Strictly monotone and smooth nonparametric regression for two or more variables

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

FAX: +49 234 3214 559

Regine Scheder

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: regine.scheder@rub.de

June 2, 2005

Abstract

In this article a new monotone nonparametric estimate for a regression function of two or more variables is proposed. The method starts with an unconstrained nonparametric regression estimate and uses successively one-dimensional isotonization procedures. In the case of a strictly monotone regression function, it is shown that the new estimate is first order asymptotic equivalent to the unconstrained estimate, and asymptotic normality of an appropriate standardization of the estimate is established. Moreover, if the regression function is not monotone in one of its arguments, the constructed estimate has approximately the same L^p-norm as the initial unconstrained estimate. The methodology is also illustrated by means of a simulation study, and two data examples are analyzed.

AMS Subject Classification: 62G05, 62G20
Keywords and Phrases: multivariate nonparametric regression, isotonic regression, order restricted inference, nondecreasing rearrangement

1   Introduction

The estimation of relationships among observable variables is an important problem in many areas of applied statistics. The dependence between a real random variable Z and an explanatory variable T can often be decomposed into a deterministic and stochastic part, that is

$$(1.1) \qquad\qquad Z = m(T) + \sigma(T)\varepsilon,$$

where m and σ are smooth (unknown) functions (the regression and variance function), ε denotes a random error, and T is the observed value of independent variables. In many settings, there

1

are restrictions on the direction of the relationships between variables, but not on the particular functional form of their relationships. Typical examples appear in economics where curvature and monotonicity apply to indirect utility, expenditure, production, profit, and cost function [see, e.g., Gallant and Golub (1984), Matzkin (1994), or Aït-Sahalia and Duarte (2003) among many others] or in medicine where the probability of contracting a certain disease (say cancer) depends monotonically on certain factors (say smoking frequency, drinking frequency, and weight). The present paper is devoted to the problem of constructing a smooth nonparametric estimate of the regression function m which is strictly monotone in all (or a subset) of its arguments.

The common approach to obtain monotone estimates in nonparametric regression with an one-dimensional predictor applies isotonic regression, which is easily implemented for functions of one variable using the pool-adjacent violator algorithm (PAVA). Various modifications of this algorithm have been proposed in the literature [see Brunk (1955), Mukerjee (1988), Mammen (1991), Friedman and Tibshirani (1984), Ramsay (1988), or Hall and Huang (2001) among many others]. However for regression functions of more than one variable, the problem of constructing monotone estimates is substantially more difficult, and good algorithms are not available except in special cases. Most of the literature deals with algorithms for multi-argument isotonic regression on a grid [see Gebhardt (1970), Dykstra and Robertson (1982), Lee (1983), or Qian and Eddy (1996) among others], but to the knowledge of the authors, the presence of multivariate random predictors has not been discussed so far in the literature. Because of the computational difficulties, several authors propose alternative methodologies such as additive isotonic regression [see Bacchetti (1989)] or hybrid-type estimators [see Mukarjee and Stern (1994)].

It is the purpose of the present paper to construct a smooth estimate for the function m in the nonparametric regression model (1.1) which is strictly monotone in all (or a subset) of its arguments. The approach uses successively one dimensional strictly monotone estimates, which were recently proposed by Dette et al. (2005) in the case of an one-dimensional explanatory variable. The procedure applies an unconstrained nonparametric estimate of the regression function as preliminary estimate which is successively monotonized in each variable without changing the monotonicity of the other directions. The resulting estimate is strictly isotone in each of the specified coordinates and has the same smoothness properties as the unconstrained estimate of the initial step. This is an important advantage to the currently available procedures; because in the literature, isotonic regression has been critizised for two reasons. On the one hand, practitioners do not believe in the "flat" spots of the resulting estimators [see Wright (1978)]; on the other hand, there usually appears some lack of smoothness in the resulting isotone estimates [see Hall and Huang (2001)]. The method is carefully described in Section 2, where also some of its basic properties are discussed. In Section 3, we prove asymptotic normality of the new estimate and demonstrate that it has precisely the same first order asymptotic properties as the unconstrained estimate used as preliminary estimate. In Section 4, we present a small simulation study and two data examples in order to illustrate the applicability of our method. Finally, all technical details will be given in the Appendix.

2   A strictly monotone regression estimate

In order to fix ideas, we describe a general procedure of monotonizing a given function, lately introduced by Dette et al. (2005). To be precise, let $g : [0, 1] \to \mathbb{R}$ denote an arbitrary function

on a given interval, say $[0, 1]$, we define for $t \in \text{Im}(g)$

$$(2.1) \qquad g_I^{-1}(t) = \int_0^1 I\{g(x) \leq t\} dx.$$

Note that this function is always increasing and coincides with the inverse of the function g, say g^{-1}, in the case where g is strictly increasing. Some properties of a related function have been discussed in the context of nondecreasing/nonincreasing rearrangements [see, e.g., Ryff (1965), Ryff (1970), and Bennett and Sharpley (1988) among others]. The function g_I^{-1} is not necessarily smooth, but smoothing can easily be accomplished by considering the function

$$(2.2) \qquad g_I^{-1}(t, h_d) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\Big(\frac{g(x) - u}{h_d}\Big) du\, dx,$$

where h_d is a bandwidth satisfying $h_d \to 0$ with increasing sample size, and K_d is a positive two times continuously differentiable, symmetric kernel with compact support, say $[-1, 1]$. Note that for $h_d \to 0$, we have

$$(2.3) \qquad \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\Big(\frac{g(x) - u}{h_d}\Big) du\, dx = \int_0^1 I\{g(x) \leq t\} dx + o(1) \;=\; g_I^{-1}(t) + o(1),$$

and that for a positive kernel K_d, the function $g_I^{-1}(t, h_d)$ is always strictly increasing, independently whether the original function g has this property because of its positive derivative. For computational purposes, we also consider an approximation of the integral $g_I^{-1}(t, h_d)$, i.e.

$$(2.4) \qquad g_N^{-1}(t, h_d) = \frac{1}{N h_d} \sum_{i=1}^N \int_{-\infty}^t K_d\Big(\frac{g(\frac{i}{N}) - u}{h_d}\Big) du$$

and its inverse $g_N(t, h_d)$. The role of the inverse of g_I^{-1} or $g_I^{-1}(t, h_d)$ as a monotone and smooth approximation of the function g is carefully discussed in Dette et al. (2005). In particular, the following result is useful for our purposes.

Lemma 2.1. *If the function g is strictly increasing and two times continuously differentiable, we have for any $t \in (g(0), g(1))$ with $g'(g^{-1}(t)) > 0$*

$$(2.5) \qquad g_N^{-1}(t, h_d) \;=\; g^{-1}(t) + \kappa_2(K_d) h_d^2 (g^{-1})''(t) + o(h_d^2) + O\Big(\frac{1}{N h_d}\Big),$$

$$(2.6) \qquad g_N(t, h_d) \;=\; g(t) + \kappa_2(K_d) h_d^2 \frac{g''(t)}{(g'(t))^2} + o(h_d^2) + O\Big(\frac{1}{N h_d}\Big),$$

where the constant $\kappa_2(K_d)$ is given by $\kappa_2(K_d) = \frac{1}{2} \int_{-1}^1 v^2 K_d(v) dv$.

In the present context, we will use this concept successively to obtain monotone estimates of a smooth regression function for two or more variables. For the sake of transparency, we restrict ourselves to the problem of estimating regression functions with two arguments, but the corresponding generalization to the higher dimensional case is obvious [see also Remark 2.7 at the end of this section]. We further assume that the function m in (1.1) is increasing in both arguments.

3

More precisely, we assume that n independent observation according to the model

$$(2.7) \qquad Z_i = m(X_i, Y_i) + \sigma(X_i, Y_i)\varepsilon_i$$

are available. $\{(T_i, Z_i)\}_{i=1}^n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$ is a sample of i.i.d. observations, where $T_i = (X_i, Y_i)$ has a two times continuously differentiable positive density, say f, supported on the square $[0,1]^2$. $\sigma : [0,1]^2 \to \mathbb{R}^+$ is a continuous variance function, and the regression function $m : [0,1]^2 \to \mathbb{R}$ is two times continuously differentiable with respect to both arguments. We finally assume that the random variables (X_i, Y_i) and ε_i are independent such that $E[\varepsilon_i] = 0$, $E[\varepsilon_i^2] = 1$, and $E[\varepsilon_i^4] \leq c < \infty$ for some constant $c > 0$. To obtain strictly monotone estimates, the method proposed in this paper is based on a successive monotonization of an unconstrained preliminary estimate with respect to each of its arguments. For the preliminary estimate, any smoothing procedure (e.g. smoothing splines, kernel estimates or local polynomials) could be used, but for the sake of transparency and definiteness, we restrict ourselves to the classical Nadaraya-Watson estimate, that is

$$(2.8) \qquad \hat{m}(x,y) = \frac{\sum_{i=1}^n K_r\left(\frac{X_i-x}{h_{r,1}}, \frac{Y_i-y}{h_{r,2}}\right) Z_i}{\sum_{i=1}^n K_r\left(\frac{X_i-x}{h_{r,1}}, \frac{Y_i-y}{h_{r,2}}\right)},$$

where K_r is a two dimensional square integrable kernel with compact support, say $[-1,1]^2$, existing second moments satisfying

$$\kappa_2(K_r) = \frac{1}{2}\int_{[0,1]^2} x^2 K_r(x,y)dxdy = \frac{1}{2}\int_{[0,1]^2} y^2 K_r(x,y)dxdy,$$

$$(2.9)$$

$$0 = \int_{[0,1]^2} xy K_r(x,y)dxdy,$$

and $h_{r,1}, h_{r,2}$ denote the bandwidths for each coordinate. The isotonization of the nonparametric regression estimate $\hat{m}$ is now performed in four steps using a kernel density estimate with positive kernel K_d and bandwidth h_d.

- **Step 1** *(isotonization with respect to the first coordinate)*

  For each $y \in (0,1)$ define

  $$(2.10) \qquad \hat{m}_{I_1}^{-1}(z|y) = \frac{1}{Nh_d}\sum_{i=1}^N \int_{-\infty}^z K_d\left(\frac{\hat{m}(\frac{i}{N},y)-u}{h_d}\right)du,$$

  then $\hat{m}_{I_1}^{-1}(z \mid y)$ is obviously a strictly increasing function of z (for fixed y) (provided that N is sufficiently large) on the interval $(m(0,y), m(1,y))$. We also note that the summation in (2.10) can be interpreted as a quadrature formula for the corresponding integral

  $$\frac{1}{h_d}\int_0^1 \int_{-\infty}^z K_d\left(\frac{\hat{m}(x,y)-u}{h_d}\right)dudx.$$

- **Step 2** *(inversion with respect to the first coordinate)*

For each (fixed) $y \in (0,1)$ the inverse of the strictly increasing function $\hat{m}_{I_1}^{-1}(z \mid y)$ (as a function of z) is calculated and denoted by $\hat{m}_{I_1}(x,y)$. This function is strictly increasing with respect to its first argument.

- **Step 3** *(isotonization with respect to the second coordinate)*

  For each (fixed) $x \in (0,1)$ define

$$(2.11) \qquad \hat{m}_{I_{1,2}}^{-1}(t|x) = \frac{1}{Nh_d} \sum_{j=1}^{N} \int_{-\infty}^{t} K_d\Big(\frac{\hat{m}_{I_1}(x,\frac{j}{N}) - u}{h_d}\Big) du,$$

  which is again a strictly increasing function of t (for fixed x) and denoted by $\hat{m}_{I_{1,2}}^{-1}$.

- **Step 4** *(inversion with respect to the second coordinate)*

  For each (fixed) $x \in (0,1)$ the inverse of the strictly increasing function $\hat{m}_{I_{1,2}}^{-1}(y \mid x)$ (considered as a function of y) is calculated and denoted by $\hat{m}_{I_{1,2}}(x,y)$. This function is strictly increasing with respect to both arguments.

In the following, we will discuss some properties of this procedure, which demonstrate that the estimate $\hat{m}_{I_{1,2}}$ is a good smooth approximation of the regression function, which is increasing with respect to both of its arguments. We will begin with a result which provides some monotonicity properties of the function $\hat{m}_{I_1}^{-1}(x \mid y)$ with respect to both arguments x <u>and</u> y. The proof can be found in the Appendix.

Lemma 2.2. *Assume that the preliminary estimate $\hat{m}(x,y)$ is strictly increasing with respect to its second argument, then for sufficiently large N the estimate $\hat{m}_{I_1}^{-1}(x \mid y)$ obtained in the first step of the above isotonization procedure is*

- *strictly increasing in x for fixed y*

- *strictly decreasing in y for fixed x*

The preceding lemma is the basic tool to prove the statement made at the end of Step 4 of our procedure, i.e., the monotonicity of the regression estimate $\hat{m}_{I_{1,2}}$ with respect to both arguments. In other words, the monotonicity of the estimate $\hat{m}_{I_1}$ with respect to the first coordinate obtained in Step 1 and 2 of the procedure are not destroyed by the isotonization of $\hat{m}_{I_1}$ with respect to the second coordinate.

Lemma 2.3. *For sufficiently large N, the estimate $\hat{m}_{I_{1,2}}$ is strictly increasing with respect to both arguments.*

A few further remarks may be appropriate at this point for a better understanding of the proposed procedure. If the regression function m in (2.7) is strictly increasing with respect to both

arguments, it is heuristically clear that the estimate from the first step is an approximation of the quantity

$$(2.12) \qquad m_{N_1}^{-1}(t|y) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{t} K_d\Big(\frac{m(\frac{i}{N}, y) - u}{h_d}\Big) du$$

$$\approx \int_0^1 I\{m(x,y) \leq t\} dx =: m_1^{-1}(t \mid y),$$

where the approximation is justified if $N \to \infty, h_d \to 0$ sufficiently fast. For fixed $y \in (0,1)$ the function m is strictly isotone with respect to the first argument. It follows that $m_1^{-1}(\cdot \mid y)$ is the inverse of $m(\cdot, y)$ (for fixed $y \in (0,1)$), and the inversion in Step 2 of the algorithm reproduces m. Because the same argument applies to the third and fourth step of our procedure, it is intuitively clear that the estimate $\hat{m}_{I_{1,2}}$ is a consistent estimate of a strictly isotone regression function in two variables. These heuristic arguments will be made precise in the following section, where the asymptotic properties of the new estimate are discussed.

The following result gives some information regarding the behaviour of the estimate $\hat{m}_{I_{1,2}}$ in the case where the regression function is not necessarily increasing with respect to both of its arguments.

Theorem 2.4. *Assume that in Step 1 and 3 of the isotonization procedure the definitions*

$$(2.13) \qquad \tilde{m}_{I_1}^{-1}(z \mid y) = \int_0^1 I\{\hat{m}(u,y) \leq z\} du$$

$$(2.14) \qquad \tilde{m}_{I_{1,2}}^{-1}(t \mid x) = \int_0^1 I\{\tilde{m}_{I_1}(x,u) \leq t\} du$$

are used instead of (2.10) and (2.11), respectively, and define $\tilde{m}_{I_{1,2}}$ as the inverse of the function $\tilde{m}_{I_{1,2}}^{-1}$, then we have for all $p > 0$

$$\int_{[0,1]^2} |\tilde{m}_{I_{1,2}}(x,y)|^p dxdy = \int_{[0,1]^2} |\hat{m}(x,y)|^p dxdy.$$

Note that by Theorem 2.4 and the approximation (2.12) the original estimate $\hat{m}_{I_{1,2}}$ has approximately (if N is sufficiently large and h_d sufficiently small) the same L^p-norm as the unconstrained Nadaraya-Watson estimate $\hat{m}$. Consequently, if the regression function m is not isotone in both coordinates, the estimate $\hat{m}_{I_{1,2}}$ converges to a function, say $m_{I_{1,2}}$ with the same L^p-norm as the function m for all $p > 0$.

Remark 2.5. Note that for finite sample sizes the estimates may differ if the order of monotonization is interchanged. To be precise, let $\hat{m}_{I_{2,1}}$ denote the estimate which is obtained by first isotonizing with respect to the second argument and then with respect to the first argument. Then for a specific point $(x,y) \in (0,1)^2$ the estimates $\hat{m}_{I_{1,2}}(x,y)$ and $\hat{m}_{I_{2,1}}(x,y)$ may differ slightly. In order to avoid effects caused by the order of isotonization, the average estimate

$$\hat{m}_I(x,y) = (\hat{m}_{I_{1,2}}(x,y) + \hat{m}_{I_{2,1}}(x,y))/2$$

6

could be used. On the other hand, it is shown in Section 3 that the estimates $\hat{m}_{1,2}$ and $\hat{m}_{2,1}$ are first order asymptotic equivalent if the regression function m is monotone with respect to both arguments. Therefore for moderate sample sizes substantial differences between the estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ will rarely be observable in this case [see also Section 4 for some numerical examples].

Remark 2.6. It is also worth mentioning that the procedure can easily be adapted for the nonparametric estimation of regression functions which are increasing in some and decreasing in other components. More precisely, consider the model (2.7) and assume that the regression function m is antitone in x and isotone in y. Then a shape restricted strictly antitone (with respect to the first argument) estimate for the regression function m is obtained by using the estimate

$$(2.15) \qquad \hat{m}_{A_1}^{-1}(z|y) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_z^{\infty} K_d\Big(\frac{\hat{m}(\frac{i}{N}, y) - u}{h_d}\Big) du$$

in Step 1 and 2 of our procedure. Obviously, $\hat{m}_{A_1}$ is decreasing with respect to the first argument z, and an application of Step 3 and 4 (i.e. isotonization with respect to the second argument) yields a smooth estimate with the desired monotonicity properties, that is, an antitone and isotone function with respect to the first and second argument, respectively.

Remark 2.7. The extension of the procedure to the monotone estimation of a regression depending on more than two variables is straightforward. For each coordinate, for which the relation between response and variable is assumed to be monotone, a corresponding isotonization step and inversion is performed. Note, however, that nonparametric estimation of a multivariate regression function is a difficult problem due to the "curse of dimensionality".

On the other hand, the application of parametric regression models avoids the problem of a high dimensional predictor and unconstrained parametric estimates can easily be isotonized by the method described in Step 1 - 4 of this section (we only replace the nonparametric initial estimate by its corresponding parametric counterpart).

3 Asymptotic theory

In this section, we present some asymptotic results for the statistic $\hat{m}_{I_{1,2}}$ introduced in the previous section. In particular, we prove consistency and asymptotic normality of the new estimate. For this, we need several conditions on the bandwidths, that is

$$(3.1) \qquad h_{r,i} \to 0, \ (i = 1, 2), \quad nh_{r,1}h_{r,2} \to \infty, \ h_d \to 0, \ Nh_d \to \infty$$

$$(3.2) \qquad n(h_{r,1} + h_{r,2})^6 = O(1), \ n = O(N),$$

$$(3.3) \qquad \lim_{n \to \infty} \frac{h_{r,i}}{h_d} = \infty \ , i = 1, 2,$$

$$(3.4) \qquad \frac{1}{nh_{r,1}h_{r,2}h_d^2} = o(1),$$

$$(3.5) \qquad \frac{\log(h_{r,1} + h_{r,2})^{-1}}{nh_{r,1}h_{r,2}h_d^2} = o(1)\,.$$

The last bandwidth condition includes condition (3.4) and is used in the proofs of Theorem 3.2 and Theorem 3.4 where the asymptotic normality of the inverse of the functions $\hat{m}_{I_1}^{-1}$ and $\hat{m}_{I_{1,2}}^{-1}$, respectively, is etablished. This condition is a slight improvement to the assumption made in Dette et al. (2005).

Recall that for a strictly increasing function in the first argument $m_1^{-1}(t|y)$ is the inverse of m if the inversion is performed with respect to first argument where the second argument is fixed. Similarly $m_2^{-1}(t|x))$ denotes the inverse with respect to the second argument for fixed first argument. Moreover, $\frac{\partial}{\partial x_1} m(x,y)$ and $\frac{\partial}{\partial y_1} m(x,y)$ indicate the partial derivatives regarding the first and second argument, respectively. The following theorem is proven in the appendix.

Theorem 3.1. *Assume that the conditions stated at the beginning of Section 2, and the assumptions (3.1) - (3.4) are satisfied. If the regression function m in (2.7) is strictly increasing with respect to x, then we have for any fixed y and for all $t \in (m(0,y), m(1,y))$ with $\frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y) > 0$*

$$\sqrt{nh_{r,1}h_{r,2}}\Big(\hat{m}_{I_1}^{-1}(t|y) - m_1^{-1}(t|y) + b_1(t,y)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_1^2(t,y)),$$

where the asymptotic bias and variance are given by

$$(3.6) \qquad b_1(t,y) = \frac{\kappa_2(K_r)}{\frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y)} \Big[h_{r,1}^2 \frac{\partial^2}{\partial x_1^2} m(m_1^{-1}(t|y), y) + h_{r,2}^2 \frac{\partial^2}{\partial y_1^2} m(m_1^{-1}(t|y), y)$$

$$+ 2h_{r,1}^2 \frac{\frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y) \frac{\partial}{\partial x_1} f(m_1^{-1}(t|y), y)}{f(m_1^{-1}(t|y), y)}$$

$$+ 2h_{r,2}^2 \frac{\frac{\partial}{\partial y_1} m(m_1^{-1}(t|y), y) \frac{\partial}{\partial y_1} f(m_1^{-1}(t|y), y)}{f(m_1^{-1}(t|y), y)} \Big],$$

$$(3.7) \qquad s_1^2(t,y) = \frac{\sigma^2(m_1^{-1}(t|y), y)}{f(m_1^{-1}(t|y), y)(\frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y))^2} \int_{[-1,1]^2} K_r^2(u,v) du dv,$$

respectively, and the quantity $\kappa_2(K_r)$ is defined in (2.9).

Our next result gives a corresponding statement of asymptotic normality of the estimate $\hat{m}_{I_1}(\cdot, y)$, which is obtained by the inversion of the function $\hat{m}_{I_1}^{-1}(\cdot \mid y)$ considered as function of the variable x for fixed y. The proof follows similar arguments as presented in Dette et al. (2005) for the one-dimensional case, but we will give some hints in the appendix how the improved bandwidth condition (3.5) can be achieved.

Theorem 3.2. *If the assumptions of Theorem 3.1 and condition (3.5) are satisfied, then it follows for any $y \in (0,1)$ and $x \in (0,1)$ with $\frac{\partial}{\partial x_1} m(x,y) > 0$ that*

$$(3.8) \qquad \sqrt{nh_{r,1}h_{r,2}}\Big(\hat{m}_{I_1}(x,y) - m(x,y) - b(x,y)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x,y)),$$

where the asymptotic bias and variance are given by

$$(3.9) \qquad b(x,y) = \kappa_2(K_r)\Big[h_{r,1}^2 \frac{\partial^2}{\partial x_1^2} m(x,y) + h_{r,2}^2 \frac{\partial^2}{\partial y_1^2} m(x,y)$$

8

$$+ 2h_{r,1}^2 \frac{\frac{\partial}{\partial x_1}m(x,y)\frac{\partial}{\partial x_1}f(x,y)}{f(x,y)} + 2h_{r,2}^2 \frac{\frac{\partial}{\partial y_1}m(x,y)\frac{\partial}{\partial y_1}f(x,y)}{f(x,y)}\Bigg],$$

$$(3.10) \qquad s^2(x,y) = \frac{\sigma^2(x,y)}{f(x,y)}\int_{[-1,1]^2} K_r^2(u,v)\,du\,dv,$$

respectively, and the quantity $\kappa_2(K_r)$ is defined in (2.9).

Step 3 of the proposed algorithm in Section 2 uses the implicitly defined estimate $\hat{m}_{I_1}$ again as a preliminary estimate where now the role of the first and second argument is interchanged. Our next theorem gives the asymptotic normality of the resulting estimate $\hat{m}_{I_{1,2}}^{-1}$. The proof is more complicated because, in contrast to the estimate $\hat{m}_{I_1}^{-1}$ in (2.10) (which uses the Nadaraya-Watson estimate as preliminary estimate), the preliminary estimate $\hat{m}_{I_1}$ is only defined implicitly as the inverse of the function $\hat{m}_{I_1}^{-1}$. The details are given in the Appendix.

Theorem 3.3. *If the assumptions of Theorem 3.2 are satisfied, and the regression function m in model (2.7) is strictly increasing with respect to both arguments, then it follows for any fixed x and for all $t \in (m(x,0), m(x,1))$ with $\frac{\partial}{\partial y_1}m(x, m_2^{-1}(t|x)) > 0$ and $\frac{\partial}{\partial x_1}m(x, m_2^{-1}(t|x)) > 0$*

$$(3.11) \qquad \sqrt{nh_{r,1}h_{r,2}}\Big(\hat{m}_{I_{1,2}}^{-1}(t|x) - m_2^{-1}(t|x) + b_2(x,t)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_2^2(x,t)),$$

where the asymptotic bias and variance are given by

$$(3.12) \quad b_2(x,t) = \frac{\kappa_2(K_r)}{\frac{\partial}{\partial y_1}m(x, m_2^{-1}(t|x))}\Big[h_{r,1}^2 \frac{\partial^2}{\partial x_1^2}m(x, m_2^{-1}(t|x)) + h_{r,2}^2 \frac{\partial^2}{\partial y_1^2}m(x, m_2^{-1}(t|x))$$

$$+ 2h_{r,1}^2 \frac{\frac{\partial}{\partial x_1}m(x, m_2^{-1}(t|x))\frac{\partial}{\partial x_1}f(x, m_2^{-1}(t|x))}{f(x, m_2^{-1}(t|x))}$$

$$+ 2h_{r,2}^2 \frac{\frac{\partial}{\partial y_1}m(x, m_2^{-1}(t|x))\frac{\partial}{\partial y_1}f(x, m_2^{-1}(t|x))}{f(x, m_2^{-1}(t|x))}\Bigg],$$

$$(3.13) \quad s_2^2(x,t) = \frac{\sigma^2(x, m_2^{-1}(t|x))}{f(x, m_2^{-1}(t|x))(\frac{\partial}{\partial y_1}m(x, m_2^{-1}(t|x)))^2}\int_{[-1,1]^2} K_r^2(u,v)\,du\,dv,$$

respectively, and the quantity $\kappa_2(K_r)$ is defined in (2.9).

Our final asymptotic result refers to the stochastic properties of the estimate $\hat{m}_{1,2}$ obtained in Step 4 of our algorithm. The proof is derived from Theorem 3.3 by the same arguments as used for the proof of Theorem 3.2 and therefore omitted.

Theorem 3.4. *If the assumptions of Theorem 3.3 are satisfied, then we have for any x and y with $\frac{\partial}{\partial x_1}m(x,y) > 0$, $\frac{\partial}{\partial y_1}m(x,y) > 0$*

$$(3.14) \qquad \sqrt{nh_{r,1}h_{r,2}}\Big(\hat{m}_{I_{1,2}}(x,y) - m(x,y) - b(x,y)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x,y)),$$

where the asymptotic bias and variance are given by (3.9) and (3.10), respectively.

9

Remark 3.5. It follows from Theorem 3.4 that the estimate $\hat{m}_{I_{1,2}}$ is first order asymptotic equivalent the to the unconstrained Nadaraya-Watson estimate $\hat{m}$ [see Härdle et al. (2004) and observe the symmetry condition for the kernel K_r in (2.9)]. Similarly, by Theorem 3.2, the estimate $\hat{m}_{I_1}$ is first order asymptotic equivalent to $\hat{m}$ and $\hat{m}_{I_{1,2}}$. Thus each pair of steps in the algorithm produces a first order asymptotic equivalent estimate with one additional isotonized coordinate. Moreover, Theorem 3.4 also shows that from an asymptotic point of view the order of isotonization in the procedure can be interchanged. In other words, the nonparametric regression estimate $\hat{m}_{I_{2,1}}$ obtained by first isotonizing with respect to the second and then with respect to the first coordinate exhibits the same asymptotic behaviour as described for the estimate $\hat{m}_{I_{1,2}}$ in Theorem 3.4.

Remark 3.6. It is worthy of mention that similar asymptotic results can be obtained if alternative unconstrained nonparametric regression estimates are used as preliminary estimate $\hat{m}$. Consider, for example, the local linear estimator for the function m in the regression model (2.7), then a similar analysis as presented in the Appendix shows that the statement (3.14) in Theorem 3.4 is still valid if the bias $b(x, y)$ is replaced by

$$(3.15) \qquad \tilde{b}(x, y) = \kappa_2(K_r)\left[h_{r,1}^2 \frac{\partial^2}{\partial x_1^2} m(x, y) + h_{r,2}^2 \frac{\partial^2}{\partial y_1^2} m(x, y)\right]$$

[see equations (5.8) and (5.9) in the Appendix]. Again the monotonized estimate is first order equivalent to the unconstrained local linear estimate [see e.g. Ruppert and Wand (1994)]. Because of its better performance at the boundary, the local linear estimate was in fact used in our numerical examples, which will be described in the following section.

4 Finite sample properties and data analysis

In this section, we investigate the finite sample properties of the new monotone estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ of the regression function m by means of a simulation study and illustrate the performance of the new procedure in two data examples.

4.1 A small simulation study

In order to avoid boundary effects, we implemented the new procedure using a two dimensional local linear estimator [see Ruppert and Wand (1994)] and the bandwidth

$$(4.1) \qquad h_{r,1} = h_{r,2} = \left(\frac{\tilde{\sigma}^2}{n}\right)^{1/6}$$

where $\tilde{\sigma}^2 = \int \sigma^2(u, v) du dv$ denotes the integrated variance and n is the sample size. For the kernel K_r, we used a product kernel based on two Epanechnikov kernels $k(x) = \frac{3}{4}(1 - x^2)I_{[-1,1]}(x)$ and the same kernel was used for K_d in step 1 and 3 of the procedure. The number N of nodes in (2.10) used for the evalutation of the integral is chosen as $N = 51$ and the bandwidth in step 1 and 3 is given by $h_d = h_r^3$.

Example 4.1 We consider a regression model with

$$(4.2) \qquad m(x,y) = \frac{1}{2}(y + \frac{1}{6\pi}\sin(6\pi y))(1 + (2x-1)^3),$$

which is obviously strictly isotone in each coordinate. For the error distribution in model (2.7) we use a normal distribution, whereas the variance function is constant and given by $\sigma^2 = 0.1$. The sample size is $n = 400$ and the design density is a uniform distribution on the square $[0,1]^2$. In Figure 1, we display a typical picture of the regression estimates based on one simulation. In order to avoid the domination by boundary effects, all graphical results in this example will be displayed on the square $[0.05, 0.95]^2$.

We observe that the local linear estimate is obviously not isotone with regard to both arguments [see the upper right panel in Figure 1]. The estimates $\hat{m}_{I_1}$ and $\hat{m}_{I_2}$ with respect to the coordinates x and y, respectively, look substantially different because isotonization is only performed with respect to one coordinate [the middle panel in Figure 1]. On the other hand, substantial differences between the isotone estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ (obtained by interchanging the order of isotonization) are not visible as predicted by our asymptotic theory [see the lower panel in Figure 1].

For a better understanding of our procedure and a direct comparison of the two estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$, we present also two dimensional plots of the estimates of the regression function $m(x,y)$, where one coordinate has been fixed. In Figure 2, we show three typical simulations of the estimate $\hat{m}_{I_{1,2}}$, $\hat{m}_{I_{2,1}}$, and the "true" curve, where the coordinate y is fixed at 0.3, 0.5, and 0.7. The left part of the figure correspond to the curves obtained by the estimate $\hat{m}_{I_{1,2}}$, whereas the right part shows the curves for the estimate $\hat{m}_{I_{2,1}}$. The corresponding results for fixed $x = 0.3$, 0.5, and 0.7 are depicted in Figure 3. We observe a reasonable performance of the estimates in all cases and again no substantial differences between the estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ obtained by interchanging the order of isotonization.

Example 4.2 In this example, we compare the isotonized estimates with the unconstrained local linear estimates. Again to avoid the domination by boundary effect, we calculate the mean squared error of both estimates over the square $[0.1, 0.9] \times [0.1, 0.9]$. In Figure 4, we display the ratios

$$(4.3) \qquad r(x,y) = \frac{MSE[\hat{m}(x,y)]}{MSE[\hat{m}_{1,2}(x,y)]}$$

in this region for the regression function (4.2) and the functions

$$(4.4) \qquad m(x,y) = \frac{1}{2}[x + y + \frac{1}{6\pi}\sin(6\pi(x+y))]$$

$$(4.5) \qquad m(x,y) = \sin\left(\frac{\pi}{4}(x+y)\right)$$

$$(4.6) \qquad m(x,y) = (2xy + \frac{1}{2}x^2 - 2y^2 + 2y - \frac{1}{2})I\{1 \le 2y \le x+1\}$$
$$+ (2yx - \frac{1}{2}x^2)I\{x \le 2y < 1\} + 2y^2 I\{2y < x\} + xI\{x+1 \le 2y\}.$$
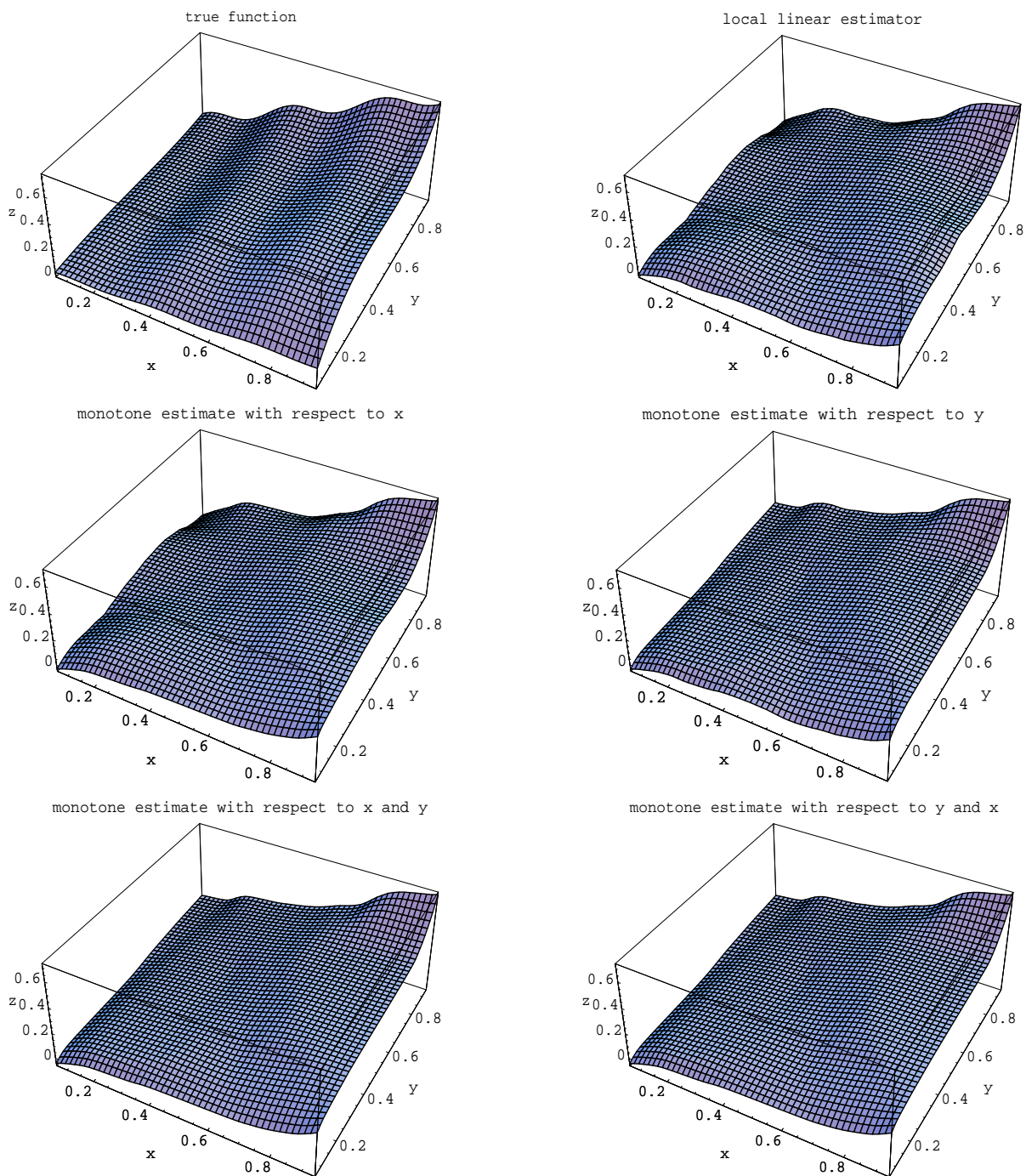
11

Figure 1: *The regression function (4.2) and its different estimates. Upper panel: the regression function m (left) and the local linear estimator $\hat{m}$ (right) based on $n = 400$ observations with variance $\sigma^2 = 0.1$. Middle panel: monotone estimate $\hat{m}_{I_1}$ with respect to the first coordinate (left) and monotone estimate $\hat{m}_{I_2}$ with respect to second coordinate (right). Lower panel: monotone estimates with respect to both coordinates $\hat{m}_{I_{1,2}}$ (left) and $\hat{m}_{I_{2,1}}$ (right).*
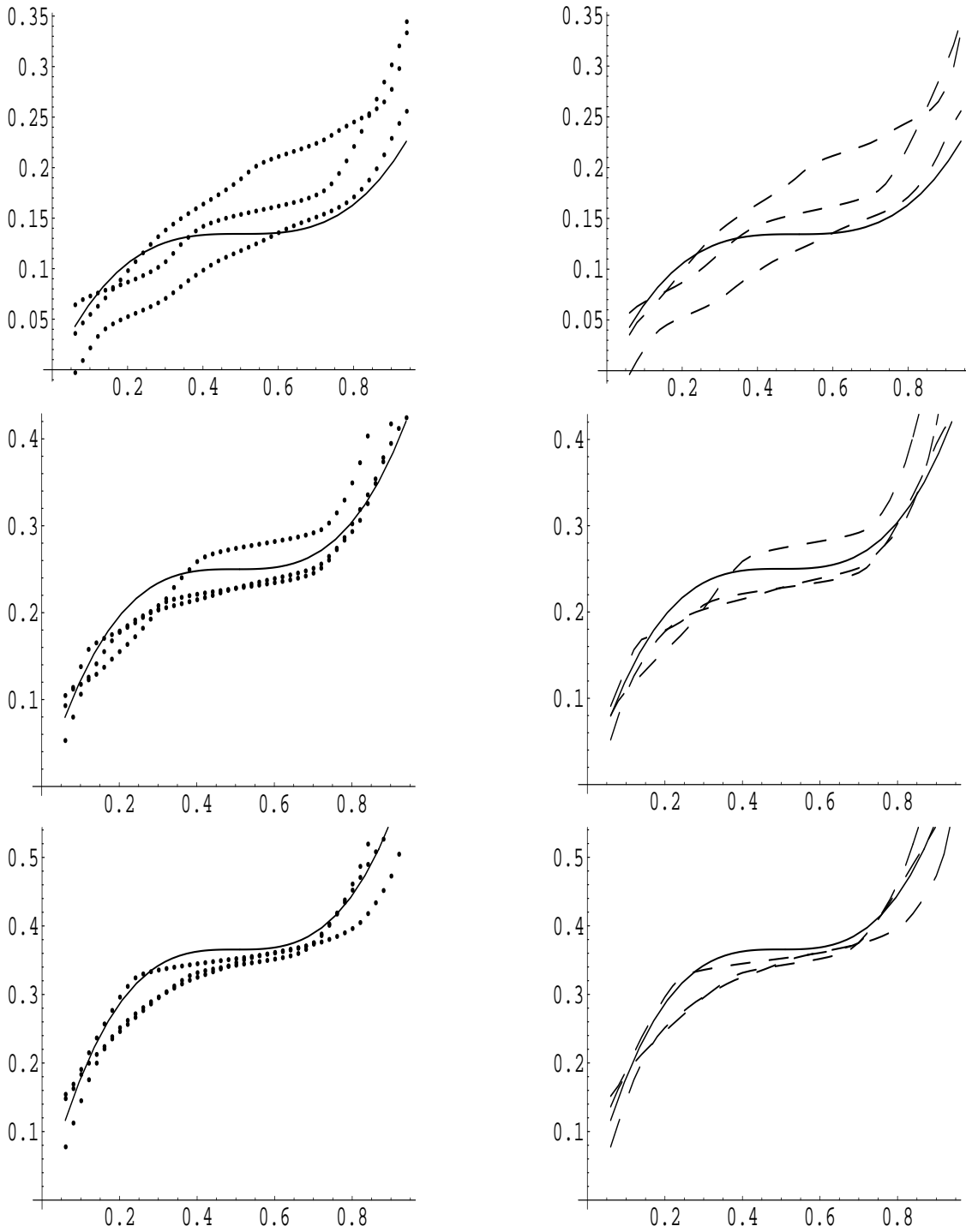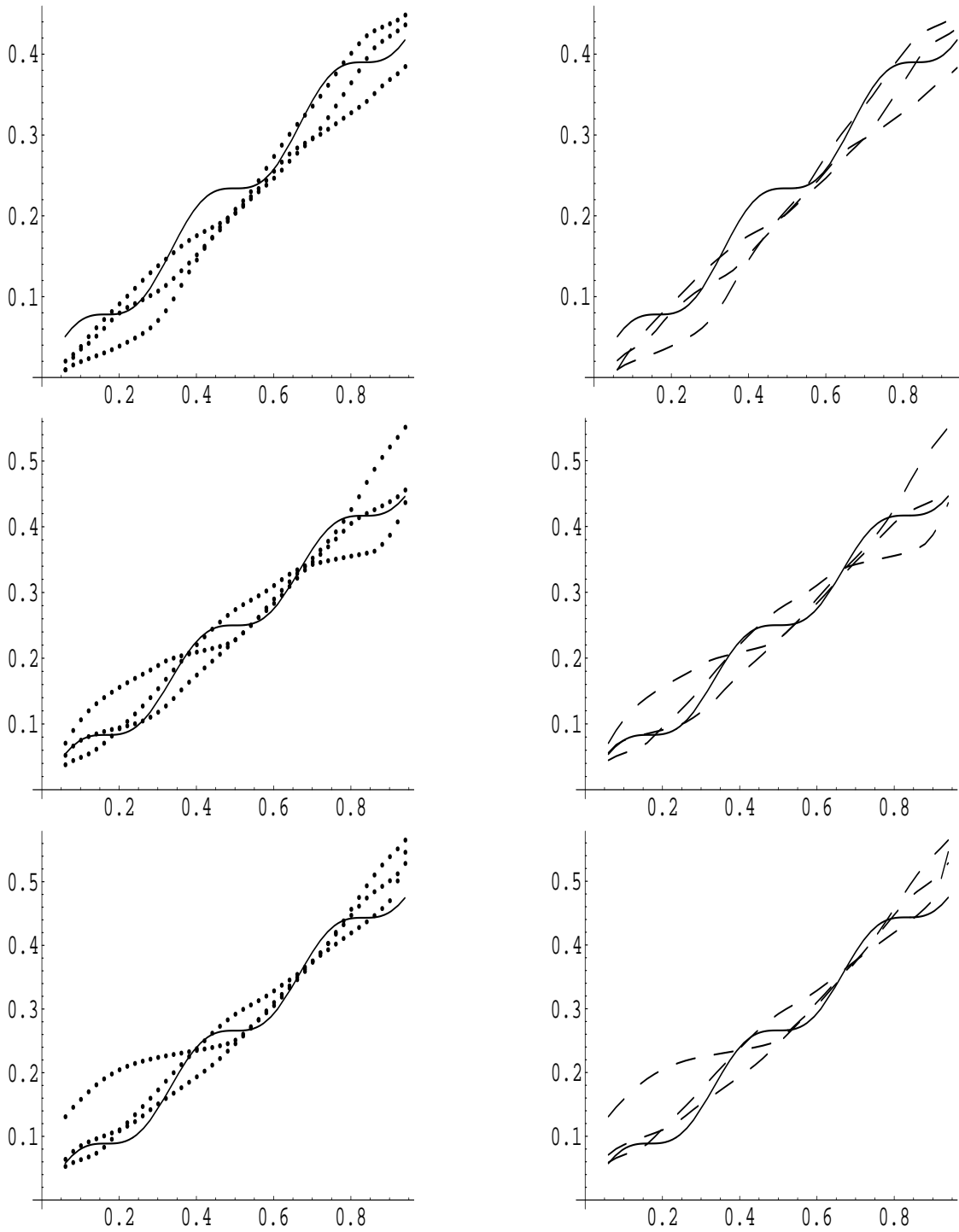
Figure 2: *The regression function m and its monotone estimates $\hat{m}_{I_{1,2}}$ (dotted lines) and $\hat{m}_{I_{2,1}}$ (dashed lines) in comparison to each other. The three estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$, respectively, are obtained from three different simulation runs. The Figure shows the two dimensional functions fixed in y: $y = 0.3$ (upper panel), $y = 0.5$ (middle panel) and $y = 0.7$ (lower panel). The solid curve corresponds to the "true" regression function, whereas the estimates $\hat{m}_{I_{1,2}}$ are represented by the dotted curves and $\hat{m}_{I_{2,1}}$ by the dashed curves.*

Figure 3: *The regression function m and its monotone estimates $\hat{m}_{I_{1,2}}$ (dotted lines) and $\hat{m}_{I_{2,1}}$ (dashed lines) in comparison to each other. The three estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$, respectively, are obtained from three different simulation runs. The Figure shows the two dimensional functions fixed in x: $x = 0.3$ (upper panel), $x = 0.5$ (middle panel) and $x = 0.7$ (lower panel). The solid curve corresponds to the "true" regression function, whereas the estimates $\hat{m}_{I_{1,2}}$ are represented by the dotted curves and $\hat{m}_{I_{2,1}}$ by the dashed curves.*

14

Again the sample size is $n = 400$ and the errors are standard normal with variance $\sigma^2 = 0.1$. The ratio of the mean squared errors was estimated by 1000 simulation runs and is larger than 1 if and only if the monotonized estimate has a better performance. We observe that for all regression functions the surfaces usually exceed the value 1, which indicates that the additional information of monotonicity in the nonparametric regression estimate can improve its finite sample properties. In most cases, the improvement is substantial. For example in model (4.2), there are several regions in the square $[0.1, 0.9]^2$, where the ratio is larger than 1.7. The improvement caused by the monotonization in model (4.4) and (4.5) is not so large but still clearly visible.

Note that model (4.6) corresponds to the distribution function of the random variable $Z = \frac{1}{2}(X + Y)$, where X, Y are independent with a uniform distribution on the interval $[0, 1]$. This model does not satisfy the assumptions of our theoretical results because for $2y \geq x + 1$, we have $m(x, y) = x$, and this function is not strictly increasing with respect to y. Similarly, we have $m(x, y) = 2y^2$ if $2y < x$ in model (4.6). Nevertheless, the statistic $\hat{m}_{I_{1,2}}$ yields also a reliable estimate of the regression function in this case. Observe also that there are some (small) areas, where the unconstrained local linear estimate yields a smaller mean squared error in model (4.6). However, in most cases, the incorporation of the additional information of monotonicity yields a reduction of the mean squared error, in particular, if the regression function is strictly increasing with respect to both arguments.

4.2   Two data examples

To investigate the performance of our procedure in practice, we consider two typical examples for monotone regression functions with two independent variables. In both examples, we used Spokoiny's variance estimate for high-dimensional regression models [see Spokoiny (2002)] to obtain an estimate $\hat{\sigma}^2$ of the integrated variance. Then a two-step cross validation was used to determine the bandwidths h_r and h_d. The first step is to choose the bandwidth h_r. The leave-one-out cross validation procedure was implemented to minimize the cross validation function over an appropriate interval centered at the point $(\hat{\sigma}^2/n)^{1/6}$ to determine h_r [see Härdle (1990, page 152)], where n is the sample size of the datasets. The resulting minimizer is used as the bandwidth h_r to fit the local linear estimator for the regression curve. To choose the bandwidth for the monotonizing procedure, we performed a second cross validation and minimized the function

$$CV(h_d) = n^{-1} \sum_{i=1}^{n} (Z_i - \hat{m_{I,i}}(X_i, Y_i))^2,$$

where

$$\hat{m_{I,i}}(x, y) = (\hat{m}_{I_{1,2},i}(x, y) + \hat{m}_{I_{2,1},i}(x, y))/2$$

is the average of the two monotone estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ (obtained by interchanging the order of isotonization) in which the ith observation is left out. In order to make these examples reproducible, we provide an R-package, called 'monoProc', where an implementation of the monotonization procedure and further examples as well as Example 4.3 can be found.

Example 4.3 A reasonable criterion for obesity is to calculate the bodyfat percentage of a person. The most accurate techniques like underwater weighing to obtain the bodyfat percentage are time consuming and expensive.
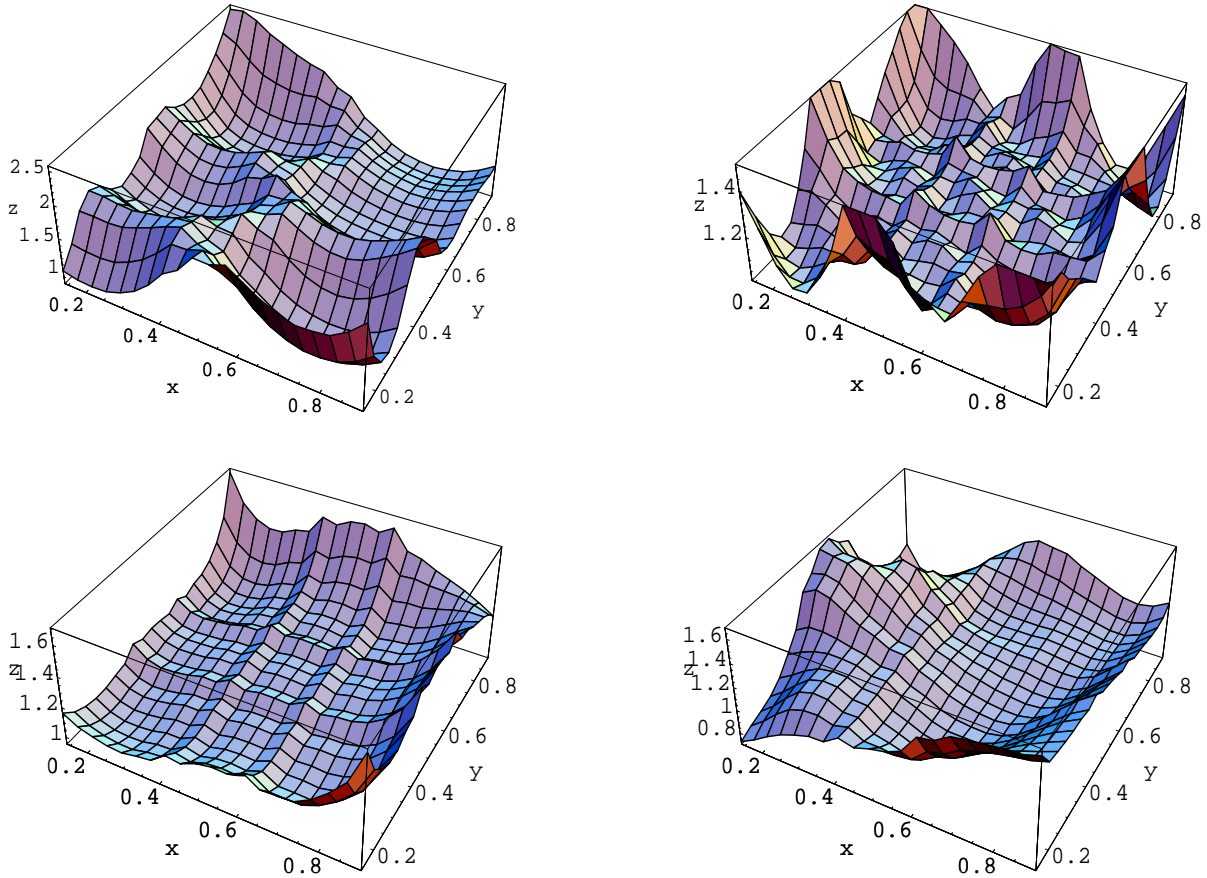
15

Figure 4: *The simulated ratio (4.3) of the mean squared errors of the local linear estimate and its monotonization with respect to the x and y coordinate. The four surfaces correspond to the regression functions (4.2): upper left panel, (4.4): upper right panel, (4.5): lower left panel and (4.6): lower right panel.*

Other methods to estimate this percentage suggest certain predictive equations depending on, e.g., weight, height, age, and several simple body measurements. The dataset, we analysed to give a simple example for a monotone regression function in more than one variable consists of age, weight, height, 10 simple body circumfence measurments, and the bodyfat percentage provided by underwater weighing using Siri's equation recorded for 252 men from 22 to 81 years [see Penrose et al. (1985)]. We restricted ourselves to consider simply weight and height. Certainly, this restriction is simplifying since the Body Mass Index

$$\text{BMI} = \frac{\text{weight(lbs)}}{(\text{height(in)})^2}$$

as an indicator for the bodyfat percentage can be quite problematic [see Summerfield (2001, page 78)]. On the other hand, there are also problems with the accuracy of body circumfence measurements. Weight and height are, in contrast to that, easily obtained quantities. Moreover, it is reasonable that in an average group of men, bodyfat should be an increasing function of weight and a decreasing function of height. It has to be mentioned that two outliers in this dataset have been disregarded [observation 39 and 42]. For the Bodyfat dataset, we obtained $\hat{\sigma}^2 = 39.48902$ using Spokoiny's variance estimate. The cross validation function suggests the bandwidth $h_r = 0.5$ for the unconstrained estimator. In the second cross validation the bandwidth $h_d = 0.75$ for the monotonizing procedure was chosen out of the set $[0.1, 3.0]$ with $CV(0.75) = 30.9634$. In this example, there is a slight difference between the cross validation value of $\hat{m}_{I_{1,2}}$ (31.063877) and $\hat{m}_{I_{2,1}}$ (30.862941). The choice $h_d = 0.75$ corresponds also to the minimum of $CV(h_d)$ if only $\hat{m}_{I_{1,2}}$ is under consideration. By comparison, the local linear estimate leads in this setup to a cross validation value of 30.952571. The (monotone) estimate of the regression function (bodyfat percentage as a function of weight and height) is depicted in Figure 5.

Example 4.4 Often a monotone dependency between education and work experience and the corresponding salary is assumed. To examine the monotonic relationship between the frequency to earn more than $50 000 per year as response and education and age as explanatory variables, we analysed a reasonably clean extract of the 1994 Census database found in Blake and Merz (1998) [see also Kohavi (1996)]. This data consists of 48813 instances after removing duplicates. Since we are only interested in age, education, and the frequency of earning more than $50 000 a year, we derived a dataset with 639 observations, age ranging from 19 to 90 and education levels from 1 to 16. This extract includes only age groups in which at least one person earns more than 50000 per year. Furthermore, we restricted age to 27-70 and analysed

Figure 5: *Isotone estimate of the Bodyfat percentage as a function of weight and age.*

therefore a sample of 516 observations. Spokoiny's variance estimate leads to $\hat{\sigma}^2 = 0.02563633$, and we obtained the bandwidth $h_r = 0.169$ by cross validation. Again we used a second cross validation to find the bandwidth h_d.
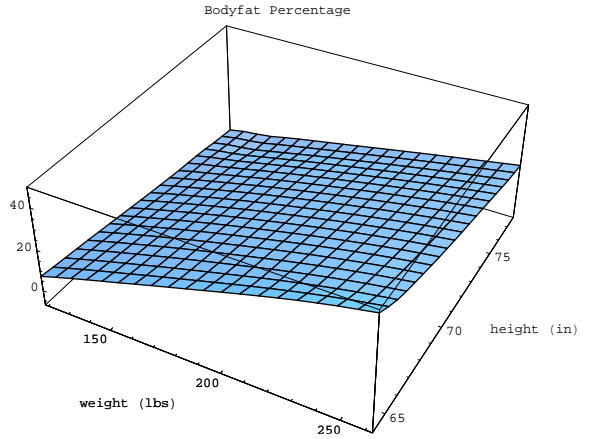
The cross validation function $CV(h_d)$ was minimized over the set $[0.001, 0.03]$. In the following table, a part of these results for the different estimates are recorded. Note that the impact of the choice of the bandwidth h_d on the cross validation function is very small. We decided to use $h_d = 0.017$ for the further analysis, which is a minimizer of $CV(h_d)$ for the average of the two monotonized procedures $\hat{m}_I = (\hat{m}_{I_{12}} + \hat{m}_{I_{21}})/2$. But the difference to the minima of the estimates $\hat{m}_{I_{12}}$ and $\hat{m}_{I_{21}}$, which are attained at $h_d = 0.013$ and in $h_d = 0.019$ respectively, is negligible.

$\hat{m}_I$	$\hat{m}_{I_{12}}$	$\hat{m}_{I_{21}}$	h_d	$\hat{m}_I$	$\hat{m}_{I_{12}}$	$\hat{m}_{I_{21}}$	h_d
0.0331355	0.034175	0.032096	0.009	0.033113	0.03419	0.032036	0.02
0.033126	0.034169	0.032083	0.01	0.0331165	0.034196	0.032037	0.021
0.033119	0.034167	0.032071	0.011	0.03312	0.034203	0.032037	0.022
0.033114	0.034166	0.032062	0.012	0.0331245	0.03421	0.032039	0.023
0.0331105	0.034165	0.032056	0.013	0.0331315	0.034221	0.032042	0.024
0.0331085	0.034166	0.032051	0.014	0.03314	0.034233	0.032047	0.025
0.0331065	0.034167	0.032046	0.015	0.033149	0.034246	0.032052	0.026
0.0331055	0.034169	0.032042	0.016	0.033159	0.03426	0.032058	0.027
0.0331055	0.034172	0.032039	0.017	0.03317	0.034275	0.032065	0.028
0.033107	0.034177	0.032037	0.018	0.0331815	0.03429	0.032073	0.029
0.0331095	0.034183	0.032036	0.019	0.033194	0.034307	0.032081	0.03

Table 1: *Scores of the cross validation function for the different estimates with $h_r = 0.169$*

The estimates for the regression function are displayed in Figure 6. Since the difference between $\hat{m}_{I_{12}}$ and $\hat{m}_{I_{21}}$ was not significant, only the estimate $\hat{m}_I$ is shown. The monotonization with respect to the education level appears to be more influential than the monotonization with respect to age.

5   Appendix: Proofs

For the sake of simplicity, we assume that the same bandwidths are used for both coordinates, i.e. $h_r = h_{r,1} = h_{r,2}$ in the regression estimate (2.8), and that the kernel K_r in this estimate is in fact a product kernel, i.e. $K_r(x)K_r(y)$, where we also use the symbol K_r for the one-dimensional kernel. For a more transparent notation, we will further assume that $N = n$. Throughout this section, we use the same notations as in the previous sections. The index "1" in m_1^{-1} means inversion of the function m with respect to its first argument, where the second argument has been fixed. Whereas the indexes in $\hat{m}_{I_{1,2}}$ refer not only to the variables which has been monotonized but also to the order of isotonization.

5.1   Proof of Lemma 2.2.

Recall the definition of the estimate $\hat{m}_{I_1}^{-1}$ in (2.10), then (for sufficiently large $N = n$) $\hat{m}_{I_1}^{-1}$ is obviously strictly increasing with respect to its first argument (for fixed second argument) because the kernel K_d is positive. For fixed x and $y_1 < y_2$, we therefore obtain

$$\hat{m}_{I_1}^{-1}(x \mid y_2) - \hat{m}_{I_1}^{-1}(x \mid y_1) = \frac{1}{N}\sum_{i=1}^{N}\left\{\bar{K}_d\left(\frac{\hat{m}(\frac{i}{N}, y_1) - x}{h_d}\right) - \bar{K}_d\left(\frac{\hat{m}(\frac{i}{N}, y_2) - x}{h_d}\right)\right\},$$
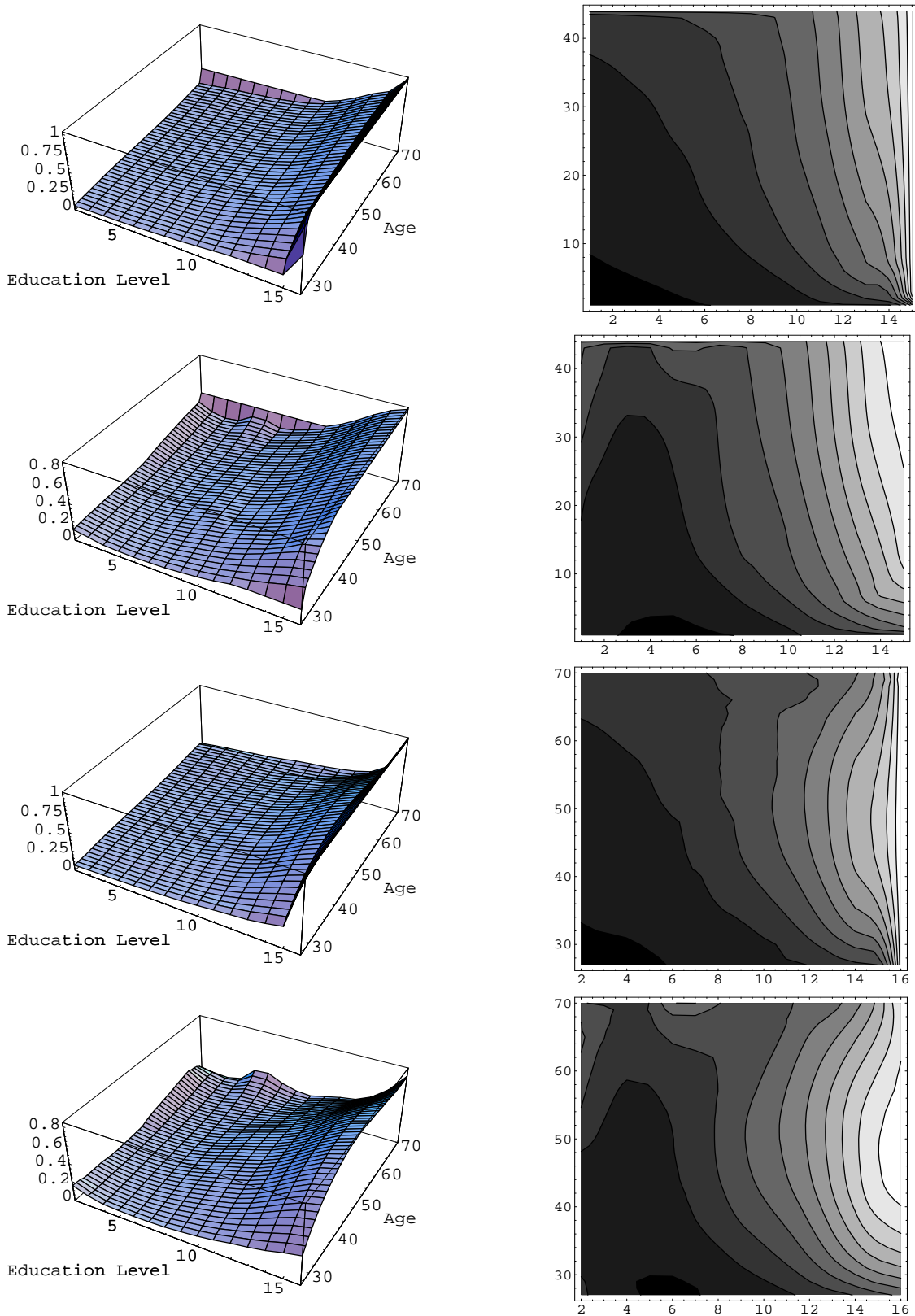
Figure 6: The results of the Census-Income Example. The estimates for the regression function and their corresponding Contour Plots from top to bottom with $h_r = 0.169$ and $h_d = 0.017$: $\hat{m}_I$, $\hat{m}_{I_1}$, $\hat{m}_{I_2}$, and $\hat{m}$

where $\bar{K}_d(w) = \int_{-1}^{w} K_d(v) dv$ is a strictly increasing function for $w \in [-1, 1]$. Now, by the assumption of the lemma, $\hat{m}$ is strictly increasing with respect to its second argument, which implies that the above difference is negative for at least one i if N is sufficiently large. Consequently, for fixed x, the estimate $\hat{m}_{I_1}^{-1}$ is strictly decreasing with respect to its second argument.

$\square$

5.2 Proof of Lemma 2.3.

From the basic construction, it is obvious that the estimate $\hat{m}_{I_{1,2}}$ is strictly increasing with respect to its second argument for sufficiently large N, and it remains to show that it has the same property with respect to its first argument. By Lemma 2.2, it follows that $\hat{m}_{I_1}$ is strictly increasing with respect to its first argument. Therefore interchanging the role of the first and second coordinate and applying the second part of Lemma 2.2, it follows that $\hat{m}_{I_{1,2}}^{-1}$ is strictly decreasing in x for fixed y and sufficiently large N, that is

(5.1)
$$\hat{m}_{I_{1,2}}^{-1}(y \mid x_1) > \hat{m}_{I_{1,2}}^{-1}(y \mid x_2)$$

if $x_1 < x_2$. Note further that the function $\hat{m}_{I_{1,2}}$ is obtained by the inversion of $\hat{m}_{I_{1,2}}^{-1}$ as a function of y for fixed x. In other words,
$$t = \hat{m}_{I_{1,2}}(x, y)$$

is the (unique) solution of the equation

$$\hat{m}_{I_{1,2}}^{-1}(t \mid x) = y$$

with respect to t. Now consider the equation (with respect to t_1, t_2)

$$y = \hat{m}_{I_{1,2}}^{-1}(t_1 \mid x_1) = \hat{m}_{I_{1,2}}^{-1}(t_2 \mid x_2)$$

with $x_1 < x_2$. We have to show that $t_1 = \hat{m}_{I_{1,2}}(x_1, y) < t_2 = m_{I_{1,2}}(x_2, y)$ and assume that the converse $t_2 \leq t_1$ is true. In this case, we would obtain from the monotonicity of $\hat{m}_{I_{1,2}}^{-1}(t \mid x)$ with respect to the variable t and from (5.1) that

$$\hat{m}_{I_{1,2}}^{-1}(t_1 \mid x_1) = m_{I_{1,2}}^{-1}(t_2 \mid x_2) \leq m_{I_{1,2}}^{-1}(t_1 \mid x_2) < m_{I_{1,2}}^{-1}(t_1 \mid x_1),$$

which yields a contradiction to the assumption

$$\hat{m}_{I_{1,2}}(x_1, y) = t_1 \geq t_2 = \hat{m}_{I_{1,2}}(x_2, y).$$

$\square$

5.3 Proof of Theorem 2.4.

It has been proven by Birke and Dette (2005) that the inverse of the function g_I^{-1} defined in (2.1) satisfies

(5.2)
$$\int_0^1 |g_I(u)|^p du = \int_0^1 |g(u)|^p du$$

for all $p > 0$. We now use Fubini's theorem and obtain by two applications of (5.2)

$$\int_{[0,1]^2} |\tilde{m}_{I_{1,2}}(x,y)|^p dxdy = \int_0^1 \int_0^1 |\hat{m}_{I_1}(x,u)|^p dudx$$

$$= \int_0^1 \int_0^1 |\hat{m}_{I_1}(x,u)|^p dxdu = \int_{[0,1]^2} |\hat{m}(x,y)|^p dxdy.$$

$\square$

5.4 Proof of Theorem 3.1.

The proof follows by adapting the arguments in Dette et al. (2005) to the two-dimensional case. Because these authors discuss the case where the bandwidths h_r and h_d are of the same order, i.e. $\lim_{n\to\infty} h_r/h_d = c > 0$, and in order to make the present paper self-contained, we present some details, which are also useful to see why the bandwidth conditions in (3.1) - (3.4) are required. Moreover, the arguments given in this subsection will help the reader to understand the proof of Theorem 3.3. We use Lemma 2.1, $h_d = o(h_r)$, and the decomposition

$$(5.3) \quad \Delta_n(t|y) = (\hat{m}_{I_1}^{-1} - m_1^{-1})(t|y)$$

$$= (\hat{m}_{I_1}^{-1} - m_{n_1}^{-1})(t|y) \ + \ o\Big(\frac{1}{\sqrt{nh_r^2}}\Big)$$

$$= \frac{1}{nh_d} \sum_{i=1}^n \int_{-\infty}^t \Big(K_d\Big(\frac{\hat{m}(\frac{i}{n},y)-u}{h_d}\Big) - K_d\Big(\frac{m(\frac{i}{n},y)-u}{h_d}\Big)\Big) du \ + \ o\Big(\frac{1}{\sqrt{nh_r^2}}\Big)$$

$$= \Delta_n^{(1)}(t|y) + \frac{1}{2}\Delta_n^{(2)}(t|y) \ + \ o\Big(\frac{1}{\sqrt{nh_r^2}}\Big),$$

where the function $m_{n_1}^{-1}(t|y)$ is defined in (2.12) (note that we assume $N = n$ throughout this proof and that the index 1 corresponds to the first coordinate),

$$(5.4) \qquad \Delta_n^{(1)}(t|y) = \frac{1}{nh_d^2} \sum_{i=1}^n \int_{-\infty}^t K_d'\Big(\frac{m(\frac{i}{n},y)-u}{h_d}\Big)\Big(\hat{m}(\frac{i}{n},y) - m(\frac{i}{n},y)\Big) du,$$

$$(5.5) \qquad \Delta_n^{(2)}(t|y) = \frac{1}{nh_d^3} \sum_{i=1}^n \int_{-\infty}^t K_d''\Big(\frac{\xi_i(y,u)-u}{h_d}\Big)\Big(\hat{m}(\frac{i}{n},y) - m(\frac{i}{n},y)\Big)^2 du,$$

and the points $\xi_i(y,u)$ satisfy for any $u \in (-\infty, t]$ the inequalities

$$|\xi_i(y,u) - m(\frac{i}{n},y)| \le |\hat{m}(\frac{i}{n},y) - m(\frac{i}{n},y)| \quad i = 1,\ldots,n.$$

Now a standard argument using the asymptotic MSE of the estimate $\hat{m}(x,y)$, the bandwidth condition (3.4), and a Taylor expansion shows that

$$(5.6) \qquad\qquad\qquad \sqrt{nh_r^2}\Delta_n^{(2)}(t|y) = o_p(1).$$

21

The remaining term $\Delta_n^{(1)}$ is decomposed as follows

$$(5.7) \qquad \Delta_n^{(1)}(t|y) = (\Delta_n^{(1.1)}(t|y) + \Delta_n^{(1.2)}(t|y))(1 + o_p(1)),$$

with

$$(5.8) \quad \Delta_n^{(1.1)}(t|y) = \frac{-1}{n^2 h_d h_r^2} \sum_{i,j=1}^{n} K_d\Big(\frac{m(\frac{i}{n}, y) - t}{h_d}\Big) K_r\Big(\frac{X_j - \frac{i}{n}}{h_r}\Big) K_r\Big(\frac{Y_j - y}{h_r}\Big) \frac{m(X_j, Y_j) - m(\frac{i}{n}, y)}{f(\frac{i}{n}, y)}$$

$$(5.9) \quad \Delta_n^{(1.2)}(t|y) = \frac{-1}{n^2 h_d h_r^2} \sum_{i,j=1}^{n} K_d\Big(\frac{m(\frac{i}{n}, y) - t}{h_d}\Big) K_r\Big(\frac{X_j - \frac{i}{n}}{h_r}\Big) K_r\Big(\frac{Y_j - y}{h_r}\Big) \sigma(X_j, Y_j) \frac{\varepsilon_j}{f(\frac{i}{n}, y)}$$

corresponding to bias and variance in Theorem 3.1. A straightforward calculation yields for the first term

$$(5.10) \quad E[\Delta_n^{(1.1)}(t|y)] = -\frac{(1 + o(1))}{h_r^2 h_d} \int K_d\Big(\frac{m(x, y) - t}{h_d}\Big) K_r\Big(\frac{u - x}{h_r}\Big) K_r\Big(\frac{v - y}{h_r}\Big)$$

$$\times \frac{(m(u, v) - m(x, y))}{f(x, y)} f(u, v) \, dx \, du \, dv$$

$$= -\frac{(1 + o(1)) h_r^2 \kappa_2(K_r)}{\frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y)} \Big(\frac{\partial^2}{\partial x_1^2} m(m_1^{-1}(t|y), y) + \frac{\partial^2}{\partial y_1^2} m(m_1^{-1}(t|y), y)$$

$$+ 2\frac{\frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y) \frac{\partial}{\partial x_1} f(m_1^{-1}(t|y), y)}{f(m_1^{-1}(t|y), y)}$$

$$+ 2\frac{\frac{\partial}{\partial y_1} m(m_1^{-1}(t|y), y) \frac{\partial}{\partial y_1} f(m_1^{-1}(t|y), y)}{f(m_1^{-1}(t|y), y)}\Big),$$

which gives the bias in Theorem 3.1 (note that we assume $h_r = h_{r,1} = h_{r,2}$ throughout this section). For the estimation of the variance of $\Delta_n^{(1.1)}(t|y)$, we use the substitution $x = m_1^{-1}(t + h_d u|y)$ and obtain

$$\text{Var}(\Delta_n^{(1.1)}(t|y))$$

$$\leq \frac{1}{n^3 h_r^4 h_d^2} E\Big[\Big(\sum_{i=1}^{n} K_d\Big(\frac{m(\frac{i}{n}, y) - t}{h_d}\Big) K_r\Big(\frac{X_j - \frac{i}{n}}{h_r}\Big) K_r\Big(\frac{Y_j - y}{h_r}\Big) \frac{m(X_j, Y_j) - m(\frac{i}{n}, y)}{f(\frac{i}{n}, y)}\Big)^2\Big]$$

$$= \frac{(1 + o(1))}{n h_r^4 h_d^2} \int \Big(\int_0^1 K_d\Big(\frac{m(x, y) - t}{h_d}\Big) K_r\Big(\frac{v - x}{h_r}\Big) K_r\Big(\frac{w - y}{h_r}\Big) \frac{m(v, w) - m(x, y)}{f(x, y)} dx\Big)^2 f(v, w) \, dv \, dw$$

$$= \frac{(1 + o(1))}{n h_r^4} \int K_r^2\Big(\frac{v - m_1^{-1}(t|y)}{h_r}\Big) K_r^2\Big(\frac{w - y}{h_r}\Big) \frac{(m(v, w) - m(m_1^{-1}(t|y), y))^2}{(f(m_1^{-1}(t|y), y) \frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y))^2} f(v, w) \, dv \, dw.$$

With the substitutions $v = m_1^{-1}(t|y) + h_r v'$ and $w = y + h_r w'$ it therefore follows

$$(5.11) \, \text{Var}(\Delta_n^{(1.1)}(t|y)) \leq \frac{(1 + o(1))}{n h_r^2} \frac{1}{f(m_1^{-1}(t|y), y)(\frac{\partial}{\partial x_1} m(m_1^{-1}(t|y), y))^2}$$

$$\times \int K_r^2(v') K_r^2(w') (m(m_1^{-1}(t|y) + h_r v', y + h_r w') - m(m_1^{-1}(t|y), y))^2 \, dv' \, dw'$$

$$= O\Big(\frac{1}{n}\Big) = o\Big(\frac{1}{n h_r^2}\Big).$$

In order to derive the last identities, we applied a first order Taylor expansion. We note at this point that in the case of a d-dimensional predictor the rate of convergence in (5.11) is of order

$$O\Big(\frac{1}{nh_r^{d-2}}\Big)$$

[see also Dette et al. (2005) for the case $d = 1$]. For the treatment of the second term $\Delta_n^{(1.2)}(t|y)$ in (5.7), we note that $E[\Delta_n^{(1.2)}(t|y)] = 0$ and obtain (using the i.i.d. assumption)

$$
\begin{aligned}
\mathrm{Var}(\sqrt{nh_r^2}\Delta_n^{(1.2)}(t|y)) &= \frac{(1+o(1))}{h_d^2 h_r^2} \int \sigma^2(u,v) K_d\Big(\frac{m(x,y)-t}{h_d}\Big) K_d\Big(\frac{m(z,y)-t}{h_d}\Big) \\
&\times K_r^2\Big(\frac{v-y}{h_r}\Big) K_r\Big(\frac{u-x}{h_r}\Big) K_r\Big(\frac{u-z}{h_r}\Big) \frac{f(u,v)}{f(x,y)f(z,y)} dx\,dz\,du\,dv \\
&= \frac{(1+o(1))}{h_r^2} \int K_r^2\Big(\frac{v-y}{h_r}\Big) K_d(x') K_d(z') K_r\Big(\frac{u-m_1^{-1}(t+h_d x'|y)}{h_r}\Big) \\
&\times K_r\Big(\frac{u-m_1^{-1}(t+h_d z'|y)}{h_r}\Big) \frac{\sigma^2(u,v)f(u,v)}{(f(m_1^{-1}(t|y),y)\frac{\partial}{\partial x_1}m(m_1^{-1}(t|y),y))^2} dx'\,dz'\,du\,dv,
\end{aligned}
$$

where we used the substitutions $x = m_1^{-1}(t+h_d x'|y)$ and $z = m_1^{-1}(t+h_d z'|y)$ in the last equation. A further substitution $v = y + h_r v'$ and $u = m_1^{-1}(t+h_d x'|y) + h_r u'$ now gives

$$
\begin{aligned}
(5.12)\quad \mathrm{Var}(\sqrt{nh_r^2}\Delta_n^{(1.2)}(t|y)) &= \frac{(1+o(1))\sigma^2(m_1^{-1}(t|y),y)}{f(m_1^{-1}(t|y),y)(\frac{\partial}{\partial x_1}m(m_1^{-1}(t|y),y))^2} \int K_r^2(v') K_d(x') K_d(z') K_r(u') \\
&\times K_r\Big(\frac{m_1^{-1}(t+h_d x'|y)-m_1^{-1}(t+h_d z'|y)}{h_r} + u'\Big) dx'\,dy'\,du'\,dv' \\
&= \frac{(1+o(1))\sigma^2(m_1^{-1}(t|y),y)}{f(m_1^{-1}(t|y),y)(\frac{\partial}{\partial x_1}m(m_1^{-1}(t|y),y))^2} \Big(\int K_r^2(u')du'\Big)^2,
\end{aligned}
$$

where we have used the condition $\lim_{n\to\infty}\frac{h_d}{h_r} = 0$ in the last step and the simplifying assumption of a product kernel made at the beginning of this section. Finally, the assertion of the Theorem follows from (5.3), (5.6), (5.7), (5.10), (5.11), (5.12), and the weak convergence

$$(5.13)\qquad\qquad \sqrt{nh_r^2}\Delta_n^{(1.2)}(t|y) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(t,y)),$$

which is an immediate consequence of Ljapunoff's Theorem. □

5.5 Sketch of the proof of Theorem 3.2.

The basic idea to prove Theorem 3.2 uses a special Taylor expansion for the operator, which maps a strictly increasing function g with t in the interior of its image to its quantile $g^{-1}(t)$. Details can be found in Dette et al. (2005). By this expansion, we obtain the following decomposition

$$(5.14)\qquad\qquad \hat{m}_{I_1}(t,y) - m(t,y) = A_n(y) + \frac{1}{2}B_n(y),$$

23

where

$$A_n(y) = -\frac{(\hat{m}_{I_1}^{-1} - m_1^{-1})}{\frac{\partial}{\partial x_1}(m_1^{-1})} \circ (m(t,y)|y)$$

$$= -\frac{\partial}{\partial x_1} m(t,y)(\hat{m}_{I_1}^{-1} - m_1^{-1}) \circ (m(t,y)|y)$$

and

$$B_n(y) = 2\frac{(\hat{m}_{I_1}^{-1} - m_1^{-1})\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_1^{-1})}{[\frac{\partial}{\partial x_1}(m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))]^2} \circ (m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))_1^{-1}(t|y)$$

$$-\frac{(\hat{m}_{I_1}^{-1} - m_1^{-1})^2\frac{\partial^2}{\partial^2 x_1}(m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))}{[\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))]^3} \circ (m^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))_1^{-1}(t|y)$$

$$= 2B_{n1}(y) - B_{n2}(y).$$

If it can be shown that

(5.15)
$$B_n(y) = o_p\Big(\frac{1}{\sqrt{nh_r^2}}\Big),$$

then the assertion of the Theorem follows directly from Theorem 3.1.
For a proof of (5.15) we first note that

$$(m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1})) \xrightarrow{\mathcal{P}} m_1^{-1}$$

and

$$(m^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))_1^{-1}(t,y) \xrightarrow{\mathcal{P}} m(t,y).$$

To deal with the terms in the numerator of $B_{n1}(y)$, we analyse $\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_1^{-1})(\zeta|y)$ in a somehow similar way as $(\hat{m}_{I_1}^{-1} - m_1^{-1})(t|y)$ in the proof of Theorem 3.1. We use the inequality

$$|\hat{m}(\frac{i}{n}, y) - m(\frac{i}{n}, y)| = O\Big(\frac{-\log h_r}{nh_r^2}\Big)^{1/2} \quad a.s.$$

[see Collomb and Härdle (1986), Theorem 2], and obtain for some $\eta_{i,n}(y)$ with

$$|\eta_{i,n}(y) - m(\frac{i}{n}, y)| < |\hat{m}(\frac{i}{n}, y) - m(\frac{i}{n}, y)|$$

the estimate

$$\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_1^{-1})(\zeta|y) = \frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_{n_1}^{-1})(\zeta|y) + \frac{\partial}{\partial x_1}(\hat{m}_{n_1}^{-1} - m_1^{-1})(\zeta|y)$$

$$= \frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_{n_1}^{-1})(\zeta|y) + O\Big(h_d + \frac{1}{nh_d}\Big)$$

$$= \frac{1}{nh_d^2}\sum_{i=1}^n\Big[K_d\Big(\frac{\hat{m}(\frac{i}{n}, y) - t}{h_d}\Big) - K_d\Big(\frac{m(\frac{i}{n}, y) - t}{h_d}\Big)\Big] + O\Big(h_d + \frac{1}{nh_d}\Big)$$

24

$$= \frac{1}{nh_d^2} \sum_{i=1}^{n} K_d' \Big(\frac{\eta_{i,n}(y) - \zeta}{h_d} \Big) \Big(\hat{m}(\frac{i}{n}, y) - m(\frac{i}{n}, y) \Big) + O\Big(h_d + \frac{1}{nh_d} \Big)$$

$$= \frac{1 + o(1)}{nh_d^2} \int_0^1 K_d' \Big(\frac{m(x, y) - \zeta}{h_d} \Big) dx \, O_p \Big(\frac{-\log h_r}{nh_r^2} \Big)^{1/2} + O\Big(h_d + \frac{1}{nh_d} \Big)$$

$$= O_p \Big(\Big(\frac{-\log h_r}{nh_r^2 h_d^2} \Big)^{1/2} + h_d + \frac{1}{nh_d} \Big).$$

From the bandwidth condition (3.5), it follows that

(5.16)
$$\frac{\partial}{\partial x_1} (\hat{m}_{I_1}^{-1} - m_1^{-1})(\zeta | y) = o_p(1).$$

We remark that in the d-dimensional case, the bandwidth condition (3.5) has to be modified into

$$\frac{-\log(h_r)}{nh_r^d h_d^2} = o(1)$$

since the rate of the strong uniform consistency of $\hat{m}$ changes [see Collomb and Härdle (1986)], whereas the above calculation stays essentially the same (i.e. the exponent of h_d does not change). This and the following estimate

$$(\hat{m}_{I_1}^{-1} - m_1^{-1}) \circ (m(t, y) | y) = O_p \Big(\frac{1}{\sqrt{nh_r^2}} \Big)$$

derived from Theorem 3.1 establish finally that

$$B_{n_1}(y) = O_p \Big(\frac{1}{\sqrt{nh_r^2}} \Big(\Big(\frac{-\log h_r}{nh_r^2 h_d^2} \Big)^{1/2} + h_d + \frac{1}{nh_d} \Big) \Big) = o_p \Big(\frac{1}{\sqrt{nh_r^2}} \Big)$$

and

$$B_{n_2}(y) = O_p \Big(\frac{1}{nh_r^2} \Big) = o_p \Big(\frac{1}{\sqrt{nh_r^2}} \Big).$$

The last equation is true, since the second derivative $\frac{\partial^2}{\partial x_1^2} m(x, y)$ is bounded. This proves the estimate (5.15) and completes the proof of Theorem 3.2. □

5.6 Proof of Theorem 3.3.

The proof uses Theorem 3.2, Lemma 2.1, $h_d = o(h_r)$ and the decomposition

(5.17) $\Delta_n(t|x) = \hat{m}_{I_{1,2}}^{-1}(t|x) - m_2^{-1}(t|x) = \hat{m}_{I_{1,2}}^{-1}(t|x) - m_{n_2}^{-1}(t|x) + o(\frac{1}{\sqrt{nh_r^2}})$

$$= \frac{1}{nh_d} \sum_{i=1}^{n} \int_{-\infty}^{t} \Big[K_d \Big(\frac{\hat{m}_{I_1}(x, \frac{i}{n}) - u}{h_d} \Big) - K_d \Big(\frac{m(x, \frac{i}{n}) - u}{h_d} \Big) \Big] du + o(\frac{1}{\sqrt{nh_r^2}})$$

$$= \Delta_n^{(1)}(t|x) + \frac{1}{2} \Delta_n^{(2)}(t|x) + o(\frac{1}{\sqrt{nh_r^2}}),$$

25

where the quantitiy $m_{n_2}^{-1}(t|x)$ is defined by

$$(5.18) \qquad m_{n_2}^{-1}(t|x) = \frac{1}{nh_d} \sum_{i=1}^{n} \int_{-\infty}^{t} K_d\Big(\frac{m(x, \frac{i}{n}) - u}{h_d}\Big) du \ ,$$

and the terms $\Delta_n^{(1)}(t|x)$ and $\Delta_n^{(2)}(t|x)$ in the last line of (5.17) are given by

$$(5.19) \qquad \Delta_n^{(1)}(t|x) = -\frac{1}{nh_d} \sum_{i=1}^{n} K_d\Big(\frac{m(x, \frac{i}{n}) - t}{h_d}\Big) \Big(\hat{m}_{I_1}(x, \tfrac{i}{n}) - m(x, \tfrac{i}{n})\Big)$$

$$(5.20) \qquad \Delta_n^{(2)}(t|x) = \frac{1}{nh_d^3} \sum_{i=1}^{n} \int_{-\infty}^{t} K_d''\Big(\frac{\xi_i(x, u) - u}{h_d}\Big) \Big(\hat{m}_{I_1}(x, \tfrac{i}{n}) - m(x, \tfrac{i}{n})\Big)^2 du.$$

The points $\xi_i(x, u)$ with $i = 1, \ldots, n$ satisfy the inequality

$$|\xi_i(x, u) - m(x, \tfrac{i}{n})| \le |\hat{m}_{I_1}(x, \tfrac{i}{n}) - m(x, \tfrac{i}{n})|$$

for all $u \in (0, 1)$. We use the fact that the estimate $\hat{m}_{I_1}(x, y)$ has the same asymptotic properties as $\hat{m}(x, y)$ established in Theorem 3.2. This yields

$$(\hat{m}_{I_1}(x, y) - m(x, y))^2 = O_p\Big(h_r^4 + \frac{1}{nh_r^2}\Big),$$

and it follows

$$(5.21) \qquad |\Delta_n^{(2)}(t|x)| = \Big|\frac{(1 + o_p(1))}{nh_d^2} \sum_{i=1}^{n} K_d'\Big(\frac{m(x, \frac{i}{n}) - t}{h_d}\Big) \Big(\hat{m}_{I_1}(x, \tfrac{i}{n}) - m(x, \tfrac{i}{n})\Big)^2\Big|$$

$$= \frac{(1 + o_p(1))}{h_d^2} \Big|\int_0^1 K_d'\Big(\frac{m(x, y) - t}{h_d}\Big) (\hat{m}_{I_1}(x, y) - m(x, y))^2 \, dy\Big|$$

$$= O_p\Big(\frac{1}{h_d}\Big(h_r^4 + \frac{1}{nh_r^2}\Big)\Big) = o_p\Big(\frac{1}{\sqrt{nh_r^2}}\Big).$$

For the first term in the decomposition (5.17), we obtain

$$(5.22) \qquad \Delta_n^{(1)}(t|x) = \Delta_n^{(1.1)}(t|x) + \Delta_n^{(1.2)}(t|x),$$

with

$$(5.23) \qquad \Delta_n^{(1.1)}(t|x) = -\frac{1}{nh_d} \sum_{i=1}^{n} K_d\Big(\frac{m(x, \frac{i}{n}) - t}{h_d}\Big) \Big(\hat{m}_{I_1}(x, \tfrac{i}{n}) - m_{n_1}(x, \tfrac{i}{n})\Big),$$

$$(5.24) \qquad \Delta_n^{(1.2)}(t|x) = -\frac{1}{nh_d} \sum_{i=1}^{n} K_d\Big(\frac{m(x, \frac{i}{n}) - t}{h_d}\Big) \Big(m_{n_1}(x, \tfrac{i}{n}) - m(x, \tfrac{i}{n})\Big)$$

$$= O\Big(h_d^2 + \frac{1}{nh_d}\Big) = o\Big(\frac{1}{\sqrt{nh_r^2}}\Big).$$

26

Note that m_{n_1} is the inverse of the function $m_{n_1}^{-1}$ defined in (2.12) (calculated with respect to the first argument), and the last estimate is obvious by Lemma 2.1 and the bandwidth conditions stated in Section 3. The first term is treated by a Taylor expansion similar to the expansion (5.14) in the sketch of the proof of Theorem 3.2, which yields

$$(5.25) \qquad \hat{m}_{I_1}(x,y) - m_{n_1}(x,y) = -\frac{(\hat{m}_{I_1}^{-1} - m_{n_1}^{-1})}{\frac{\partial}{\partial x_1}(m_1^{-1})}(m(x,y)|y) + o_p\Big(\frac{1}{\sqrt{nh_r^2}}\Big),$$

where $\frac{\partial}{\partial x_1}(m_1^{-1})$ is the derivative of the function $m_1^{-1}(\cdot|y)$ with respect to the x-coordinate for fixed y. Therefore we obtain from (5.19)

$$(5.26) \qquad \Delta_n^{(1)}(t|x) = \Delta_n^{(1.1.1)}(t|x)(1 + o_P(1)),$$

where the random variable $\Delta_n^{(1.1.1)}(t|x)$ is defined by

$$(5.27) \quad \Delta_n^{(1.1.1)}(t|x) = \frac{1}{nh_d}\sum_{i=1}^{n} K_d\Big(\frac{m(x,\frac{i}{n}) - t}{h_d}\Big)\frac{\partial}{\partial x_1}m(x,\frac{i}{n})\Big((\hat{m}_{I_1}^{-1} - m_{n_1}^{-1})(m(x,\frac{i}{n})|\frac{i}{n})\Big)$$

$$= \frac{1}{n^2 h_d^2}\sum_{i,j=1}^{n} K_d\Big(\frac{m(x,\frac{i}{n}) - t}{h_d}\Big)\frac{\partial}{\partial x_1}m(x,\frac{i}{n})$$

$$\times \int_{-\infty}^{m(x,\frac{i}{n})}\Big(K_d\Big(\frac{\hat{m}(\frac{j}{n},\frac{i}{n}) - u}{h_d}\Big) - K_d\Big(\frac{m(\frac{j}{n},\frac{i}{n}) - u}{h_d}\Big)\Big)du.$$

We now apply a similar (but substantially more intricate decomposition) as used in the proof of Theorem 3.1, that is

$$(5.28) \qquad \Delta_n^{(1.1.1)}(t|x) = \Delta_n^{(1.1.1.a)}(t|x) + \frac{1}{2}\Delta_n^{(1.1.1.b)}(t|x),$$

with

$$(5.29) \qquad \Delta_n^{(1.1.1.a)}(t|x) = \frac{1}{n^2 h_d^3}\sum_{i,j=1}^{n}\frac{\partial}{\partial x_1}m(x,\frac{i}{n})K_d\Big(\frac{m(x,\frac{i}{n}) - t}{h_d}\Big)$$

$$\times \int_{-\infty}^{m(x,\frac{i}{n})} K_d'\Big(\frac{m(\frac{j}{n},\frac{i}{n}) - u}{h_d}\Big)\Big(\hat{m}(\frac{j}{n},\frac{i}{n}) - m(\frac{j}{n},\frac{i}{n})\Big)du$$

$$= -\frac{1}{n^2 h_d^2}\sum_{i,j=1}^{n}\frac{\partial}{\partial x_1}m(x,\frac{i}{n})K_d\Big(\frac{m(x,\frac{i}{n}) - t}{h_d}\Big)$$

$$\times K_d\Big(\frac{m(\frac{j}{n},\frac{i}{n}) - m(x,\frac{i}{n})}{h_d}\Big)\Big(\hat{m}(\frac{j}{n},\frac{i}{n}) - m(\frac{j}{n},\frac{i}{n})\Big)$$

$$(5.30) \qquad \Delta_n^{(1.1.1.b)}(t|x) = \frac{1}{n^2 h_d^4}\sum_{i,j=1}^{n}\frac{\partial}{\partial x_1}m(x,\frac{i}{n})K_d\Big(\frac{m(x,\frac{i}{n}) - t}{h_d}\Big)$$

$$\times \int_{-\infty}^{m(x,\frac{i}{n})} K_d''\Big(\frac{\xi_j(\frac{i}{n},u) - u}{h_d}\Big)\Big(\hat{m}(\frac{j}{n},\frac{i}{n}) - m(\frac{j}{n},\frac{i}{n})\Big)^2 du,$$

27

and $|\xi_j(\frac{i}{n}, u) - m(\frac{j}{n}, \frac{i}{n})| \leq |\hat{m}(\frac{j}{n}, \frac{i}{n}) - m(\frac{j}{n}, \frac{i}{n})|$. Note that the terms $\Delta_n^{(1.1.1.a)}(t|x)$ and $\Delta_n^{(1.1.1.b)}(t|x)$ are the analogs of the quantities $\Delta_n^{(1)}$ and $\Delta_n^{(2)}$ in the proof of Theorem 3.1 (which was one of the reasons why we gave some more details of this proof). A straightforward argument using the asymptotic properties of the Nadaraya-Watson estimate $\hat{m}(x, y)$ shows

$$(5.31) \quad |\Delta_n^{(1.1.1.b)}(t|x)| = \frac{(1 + o_p(1))}{h_d^3} \Big| \int_0^1 \int_0^1 \frac{\partial}{\partial x_1} m(x, y) K_d\Big(\frac{m(x, y) - t}{h_d}\Big)$$

$$\times K_d'\Big(\frac{m(u, y) - m(x, y)}{h_d}\Big) (\hat{m}(u, y) - m(u, y))^2 \, du dy \Big|$$

$$= O\Big(\frac{1}{h_d}\Big(h_r^4 + \frac{1}{nh_r^2}\Big)\Big) = o_p\Big(\frac{1}{\sqrt{nh_r^2}}\Big),$$

where we used the assumption (3.4) in the last estimate. For the term in (5.29) we derive a further decomposition, that is

$$(5.32) \quad \Delta_n^{(1.1.1.a)}(t|x) = -\Big[\Delta_n^{(1.1.1.a.a)}(t|x) + \Delta_n^{(1.1.1.a.b)}(t|x)\Big](1 + o_p(1)),$$

where the terms $\Delta^{(1.1.1.a.a)}$ and $\Delta^{(1.1.1.a.b)}$ are defined by

$$(5.33) \quad \Delta_n^{(1.1.1.a.a)}(t|x) = \frac{1}{n^3 h_d^2 h_r^2} \sum_{i,j,k=1}^n \frac{\partial}{\partial x_1} m\Big(x, \frac{i}{n}\Big) K_d\Big(\frac{m(x, \frac{i}{n}) - t}{h_d}\Big) K_d\Big(\frac{m(\frac{j}{n}, \frac{i}{n}) - m(x, \frac{i}{n})}{h_d}\Big)$$

$$\times K_r\Big(\frac{X_k - \frac{j}{n}}{h_r}\Big) K_r\Big(\frac{Y_k - \frac{i}{n}}{h_r}\Big) \frac{m(X_k, Y_k) - m(\frac{j}{n}, \frac{i}{n})}{f(\frac{j}{n}, \frac{i}{n})},$$

$$(5.34) \quad \Delta_n^{(1.1.1.a.b)}(t|x) = \frac{1}{n^3 h_d^2 h_r^2} \sum_{i,j=1}^n \frac{\partial}{\partial x_1} m\Big(x, \frac{i}{n}\Big) K_d\Big(\frac{m(x, \frac{i}{n}) - t}{h_d}\Big) K_d\Big(\frac{m(\frac{j}{n}, \frac{i}{n}) - m(x, \frac{i}{n})}{h_d}\Big)$$

$$\times \sum_{k=1}^n K_r\Big(\frac{X_k - \frac{j}{n}}{h_r}\Big) K_r\Big(\frac{Y_k - \frac{i}{n}}{h_r}\Big) \sigma(X_k, Y_k) \frac{\varepsilon_k}{f(\frac{j}{n}, \frac{i}{n})},$$

and correspond to the bias and variance of the estimate $\hat{m}_{I_{1,2}}^{-1}$. For the expectation of (5.33), it follows by a straightforward but tedious calculation

$$(5.35) \quad E[\Delta_n^{(1.1.1.a.a)}(t|x)] = \frac{(1 + o(1))}{h_d^2 h_r^2} \int \frac{\partial}{\partial x_1} m(x, y) K_d\Big(\frac{m(x, y) - t}{h_d}\Big) K_r\Big(\frac{u - x'}{h_r}\Big) K_r\Big(\frac{v - y}{h_r}\Big)$$

$$\times K_d\Big(\frac{m(x', y) - m(x, y)}{h_d}\Big) f(u, v) \frac{m(u, v) - m(x', y)}{f(x', y)} dx' dy du dv$$

$$= \frac{(1 + o(1))}{h_d^2} h_r^2 \kappa_2(K_r) \int K_d\Big(\frac{m(x, y) - t}{h_d}\Big) K_d\Big(\frac{m(x', y) - m(x, y)}{h_d}\Big)$$

$$\times \frac{\partial}{\partial x_1} m(x, y) \Big[\frac{\partial^2}{\partial x_1^2} m(x', y) + \frac{\partial^2}{\partial y_1^2} m(x', y) +$$

$$2 \frac{\frac{\partial}{\partial x_1} m(x', y) \frac{\partial}{\partial x_1} f(x', y)}{f(x', y)} + 2 \frac{\frac{\partial}{\partial y_1} m(x', y) \frac{\partial}{\partial y_1} f(x', y)}{f(x', y)}\Big] dx' dy$$

$$= \frac{(1+o(1))h_r^2\kappa_2(K_r)}{\frac{\partial}{\partial y_1}m(x,m_2^{-1}(t|x))}\left[\frac{\partial^2}{\partial x_1^2}m(x,m_2^{-1}(t|x)) + \frac{\partial^2}{\partial y_1^2}m(x,m_2^{-1}(t|x))\right.$$

$$+ 2\frac{\frac{\partial}{\partial x_1}m(x,m_2^{-1}(t|x))\frac{\partial}{\partial x_1}f(x,m_2^{-1}(t|x))}{f(x,m_2^{-1}(t|x))}$$

$$\left.+ 2\frac{\frac{\partial}{\partial y_1}m(x,m_2^{-1}(t|x))\frac{\partial}{\partial y_1}f(x,m_2^{-1}(t|x))}{f(x,m_2^{-1}(t|x))}\right],$$

where we used successively the substitutions $u' = \frac{u-x'}{h_r}$, $v' = \frac{v-y}{h_r}$, $x'' = \frac{m(x',y)-m(x,y)}{h_d}$ with $x' = m_1^{-1}(m(x,y)+h_dx''|y)$, $y' = \frac{m(x,y)-t}{h_d}$ with $y = m_2^{-1}(t+h_dy'|x)$ and several Taylor expansions. For the variance of $\Delta_n^{(1.1.1.a.a)}(t|x)$, we obtain by a similar calculation

$$\text{Var}(\Delta_n^{(1.1.1.a.a.)}(t|x)) \leq \frac{1}{n^5h_d^4h_r^4}E\left[\left(\sum_{i,j=1}^n \frac{\partial}{\partial x_1}m(x,\tfrac{i}{n})K_d\left(\frac{m(x,\tfrac{i}{n})-t}{h_d}\right)K_d\left(\frac{m(\tfrac{j}{n},\tfrac{i}{n})-m(x,\tfrac{i}{n})}{h_d}\right)\right.\right.$$

$$\left.\left.\times K_r\left(\frac{X_k-\tfrac{j}{n}}{h_r}\right)K_r\left(\frac{Y_k-\tfrac{i}{n}}{h_r}\right)\frac{m(X_k,Y_k)-m(\tfrac{j}{n},\tfrac{i}{n})}{f(\tfrac{j}{n},\tfrac{i}{n})}\right)^2\right]$$

$$= \frac{(1+o(1))}{nh_r^4h_d^2}\int\left(\int_0^1 K_r\left(\frac{u-x}{h_r}\right)K_r\left(\frac{v-y}{h_r}\right)K_d\left(\frac{m(x,y)-t}{h_d}\right)\right.$$

$$\left.\times K_d\left(x''\right)\frac{m(u,v)-m(x,y)}{f(x,y)}dx''dy\right)^2 f(u,v)dudv$$

$$= \frac{1}{nh_r^4}\int\left(K_r\left(\frac{u-x}{h_r}\right)K_r\left(\frac{v-m_2^{-1}(t|x)}{h_r}\right)\frac{m(u,v)-m(x,m_2^{-1}(t|x))}{f(x,m_2^{-1}(t|x))}\right)^2$$

$$f(u,v)dudv \times \frac{(1+o(1))}{(\frac{\partial}{\partial y_1}m(x,m_2^{-1}(t|x)))^2}$$

$$= \frac{(1+o(1))}{nh_r^2(\frac{\partial}{\partial y_1}m(x,m_2^{-1}(t|x)))^2 f(x,m_2^{-1}(t|x))}\int K_r^2(u')K_r^2(v')$$

$$\times (m(x+h_ru',m_2^{-1}(t|x)+h_rv')-m(x,m_2^{-1}(t|x)))^2du'dv'$$

$$= o(\frac{1}{nh_r^2}),$$

where the substitutions $x' = m_1^{-1}(m(x,y)+h_dx''|y)$, $y = m_2^{-1}(t+h_dy'|x)$ and $v' = \frac{v-m_2^{-1}(t|x)}{h_r}$ are used successively. Therefore it follows that

$$(5.36) \qquad \Delta_n^{(1.1.1.a.a)}(t|x) - E[\Delta^{(1.1.1.a.a)}(t|x)] = o_P(\frac{1}{\sqrt{nh_r^2}}).$$

For the variance of the remaining term in (5.34), we obtain

$$\text{Var}(\sqrt{nh_r^2}\Delta_n^{(1.1.1.a.b)}(t|x)) = \frac{(1+o(1))}{h_d^4h_r^2}\int\sigma^2(u,v)f(u,v)\left(\int\frac{\partial}{\partial x_1}m(x,y)K_d\left(\frac{m(x,y)-t}{h_d}\right)\right.$$

$$\left.\times K_d\left(\frac{m(x',y)-m(x,y)}{h_d}\right)K_r\left(\frac{u-x'}{h_r}\right)K_r\left(\frac{v-y}{h_r}\right)\frac{1}{f(x',y)}dx'dy\right)^2 dudv$$

29

$$= \frac{(1+o(1))}{h_d^2 h_r^2} \int \sigma^2(u,v) f(u,v) \left(\int \frac{\frac{\partial}{\partial x_1} m(x, m_2^{-1}(t+h_d y'|x))}{\frac{\partial}{\partial y_1} m(x, m_2^{-1}(t+h_d y'|x))} K_d(y') \right.$$

$$\times \ K_d \left(\frac{m(x', m_2^{-1}(t+h_d y'|x)) - m(x, m_2^{-1}(t+h_d y'|x))}{h_d} \right) K_r \left(\frac{u-x'}{h_r} \right)$$

$$\left. \times \ K_r \left(\frac{v - m_2^{-1}(t+h_d y'|x)}{h_r} \right) \frac{dx' dy'}{f(x', m_2^{-1}(t+h_d y'|x))} \right)^2 du dv,$$

where the substitution $y = m_2^{-1}(t+h_d y'|x)$ is used in the second step. Similarly, putting

$$x'' = \frac{m(x', m_2^{-1}(t+h_d y'|x)) - m(x, m_2^{-1}(t+h_d y'|x))}{h_d},$$

it follows that

$$\mathrm{Var}(\sqrt{n h_r^2} \Delta_n^{(1.1.1.a.b)}(t|x)) = \frac{(1+o(1))}{h_r^2} \int \sigma^2(u,v) f(u,v) \left(\int \frac{K_d(y') K_d(x'')}{\frac{\partial}{\partial y_1} m(x, m_2^{-1}(t+h_d y'|x))} \right.$$

$$\times \ K_r \left(\frac{u - m_1^{-1}(m(x, m_2^{-1}(t+h_d y'|x)) + h_d x''|m_2^{-1}(t+h_d y'|x))}{h_r} \right)$$

$$\left. \times \ K_r \left(\frac{v - m_2^{-1}(t+h_d y'|x)}{h_r} \right) \frac{dx'' dy'}{f(x, m_2^{-1}(t|x))} \right)^2 du dv$$

$$= \frac{(1+o(1))}{h_r^2 (f(x, m_2^{-1}(t|x)) \frac{\partial}{\partial y_1} m(x, m_2^{-1}(t|x)))^2} \int \sigma^2(u,v) f(u,v) K_d(y_1')$$

$$\times \ K_d(x_1'') K_r \left(\frac{u - m_1^{-1}(m(x, m_2^{-1}(t|x)) + h_d x_1''|m_2^{-1}(t|x))}{h_r} \right) K_d(y_2')$$

$$\times \ K_d(x_2'') K_r \left(\frac{u - m_1^{-1}(m(x, m_2^{-1}(t|x)) + h_d x_2''|m_2^{-1}(t|x))}{h_r} \right)$$

$$\times \ K_r \left(\frac{v - m_2^{-1}(t+h_d y_1'|x)}{h_r} \right) K_r \left(\frac{v - m_2^{-1}(t+h_d y_2'|x)}{h_r} \right) dx_1'' dx_2'' dy_1' dy_2' du dv$$

Finally, a further substitution

$$v' = \frac{v - m_2^{-1}(t+h_d y_1'|x)}{h_r}, \quad u' = \frac{u - m_1^{-1}(m(x, m_2^{-1}(t|x)) + h_d x_1''|m_2^{-1}(t|x))}{h_r}$$

and a Taylor expansion yields

$$\mathrm{Var}(\sqrt{n h_r^2} \Delta_n^{(1.1.1.a.b)}(t|x))$$

$$= \frac{(1+o(1)) \sigma^2(x, m_2^{-1}(t|x))}{f(x, m_2^{-1}(t|x)) (\frac{\partial}{\partial y_1} m(x, m_2^{-1}(t|x)))^2} \int K_d(y_1') K_r(v') K_d(x_1'') K_r(u') K_d(y_2') K_d(x_2'')$$

$$\times \ K_r \left(\frac{m_1^{-1}(m(x, m_2^{-1}(t|x)) + h_d x_1''|m_2^{-1}(t|x)) - m_1^{-1}(m(x, m_2^{-1}(t|x)) + h_d x_2''|m_2^{-1}(t|x))}{h_r} + u' \right)$$

$$\times \ K_r \left(\frac{m_2^{-1}(t+h_d y_1'|x) - m_2^{-1}(t+h_d y_2'|x)}{h_r} + v' \right) dx_1'' dx_2'' dy_1' dy_2' du' dv'$$

$$= \frac{(1+o(1)) \sigma^2(x, m_2^{-1}(t|x))}{f(x, m_2^{-1}(t|x)) (\frac{\partial}{\partial y_1} m(x, m_2^{-1}(t|x)))^2} \left(\int K_r^2(u) du \right)^2.$$

The Ljapunoff condition can be checked by similar arguments, that is

$$\sum_{k=1}^{n} E\Big[\Big(\frac{\sigma(X_k, Y_k)}{n^{5/2}h_d^2 h_r}\varepsilon_k \sum_{i,j=1}^{n} \frac{\partial}{\partial x_1}m(x, \tfrac{i}{n})K_d\Big(\frac{m(x, \tfrac{i}{n}) - t}{h_d}\Big)K_d\Big(\frac{m(\tfrac{j}{n}, \tfrac{i}{n}) - m(x, \tfrac{i}{n})}{h_d}\Big)$$

$$\times\ K_r\Big(\frac{X_k - \tfrac{j}{n}}{h_r}\Big)K_r\Big(\frac{Y_k - \tfrac{i}{n}}{h_r}\Big)\frac{1}{f(\tfrac{j}{n}, \tfrac{i}{n})}\Big)^4\Big]$$

$$=\ \frac{E[\varepsilon_1^4](1 + o(1))}{nh_d^8 h_r^4}\int\int\sigma^4(u, v)f(u, v)\Big\{\prod_{l=1}^{4}\int\int\frac{\partial}{\partial x_1}m(x, y_l)K_d\Big(\frac{m(x, y_l) - t}{h_d}\Big)$$

$$\times\ K_d\Big(\frac{m(x'_l, y_l) - m(x, y_l)}{h_d}\Big)K_r\Big(\frac{u - x'_l}{h_r}\Big)K_r\Big(\frac{v - y_l}{h_r}\Big)\frac{dx'_l dy_l}{f(x'_l, y_l)}\Big\}dudv$$

$$=\ O\Big(\frac{1}{nh_r^2}\Big).$$

The asymptotic normality of the random variable $\sqrt{nh_r^2}\Delta_n^{(1.1.1.a.b)}(t|x)$ is obtained from Ljapunoffs Theorem, that is

$$\sqrt{nh_r^2}\Delta_n^{(1.1.1.a.b)}(t|x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_2^2(x, t)),$$

where the quantity $s_2^2(x, t)$ is defined in (3.13). The assertion of Theorem 3.3 finally follows from (5.17), (5.21), (5.22), (5.24), (5.26), (5.28), (5.32), and (5.36). $\qquad\square$

References

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *J. Econometrics*, 116(1-2):9–47. Frontiers of financial econometrics and financial engineering.

Bacchetti, P. (1989). Additive isotonic models. *J. Amer. Statist. Assoc.*, 84(405):289–294.

Bennett, C. and Sharpley, R. (1988). *Interpolation of operators*, volume 129 of *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA.

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

Blake, C. and Merz, C. (1998). UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html.

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

Collomb, G. and Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Process. Appl.*, 23(1):77–89.

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

Dykstra, R. L. and Robertson, T. (1982). An algorithm for isotonic regression for two or more independent variables. *Ann. Statist.*, 10(3):708–716.

Friedman, J. and Tibshirani, R. (1984). The monotone smoothing of scatterplots. *Technometrics*, 26:243–250.

Gallant, A. and Golub, G. H. (1984). Imposing curvature restrictions on flexible functional forms. *J. Econ.*, 26:295–321.

Gebhardt, F. (1970). An algorithm for monotone regression with one or more independent variables. *Biometrika*, 57:263–271.

Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.*, 29(3):624–647.

Härdle, W. (1990). *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and semiparametric models*. Springer Series in Statistics. Springer-Verlag, New York.

Kohavi, R. (1996). Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 202–207.

Lee, C. I. C. (1983). The min-max algorithm and isotonic regression. *Ann. Statist.*, 11(2):467–477.

Mammen, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.*, 19(2):724–740.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2523–2558. North-Holland, Amsterdam.

Mukarjee, H. and Stern, S. (1994). Feasible nonparametric estimation of multiargument monotone functions. *J. Amer. Statist. Assoc.*, 89(425):77–80.

Penrose, K., Nelson, A., and Fisher, A. (1985). Generalized body composition prediction equation for men using simple measurement techniques. *Medicine and Science in Sports and Exercise*, 17(2):189.

Qian, S. and Eddy, W. F. (1996). An algorithm for isotonic regression on ordered rectangular grids. *J. Comput. Graph. Statist.*, 5(3):225–235.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, 3(4):425–441.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22(3):1346–1370.

Ryff, J. V. (1965). Orbits of L^1-functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.*, 117:92–100.

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

Spokoiny, V. (2002). Variance estimation for high-dimensional regression models. *J. Multivariate Anal.*, 82(1):111–133.

Summerfield, L. M. (2001). *Nutrition, Exercise, and Behavior.* Wadsworth/Thomson Learning, c2001.

Wright, F. (1978). Estimating strictly increasing regression functions. *J. Am. Stat. Assoc.*, 73:636–639.