A note on estimating a 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 22, 2005

Abstract

In a recent paper Dette, Neumeyer and Pilz (2005) proposed a new nonparametric estimate of a 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 monotone estimates. In this note we provide a more refined asymptotic analysis of the monotone regression estimate. It is shown that in the case of a non-decreasing regression function the new method produces an estimate with nearly the same L^p-norm as the given function 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 arbitrarily small. Therefore the rate of convergence of the new estimate is arbitrarily close to the rate of the estimate obtained from monotone least squares estimation, but the asymptotic distribution of the new estimate is substantially simpler.

AMS Subject Classification: 62G05, 62G20

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

1 Introduction

One of the most important problems in applied statistics is the estimation of relationships among observable variables. In many cases 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 (2003).

In a recent paper Dette, Neumeyer and Pilz (2005) introduced an alternative monotone estimate of the regression function, which is based on a non-decreasing rearrangement of the Nadaray-Watson estimate. This method is called density-regression estimate, because it is based on the combination of a density and regression estimator. In a first step an estimate of the inverse of the monotone regression function is constructed using a density estimator, 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, Neumeyer and Pilz (2005). In particular we show that the isotone estimate is an approximation of the unconstrained estimate in the sense that both estimates have the same L^p-norm for all $p \geq 1$ (the result is in fact slightly stronger - see Theorem 2.1). On the other hand we investigate the properties of this estimate in the case where the regression function is only once continuously differentiable. Moreover, 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. 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 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, Neumeyer and Pilz (2005) 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 arbitrarily small. In other words, this estimate has nearly the same asymptotic mean squared error and variance as the least squares isotone regression estimate but is still asymptotically normal distributed. 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, Neumeyer and Pilz (2005) and it is proved that this estimate has the same L^p-norm as the unconstrained preliminary estimate. In other words, if the "true" regression estimate is not isotone, the density regression estimate of Dette, Neumeyer and Pilz (2005) converges to an isotone function with the same L^p-norm as the "true" function (for all $p \geq 1$). Section 3 contains our asymptotic main results and we establish asymptotic normality of the density-regression estimate in the case of a once continuously differentiable regression function. We also establish uniform almost sure consistency of the estimate in this case, which extends the results of Dette, Neumeyer and Pilz (2005) in a further direction.

2 Monotone smoothing by inversion

Consider the nonparametric regression model

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

where $\{(X_i, Y_i)\}_{i=1}^n$ is a bivariate sample of i.i.d. observations such that 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, finite variance and existing fourth moment. 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, Neumeyer and Pilz (2005) we consider a transformation of the regression function defined by

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

where K_d is a given density and h_d denotes a bandwidth converging to 0 with increasing sample size. If $h_d \to 0$ it is easy to see that m_I^{-1} can be approximated as $m_I^{-1}(t) = \tilde{m}_I^{-1}(t) + o(1)$, where

$$\tilde{m}_I^{-1}(t) = \int_0^1 I\{m(x) \leq t\} dx, \tag{2.3}$$

and the precise order of the error of this approximation depends on the smoothness of the regression function m. Note that m_I^{-1} and $\tilde{m}_I^{-1}$ are isotone even if m is not isotone. Therefore we can calculate the inverse of these functions, which will be denoted by m_I and $\tilde{m}_I$ throughout this paper. The function $\tilde{m}_I$ is called a nondecreasing rearrangement of the function m [see e.g. Ryff (1965, 1970) or Bennett and Sharpley (1988) among others]. Our first result shows that this function is an approximation to the function m in the sense that it has the same L^p-norm on the interval $[0, 1]$.

Theorem 2.1. *Let*

$$\tilde{m}_I(x) = \inf\{u \mid \tilde{m}_I^{-1}(u) \geq x\}$$

denote the inverse of the function $\tilde{m}_I^{-1}$ defined by (2.3), then we have for all $0 < p < \infty$

$$\Big(\int_0^1 |m(x)|^p dx \Big)^{1/p} = \Big(\int_0^1 |\tilde{m}_I(x)|^p dx \Big)^{1/p}.$$

Proof. Consider a step function of the form

$$m(x) = \sum_{j=1}^{n} a_j I_{E_j}(x), \tag{2.4}$$

where $-\infty < a_1 < \ldots < a_n < \infty$ and $E_1, E_2, \ldots, E_n \subset \mathbb{R}$ are pairwise disjoint sets with finite Lebesgue measure such that $[0,1] = \cup_{j=1}^n E_j$. Observing the definition (2.3) it is easy to see that

$$\tilde{m}_I^{-1}(t) = \sum_{j=1}^{n} m_j I_{[a_j, a_{j+1})}(t), \tag{2.5}$$

where $m_j = \sum_{i=1}^{j} \lambda(E_i)$ $(j \geq 1)$, $m_0 = 0$, and λ denotes the Lebesgue measure. This implies for the inverse of the function $\tilde{m}_I^{-1}$

$$\tilde{m}_I(x) = \inf\{u \mid \tilde{m}_I^{-1}(u) \geq x\} = \sum_{j=1}^{n} a_j I_{(m_{j-1}, m_j]}(x). \tag{2.6}$$

Consequently, we obtain

$$\int_0^1 |m(x)|^p dx = \sum_{j=1}^{n} |a_j|^p \lambda(E_j) = \sum_{j=1}^{n} |a_j|^p (m_j - m_{j-1}) = \int_0^1 |m_I(x)|^p dx$$

which proves the assertion of Theorem 2.1 for step functions. The general statement now follows from the fact that for any decreasing sequence of functions $(m^{(n)})_{n\in\mathbb{N}}$ with limit m, the corresponding sequence $(m_I^{(n)})_{n\in\mathbb{N}}$ is a decreasing sequence with limit $\tilde{m}_I$, i.e.

$$m^{(n)} \searrow m \Rightarrow \tilde{m}_I^{(n)} \searrow \tilde{m}_I. \tag{2.7}$$

For a proof of the property (2.7) assume $m^{(n)} \searrow m$, define the sets

$$E_n = \{x \in [0,1] \mid m^{(n)}(x) \leq t\}, E = \{x \in [0,1] \mid m(x) \leq t\},$$

then $(E_n)_{n\in\mathbb{N}}$ defines an increasing sequence of events with limit

$$E = \lim_{n\to\infty} E_n.$$

Consequently, we obtain

$$(\tilde{m}_I^{(n)})^{-1}(t) = \int_0^1 I\{m^{(n)}(x) \leq t\} dx = \lambda(E_n) \nearrow \lambda(E) = \tilde{m}_I^{-1}(t).$$

But this implies

$$\tilde{m}_I^{(n)}(x) = \inf\{u \mid (\tilde{m}_I^{(n)})^{-1}(u) \geq x\} \searrow \inf\{u \mid \tilde{m}_I^{-1}(u) \geq x\} = \tilde{m}_I(x),$$

which proves (2.7) and completes the proof of Theorem 2.1.

□

Example 2.2. In order to fix ideas we consider the regression function

$$m(x) = 3(2x-1)^2 \quad x \in [0,1], \tag{2.8}$$

which is obviously neither decreasing nor increasing. In this case we have

$$\tilde{m}_I^{-1}(t) = \sqrt{\frac{t}{3}}; \quad t \in [0,3],$$

which gives $\tilde{m}_I(x) = 3x^2$ as non-decreasing rearrangement of the function m. Note that the functions $\tilde{m}_I$ and m have for all $p > 0$ the same L^p-norm on the interval $[0,1]$.

Because the regression function in (2.1) is unknown, we replace it by a 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

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_r\left(\frac{X_i - x}{h_r}\right) Y_i}{\sum_{i=1}^n K_r\left(\frac{X_i - x}{h_r}\right)}, \tag{2.9}$$

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

$$\hat{m}_I^{-1}(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{\hat{m}(v) - u}{h_d}\right) du dv, \tag{2.10}$$

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 Dette, Neumeyer and Pilz (2005) replaced the integral with respect to dv in (2.10) by a discrete approximation of the Riemann integral, but in this paper we will work with the representation (2.10) 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, Neumeyer and Pilz (2005).

It is also worthwhile to mention 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. Similary, the index r in (2.9) 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.

We assume that the kernels K_d and K_r are symmetric with compact support, say $[-1,1]$, existing second moment and that the corresponding bandwidths h_d, h_r converge to 0 with increasing sample size n. We also assume that K_d is twice continuously differentiable on its support and that the kernel K_r has been appropriately modified in order to address for boundary effects [see Müller (1985)]. If m and f are twice continuously differentiable, Dette, Neumeyer and Pilz (2005) proved the asymptotic normality of the density-regression estimate and we mention their result here for the sake of completeness.

Theorem 2.3. [Dette, Neumeyer, Pilz (2005)] *Assume that $m, f \in C^2([0,1])$, $\sigma^2 \in C([0,1])$, and that $\log h_r^{-1}/(nh_rh_d^3) = o(1)$, $h_d, h_r \to 0$ as $n \to \infty$.*

(a) If $\lim_{n\to\infty} \frac{h_r}{h_d} = c \in [0,\infty)$ exists, then we have for every $t \in (0,1)$ with $m'(t) > 0$

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

where $\kappa_2(K_r) = \frac{1}{2}\int_{-1}^{1} u^2K_r(u)du$ and the asymptotic variance is given by

$$s^2(t) = \frac{\sigma^2(t)m'(t)}{f(t)} \int\int\int K_d(w + cm'(t)(v-u))K_d(w)K_r(u)K_r(v)dwdudv. \tag{2.11}$$

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

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

where the asymptotic variance is given by

$$\tilde{s}^2(t) = \frac{\sigma^2(t)}{f(t)} \int K_r^2(u)du. \tag{2.12}$$

Note that in the second case, i.e. $h_d = o(h_r)$ the isotone estimate is first order asymptotic equivalent to the Nadaraya-Watson estimate, but this is not the case if the bandwidths h_d and h_r are of the same order. In the following section we will investigate the asymptotic behaviour of the density-regression estimate $\hat{m}_I$ in the case where $m, f \in C^1([0,1])$.

3 A refined asymptotic analysis

Note that Dette, Neumeyer and Pilz (2005) assumed (among other technical assumptions) that $m \in C^2([0,1])$ and showed asymptotic normality of the random variable

$$\left(\hat{m}_I^{-1}(t) - \mathrm{E}\left[\hat{m}_I^{-1}(t)\right]\right) . \tag{3.1}$$

This result is then used to establish the asymptotic normality of $(\hat{m}_I(t) - \mathrm{E}[\hat{m}_I(t)])$ [for the precise statement see Theorem 2.3]. 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

$$\lim_{h_d \to 0, h_r \to 0} \frac{h_d}{h_r} = \infty \tag{3.2}$$

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

$$\sqrt{nh_d}\left(\hat{m}_I^{-1}(t) - \mathrm{E}\left[\hat{m}_I^{-1}(t)\right]\right). \tag{3.3}$$

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, Neumeyer and Pilz (2005). In particular we require 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

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

We begin the asymptotic analysis with a Taylor expansion of the difference $\hat{m}_I^{-1}(t) - m^{-1}(t)$, that is

$$\hat{m}_I^{-1}(t) - m^{-1}(t) = m_I^{-1}(t) - m^{-1}(t) + \Delta_n^{(1)}(t) + \frac{1}{2}\Delta_n^{(2)}(t), \tag{3.4}$$

where the quantities $\hat{m}_I^{-1}$ and m_I^{-1} are defined in (2.10) and (2.2) respectively,

$$\Delta_n^{(1)}(t) = \frac{1}{h_d^2}\int_0^1\int_{-\infty}^t K_d'\left(\frac{m(v)-u}{h_d}\right)du\,(\hat{m}(v)-m(v))\,dv \tag{3.5}$$
$$= -\frac{1}{h_d}\int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right)(\hat{m}(v)-m(v))\,dv$$

$$\Delta_n^{(2)}(t) = \frac{1}{h_d^3}\int_0^1\int_{-\infty}^t K_d''\left(\frac{\xi(u,v)-u}{h_d}\right)du\,(\hat{m}(v)-m(v))^2\,dv, \tag{3.6}$$

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

Lemma 3.1. *If the assumptions (V1)-(V4), (W1)-(W3) are satsified we have for any t with $m'(m^{-1}(t)) > 0$ and some $\lambda \in [0,1]$*

$$m_I^{-1}(t) - m^{-1}(t) = h_d\int_{-1}^1 uK_d(u)\left(m^{-1}\right)'(t+h_d\lambda u)\,du =: b_{K_d}(t) \tag{3.7}$$

Proof. Using the same arguments as in the proof of Lemma 2.1 in Dette, Neumeyer and Pilz (2005) yields

$$\begin{aligned} D_n(t) &= m_I^{-1}(t) - m^{-1}(t) \\ &= m^{-1}(t-h_d) + h_d\int_{-1}^1\left(m^{-1}\right)'(t+h_dz)\int_z^1 K_d(v)\,dvdz - m^{-1}(t). \end{aligned}$$

Therefore we obtain by integration by parts

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

for some $\lambda \in [0,1]$. □

We now investigate the second term $\Delta_n^{(1)}(t)$ in the decomposition (3.4).

Lemma 3.2. *If the assumptions (V1) - (V4), (W1) - (W3) are satisfied then 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 $h = h_r$,

$$\Delta_n^{(1.2)}(t) = -\frac{1}{nh_rh_d}\sum_{i=1}^{n}\int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right)K_r\left(\frac{v-X_i}{h_r}\right)\frac{\sigma(X_i)\varepsilon_i}{f(v)}dv, \tag{3.8}$$

and the quantity $a_{K_d,K_r}(t)$ is given by

$$a_{K_d,K_r}(t) = \int_{-1}^{1} K_d(v)\int_{-1}^{1} uK_r(u)\frac{m'(m^{-1}(t+h_dv)+h_r\mu u)}{m'(m^{-1}(t+h_dv))}dudv \tag{3.9}$$

Proof. We use the decomposition

$$\Delta_n^{(1)}(t) = \left(\Delta_n^{(1.1)}(t) + \Delta_n^{(1.2)}(t)\right)(1+o_p(1)), \tag{3.10}$$

where $\Delta_n^{(1.2)}(t)$ is defined in (3.8) and

$$\Delta_n^{(1.1)}(t) = -\frac{1}{nh_rh_d}\sum_{i=1}^{n}\int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right)K_r\left(\frac{v-X_i}{h_r}\right)\frac{m(X_i)-m(v)}{f(v)}dv. \tag{3.11}$$

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

$$\begin{aligned}
\mathrm{E}\left[\Delta_n^{(1.1)}(t)\right] &= -\frac{1}{nh_rh_d}\sum_{i=1}^{n}\mathrm{E}\left[\int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right)K_r\left(\frac{v-X_i}{h_r}\right)\frac{m(X_i)-m(v)}{f(v)}dv\right] \\
&= -\frac{1}{h_rh_d}\int_0^1\int_0^1 K_d\left(\frac{m(v)-t}{h_d}\right)K_r\left(\frac{v-y}{h_r}\right)\frac{m(y)-m(v)}{f(v)}dvf(y)dy \\
&= -\frac{1}{h_rh_d}\int_{m^{-1}(t-h_d)}^{m^{-1}(t+h_d)}\int_{v-h_r}^{v+h_r} K_d\left(\frac{m(v)-t}{h_d}\right)K_r\left(\frac{v-y}{h_r}\right)\frac{m(y)-m(v)}{f(v)}f(y)dydv \\
&= -h_r\int_{-1}^{1}\int_{-1}^{1} K_d(v)\,yK_r(y)\frac{m'(m^{-1}(t+h_dv)-h_r\mu y)}{m'(m^{-1}(t+h_dv))} \\
&\qquad \times\left\{1+h_ry\frac{f'(m^{-1}(t+h_dv)-h_r\nu y)}{f(m^{-1}(t+h_dv))}\right\}dydv \\
&= -h_r\int_{-1}^{1} K_d(v)\int_{-1}^{1} yK_r(y)\frac{m'(m^{-1}(t+h_dv)-h_r\mu y)}{m'(m^{-1}(t+h_dv))}dydv + o\left(\frac{1}{\sqrt{nh}}\right) \\
&= h_ra_{K_d,K_r}(t) + o\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned}$$

where $h = h_d$ or $h = h_r$. On the other hand it was shown by Dette, Neumeyer and Pilz (2005) that for $\lim_{h_d,h_r\to 0} h_r/h_d = c \in [0,\infty)$

$$\mathrm{Var}(\Delta^{(1.1)}(t)) = o_p\Big(\frac{1}{\sqrt{nh_d}}\Big) = o_p\Big(\frac{1}{\sqrt{nh_r}}\Big) \tag{3.12}$$

(note that the derivation of this statement in this 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

$$\lim_{h_d,h_r\to 0} \mathrm{Var}(\sqrt{nh_d}\Delta_n^{(1.2)}(t)) \;=\; \frac{\sigma^2(m^{-1}(t))}{f(m^{-1}(t))m'(m^{-1}(t))}\int_{-1}^{1} K_d^2(v)dv. \tag{3.13}$$

The assertion of the Lemma is now obvious from (3.10).

□

Our final auxiliary result deals with the term $\Delta_n^{(2)}(t)$ in the decomposition (3.4).

Lemma 3.3. *If the assumptions (V1)-(V4), (W1)-(W3) are satsified, we have*

$$\Delta_n^{(2)}(t) = \Delta_n^{(2.1)}(t)\left(1 + o_p(1)\right),$$

where the random variable $\Delta_n^{(2.1)}$ *is defined by*

$$\Delta_n^{(2.1)}(t) = -\frac{1}{h_d^2}\int_0^1 K_d'\left(\frac{m(v)-t}{h_d}\right)\left(\hat m(v) - m(v)\right)^2 dv$$

and satsifies

$$\sqrt{nh_d}\Delta_n^{(2.1)}(t) = O_p\Big(\frac{1}{\sqrt{nh_r^2h_d}} + \sqrt{\frac{nh_r^4}{h_d}}\Big).$$

Proof. Recalling the definition of the term $\Delta_n^{(2)}(t)$ in (3.6) we obtain

$$\begin{aligned}
\Delta_n^{(2)}(t) &= \frac{1}{h_d^3}\int_0^1\int_{-\infty}^{t} K_d''\Big(\frac{m(v)-u}{h_d}\Big)(\hat m(v)-m(v))^2 \\
&\quad\times\Big[1+\Big(K_d''\Big(\frac{m(v)-u}{h_d}\Big)\Big)^{-1}\Big(K_d''\Big(\frac{\xi(u,v)-u}{h_d}\Big)-K_d''\Big(\frac{m(v)-u}{h_d}\Big)\Big)\Big]dudv \\
&= \frac{1}{h_d^2}\int_0^1\int_{\frac{m(v)-t}{h_d}}^{\infty} K_d''(u)(\hat m(v)-m(v))^2 \\
&\quad\times\Big[1+(K_d''(u))^{-1}\Big(K_d''\Big(u+\frac{\xi(m(v)-h_du,v)-m(v)}{h_d}\Big)-K_d''(u)\Big)\Big]dudv,
\end{aligned} \tag{3.14}$$

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

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

[see Mack and Silverman (1982)] we obtain

$$\frac{1}{h_d}\,\mathrm{E}\Big[\sup_{u,v}|\xi(m(v)-h_du,v)-m(v)|\Big] \;\le\; \frac{1}{h_d}\,\mathrm{E}\Big[\sup_v |\hat m(v)-m(v)|\Big] \;=\; O\Big(\frac{\log h_r^{-1}}{nh_rh_d^2}\Big)^{1/2} = o(1),$$

where we use assumption (W3) for the last estimate. By Markov's inequality it follows that

$$\sup_{u,v} |\xi\,(m\,(v) - h_d u, v)| = o_p\,(h_d)$$

and by the continuity of K_d'' we have

$$\left|K_d''\left(u + \frac{\xi\,(m\,(v) - h_d u, v) - m\,(v)}{h_d}\right) - K_d''\,(u)\right| = o_p(1) \tag{3.15}$$

Therefore it follows from (3.14)

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

For the expectation of the second factor we have (using the assumption that $m \in C^1([0,1])$)

$$E\Big[|\hat{m}(v) - m(v)|^2\Big] = O\Big(\frac{1}{nh_r} + h_r^2\Big),$$

which yields

$$|\Delta_n^{(2)}(t)| = O_p\Big(\frac{1}{nh_r h_d} + \frac{h_r^2}{h_d}\Big).$$

Therefore the assertion of Lemma 3.3 follows from assumption (W3). □

We are now in a position to prove the asymptotic normality of the estimate $\hat{m}_I^{-1}$ for the inverse of the regression function.

Theorem 3.4 *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 the constants a_{K_d,K_r} and b_{K_d} are given by (3.9) and (3.7), respectively, and

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

Proof. Observing the decomposition (3.4), Lemma 3.1, 3.2 and 3.3 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 and (3.13) the remaining term is asymptotically normal with variance

$$\frac{\sigma^2(m^{-1}(t))}{f(m^{-1}(t))m'(m^{-1}(t))}\int_{-1}^{1}K_d^2(v)dv,$$

which proves the assertion of Theorem 3.4.

□

Note that the final monotone estimate of the regression function is obtained by an inversion of the function $\hat{m}_I^{-1}$. Dette, Neumeyer and Pilz (2005) 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 $m \in C^1([0,1])$ only 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 3.5. *Assume that the assumptions* $(V1)-(V4),(W1)-(W3)$ *are satisfied and that* $m'(m^{-1}(t))>0$ *for all* $t\in(m(0),m(1))$. *Let* $\delta>0$ *be an arbitrary small positive number, define*

$$J := J(\delta) = [m(0)+\delta, m(1)-\delta],$$

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

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

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

$$\sup_{t\in J}|\left(\hat{m}_I^{-1}\right)^{(s)}(t)-\left(m^{-1}\right)^{(s)}(t)| \le T_1^{(s)}+T_2^{(s)}, \tag{3.16}$$

where the quantities $T_1^{(s)}$ and $T_2^{(s)}$ are defined by

$$\begin{aligned} T_1^{(s)} &:= \sup_{t\in J}|\left(\hat{m}_I^{-1}\right)^{(s)}(t)-\left(m_I^{-1}\right)^{(s)}(t)|, \\ T_2^{(s)} &:= \sup_{t\in J}|\left(m_I^{-1}\right)^{(s)}(t)-\left(m^{-1}\right)^{(s)}(t)|. \end{aligned}$$

Observing the decomposition in (3.4) we therefore obtain

$$T_1^{(s)} \le \Delta_n^{(1),(s)}+\frac{1}{2}\Delta_n^{(2),(s)},$$

where we used the notation

$$\Delta_n^{(k),(s)} = \sup_{t\in J}\left|\left(\Delta_n^{(k)}\right)^{(s)}(t)\right| \quad , \quad k=1,2;\; s=0,1,$$

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 (3.5) and (3.6), respectively. Assume that h_d is sufficiently small such that

$$\{t + h_d v \mid t \in J(\delta), |v| \le 1\} \subset [m(0), m(1)], \tag{3.17}$$

then this term can be estimated as follows

$$\begin{aligned} \Delta_n^{(k),(s)} \quad &\le \quad \frac{1}{h_d^{k+s-1}} \int_{-1}^{1} \left| K_d^{(k+s-1)}(v) \right| \sup_{t \in J} \left| \left(m^{-1}\right)'(t + h_d v) \right| \\ &\qquad \times \left(\sup_{t \in J} \left| (\hat{m} - m) \circ m^{-1}(t + h_d v) \right| \right)^k dv \\ &\le \quad \frac{1}{h_d^{k+s-1}} \sup_z \left| \left(m^{-1}\right)'(z) \right| \sup_z |\hat{m}(z) - m(z)|^k \int_{-1}^{1} |K_d^{(k+s-1)}(v)| dv. \end{aligned} \tag{3.18}$$

Using a similar argument as in Mack and Silverman (1982) (in the case where $m, f \in C^1([0,1])$) we have under the assumption $h_r = o\left(\log h_r^{-1}/nh_r\right)^{1/2}$

$$\sup_z |\hat{m}(z) - m(z)| = O\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2}, \quad \text{a.s.}$$

which yields in (3.18) the estimate

$$\Delta_n^{(k),(s)} = O\Big(\frac{\log h_r^{-1}}{nh_r h_d^{2(k+s-1)/k}}\Big)^{k/2} \qquad \text{a.s. } (k = 1, 2, \ s = 0, 1) \tag{3.19}$$

In the case $k = 2$ we obtain

$$\Delta_n^{(2),(s)} = O\left(\frac{\log h_r^{-1}}{nh_r h_d^{1+s}}\right) = O\left(\frac{\log h_r^{-1}}{nh_r h_d^{2s}}\right)^{1/2} \qquad \text{a.s.} \quad (s = 0, 1)$$

by assumption (W3), while for the terms $\Delta_n^{(1),(s)}$ this estimate follows directly from (3.19). This yields for $s = 0, 1$

$$T_1^{(s)} = O\Big(\Big(\frac{\log h_r^{-1}}{nh_r h_d^{2s}}\Big)^{1/2}\Big) \quad \text{a.s. ,} \tag{3.20}$$

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 $T_2^{(0)}$

$$\begin{aligned} T_2^{(0)} \quad &= \quad h_d \sup_{t \in J} \Big| \int_{-1}^{1} u K_d(u) (m^{-1})'(t + h_d \lambda u) du \Big| \\ &= \quad h_d \sup_{t \in J} \Big| \int_{-1}^{1} u K_d(u) \left\{ (m^{-1})'(t + h_d \lambda u) - (m^{-1})'(t) \right\} du \Big| \\ &\le \quad h_d \int_{-1}^{1} |u| |K_d(u)| \sup_{t \in J} |(m^{-1})'(t + h_d \lambda u) - (m^{-1})'(t)| du \\ &= \quad h_d \int_{-1}^{1} |u| |K_d(u)| du \cdot o(1) = o(h_d), \end{aligned} \tag{3.21}$$

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

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

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 (3.16), (3.20) and (3.21).

□

Theorem 3.5 will be the main tool for deriving the asymptotic normality of the estimate $\hat{m}_I$. For the statement of this result we define for any $\delta > 0$, $\eta > 0$ the set

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

Theorem 3.6. *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 (3.7) and (3.9), respectively and

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

Proof. 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

(3.23) $$\{\hat{m}_I(x) \mid x \in I(\eta)\} \subset J(\delta) ,$$

where the set $I(\eta)$ has been defined in (3.22) (note that the function $\hat{m}_I^{-1}$ converges uniformly to m^{-1} on $J(\delta)$, by Theorem 3.5). By the mean value theorem we have for any $x \in I(\eta)$

(3.24) $$\hat{m}_I^{-1}(\hat{m}_I(x)) - \hat{m}_I^{-1}(m(x)) = (\hat{m}_I(x) - m(x))\left(\hat{m}_I^{-1}\right)'(\xi_{\hat{m}_I}(x)) ,$$

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 $\hat{m}_I(x)$. By assumption $(m^{-1})'$ is bounded from below by some positive constant in a neighbourhood of the point $m(x)$ and by Theorem 3.5 the same holds true for the estimate $(\hat{m}_I^{-1})'$ if n is sufficiently large. Observing the identity

$$\hat{m}_I^{-1}(\hat{m}_I(x)) = m^{-1}(m(x)),$$

we obtain from (3.24)

$$\hat{m}_I(x) - m(x) = -\frac{\hat{m}_I^{-1}(m(x)) - m^{-1}(m(x))}{(\hat{m}_I^{-1})'(\xi_{\hat{m}_I}(x))}. \tag{3.25}$$

We will finally show that the nominator in this expression converges in probability to $(m^{-1})'(m(x)) = 1/m'(x)$. The assertion of Theorem 3.6 is then obvious from Theorem 3.4. For this final step we use the estimate

$$\begin{aligned} \left|(\hat{m}_I^{-1})'(\xi_{\hat{m}_I}(x)) - (m^{-1})'(m(x))\right| \leq{} & \left|(\hat{m}_I^{-1})'(\xi_{\hat{m}_I}(x)) - (m^{-1})'(\xi_{\hat{m}_I}(x))\right| \\ & + \left|(m^{-1})'(\xi_{\hat{m}_I}(x)) - (m^{-1})'(m(x))\right|. \end{aligned} \tag{3.26}$$

It follows from the proof of Theorem 3.5 that the random variables

$$T^{(0)}(t) = \left|\hat{m}_I^{-1}(t) - m^{-1}(t)\right|$$

and

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

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 (3.26) 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 (3.26)

$$T^{(1)}(\xi_{\hat{m}_I(x)}) = o(1) \ a.s..$$

In other words the left hand side of (3.26) converges uniformly to 0 which completes the proof of Theorem 3.6. □

4 Further discussion

Note that the result of Theorem 3.6 requires the condition $h_r = o(h_d)$, which is used at several steps in the proofs of Section 3. We were not able to derive an asymptotic law in the case $\lim_{h_d,h_r\to 0} h_d/h_r = c \in [0,\infty)$ 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.

In the remaining part of this paper we discuss the case, where the bandwidth h_r of the regression estimate is chosen as

$$h_r = cn^{-1/3}. \tag{4.1}$$

This case is of particular interest, because the choice corresponds to the optimal rate (with respect to mean squared error) in nonparametric estimation of a once continuously differentiable regression function. Moreover, it is known that the appropriately normalized monotone least squares estimate 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 Robertson, Wright and Dykstra (1989), Theorem 9.2.4]. In this case the conditions in (W3) yield for the bandwidth in the density step

$$h_d = n^{-1/3}\alpha_n \tag{4.2}$$

where the sequence α_n converges to infinity such that

$$\alpha_n = O(n^{\varepsilon}) \quad \text{for some } 0 < \varepsilon < \frac{1}{12} \tag{4.3}$$

$$\log n = o(\alpha_n). \tag{4.4}$$

In this case the statement of Theorem 3.6 simplifies substantially.

Corollary 4.1. *Assume that the assumptions of Theorem 3.6 are satisfied and that the bandwidths* h_d *and* h_r *satisfy (4.1) - (4.4), then it follows for any* x *with* $m'(x) > 0$

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

where the quantities a_{K_d,K_r}, b_{K_d} *and* $s^2(x)$ *are defined in Theorem 3.6. Moreover, if* m' *is Hölder continuous of order* γ *and the constant* ε *in condition (4.3) satisfies* $\varepsilon < \frac{2\gamma}{9+6\gamma}$, *then*

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

Note that Corollary 4.1 shows that the variance of the density-regression estimate $\hat{m}_I$ is of order

$$O\Big(\frac{1}{n^{2/3}\alpha_n}\Big),$$

while the bias is of order

$$o\Big(\frac{\alpha_n}{n^{1/3}}\Big).$$

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.

Acknowledgements. The authors are grateful to Isolde Gottschlich 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.

References

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

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

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.

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 (2005). A simple nonparametric estimator of a monotone regression function. Technical report, Department of Mathematics.

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

I. Gijbels (2003). Monotone regression. Discussion paper 0334, Institute de Statistique, Université Catholique de Louvain. http://www.stat.ucl.ac.be.

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

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

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

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

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

E. Mammen, C. Thomas-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.

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)

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.

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