{1/2}\Big) \quad \text{a.s. }, \tag{A.18}$$

and it remains to derive a corresponding estimate for the quantities $T_2^{(0)}$ and $T_2^{(1)}$. In the case $s = 0$ we have by Lemma 3.1 for the term $T_2^{(0)}$

$$
\begin{aligned}
\text{(A.19)} \qquad T_2^{(0)} &= h_d \sup_{t\in J} \Big| \int_{-1}^{1} uK_d(u) \left\{(m^{-1})'(t+h_d\lambda u) - (m^{-1})'(t)\right\} du\Big| \\
&\leq h_d \int_{-1}^{1} |u||K_d(u)| \sup_{t\in J} |(m^{-1})'(t+h_d\lambda u) - (m^{-1})'(t)| du \\
&= h_d \int_{-1}^{1} |u||K_d(u)| du \cdot o(1) = o(h_d),
\end{aligned}
$$

where we used the fact that $\int_{-1}^{1} uK_d(u)du = 0$ and the uniform continuity of the function $(m^{-1})'$ on the interval $[0,1]$. Finally, the remaining term $T_2^{(1)}$ is treated as follows

$$
\begin{aligned}
T_2^{(1)} &= \sup_{t\in J} \left| \frac{\partial}{\partial t} \left(\frac{1}{h_d} \int_0^1 \int_{-\infty}^{t} K_d\left(\frac{m(v)-u}{h_d}\right) dv \right) - \left(m^{-1}\right)'(t) \right| \\
&= \sup_{t\in J} \left| \frac{1}{h_d} \int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right) dv - \left(m^{-1}\right)'(t) \right| \\
&\leq \sup_{t\in J} \left| \left(m^{-1}\right)'(t+h_d\lambda v) - \left(m^{-1}\right)'(t) \right| = o(1)
\end{aligned}
$$

for some $\lambda \in [0,1]$, where we again used the uniform continuity of $(m^{-1})'$ on the interval $[0,1]$. The assertion of Theorem 3.5 now follows from (A.14), (A.18) and (A.19).

$\square$

Proof of Theorem 3.4. Without loss of generality it is assumed that $m'(x) > 0$ for all $x \in [0,1]$ (otherwise this assumption is satisfied in a neighbourhood of the point x and an appropriate subinterval has to be considered). Recall the definition of $J(\delta)$, and assume that n is sufficiently large, h_d and h_r are sufficiently small such that $\{\hat{m}_I(x) \mid x \in I(\eta)\} \subset J(\delta)$, where the set $I(\eta)$ has been defined in (3.4) (note that the function $\hat{m}_I^{-1}$ converges uniformly to m^{-1} on $J(\delta)$, by Theorem 3.3). By the mean value theorem we have for any $x \in I(\eta)$

$$
\text{(A.20)} \qquad \hat{m}_I^{-1}\left(\hat{m}_I(x)\right) - \hat{m}_I^{-1}\left(m(x)\right) = \left(\hat{m}_I(x) - m(x)\right)\left(\hat{m}_I^{-1}\right)'\left(\xi_{\hat{m}_I}(x)\right),
$$

where $|\xi_{\hat{m}_I}(x) - m(x)| \leq |\hat{m}_I(x) - m(x)|$. Note that $\xi_{\hat{m}_I}(x) \in J(\delta)$, because it is a convex combination of $m(x)$ and $\hat{m}_I(x)$. $(m^{-1})'$ is bounded from below by some positive constant in a neighbourhood of the point $m(x)$ and by Theorem 3.3 the same holds true for the estimate $(\hat{m}_I^{-1})'$ if n is sufficiently large. Form (A.20) and the identity $\hat{m}_I^{-1}\left(\hat{m}_I(x)\right) = m^{-1}\left(m(x)\right)$, we obtain

$$
\text{(A.21)} \qquad \hat{m}_I(x) - m(x) = -\frac{\hat{m}_I^{-1}\left(m(x)\right) - m^{-1}\left(m(x)\right)}{\left(\hat{m}_I^{-1}\right)'\left(\xi_{\hat{m}_I}(x)\right)}.
$$

We will finally show that the denominator in this expression converges in probability to $(m^{-1})'(m(x))$ $= 1/m'(x)$. The assertion of Theorem 3.4 is then obvious from Theorem 3.2. For this final step we use the estimate

$$\text{(A.22)}\qquad \begin{aligned} \left|\left(\hat{m}_I^{-1}\right)'(\xi_{\hat{m}_I}(x)) - \left(m^{-1}\right)'(m(x))\right| \;\le\; & \left|\left(\hat{m}_I^{-1}\right)'(\xi_{\hat{m}_I}(x)) - \left(m^{-1}\right)'(\xi_{\hat{m}_I}(x))\right| \\ & + \left|\left(m^{-1}\right)'(\xi_{\hat{m}_I}(x)) - \left(m^{-1}\right)'(m(x))\right|. \end{aligned}$$

It follows from the proof of Theorem 3.3 that the random variables

$$T^{(0)}(t) = \left|\hat{m}_I^{-1}(t) - m^{-1}(t)\right|; \quad T^{(1)}(t) = \left|\left(\hat{m}_I^{-1}\right)'(t) - \left(m^{-1}\right)'(t)\right|$$

converge a.s. to 0 uniformly on the set $J(\delta)$. This implies the uniform a.s. convergence of $\hat{m}_I(x)$ to $m(x)$ on $I(\eta)$ and as a consequence the random variable $\xi_{\hat{m}_I(x)}$ converges to $m(x)$ a.s. The continuity of $(m^{-1})'$ now implies the a.s. convergence of $(m^{-1})'(\xi_{\hat{m}_I(x)})$ to $(m^{-1})'(m(x))$, which shows that the second term in (A.22) converges to 0. By the previous discussion we have $\xi_{\hat{m}_I}(x) \in J(\delta)$ and the uniform convergence of $T^{(1)}(t)$ on $J(\delta)$ yields for the first term in (A.22) $T^{(1)}(\xi_{\hat{m}_I(x)}) = o(1)$ *a.s.* In other words the left hand side of (A.22) converges uniformly to 0 which completes the proof of Theorem 3.4. $\square$