Shape constraints in multivariate regression

REGINE SCHEDER

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
an der Fakultät für Mathematik
der Ruhr-Universität Bochum

2007

Contents

Introduction

In many applications, the objective is to find a meaningful structure in a relationship between dependent and independent variables through a regression kind of model expressed by

$$Y = m(X) + \varepsilon,$$

where m is the so called *regression function* and ε a random error with zero mean.
Nowadays, many statistical problems are high-dimensional, since it is easier to collect data and modern computing power allows to consider massive amounts of information. Complex relations are evolved and evaluated more and more. The present thesis tries to answer some questions in this context by proposing new multivariate estimates utilizing shape constraints. Structured relationships in multivariate setups have not been considered in the literature that much probably due to inherent problems like data sparseness.
Regression analysis is concerned with fitting a curve to a finite set of points in a space. An estimated regression function enables to predict, evaluate, and interpret a relationship between an explanatory variable and a response variable. In numerous examples, order restricted inference is a reasonable approach to analyze the regression function. Order restrictions can model shape constraints like monotonicity or convexity. Consider a finite set of points $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$, where $\mathbf{X}_j \in \mathbb{R}^d$ and $Y_j \in \mathbb{R}$. The following figure illustrates a data set with a two-dimensional predictor variable. As in this data set, the scatterplot sometimes shows a monotonic trend in one or more directions. In other words, experimental evidence suggests that the underlying true regression function m is monotone increasing or decreasing with respect to some variables. In this example, it makes sense for researchers to suppose that the body fat of a person depends increasingly on weight and decreasingly on height.
To find the appropriate regression model is a challenging task, since there is a broad range available varying from parametric to nonparametric setups. An easy regression approach is to fit a linear regression line to the data, but often this simplifies the model too much. On the other end of the spectrum lies the nonparametric regression approach, where no assumption is made regarding the particular shape of the regression function. Sometimes further information or physical laws make researchers believe that the underlying regression curve belongs to an order restricted class of functions. There is a considerable amount of literature available about this topic [see Barlow et al. (1972) or Robertson

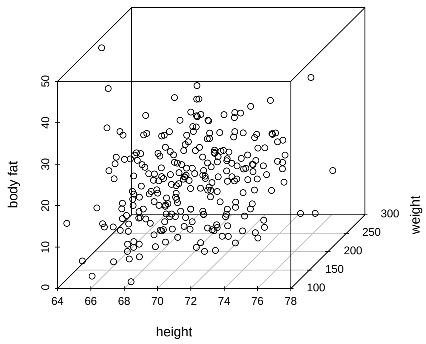

Figure 1: Data set `Fat` from `R` Package `UsingR`.

et al. (1988) as introductory compendium]. But most of this *statistical inference under order restrictions* refers to the univariate regression model. The pool-adjacent-violators algorithm (PAVA) is the most widely used algorithm for computing the isotonic regression for a univariate covariate first published by Ayer et al. (1955)[1]. The PAVA computes for a given set $\{(X_j, Y_j)\}_{j=1}^n$ with $X_j \in I\!R$ and sorted data $\{(X_{(j)}, Y_{[j]})\}_{j=1}^n$ ($Y_{[j]}$ is ordered relative to $X_{(j)}$), respectively, the values $\{\hat{m}(X_{(j)})\}_{j=1}^n$, which minimize

$$\frac{1}{n}\sum_{j=1}^{n}\left(Y_{[j]} - \hat{m}(X_{(j)})\right)^2$$

subject to the monotonicity restriction

$$\hat{m}(X_{(1)}) \leq \hat{m}(X_{(2)}) \leq \ldots \leq \hat{m}(X_{(n)}).$$

There are several extensions and modifications of this algorithm, but multivariate versions are available only for special cases [see Gebhardt (1970), Dykstra and Robertson (1982), Lee (1983), or Qian and Eddy (1996), among others].
The main results of this thesis can be divided into two parts, which are detached from each other. Nevertheless, all results give new insights into nonparametric multivariate relationships and therefore allow to analyze complex high-dimensional statistical problems under additional constraints.
In the first part of this thesis, a new strictly monotone increasing regression estimator with $d \geq 2$ explanatory variables is proposed [see Chapter 2]. This estimator is introduced and analyzed in a nonparametric regression context. But as an important feature

[1]van Eeden (1956) developed independently the same procedure and kindly sent me her paper.

of this estimate, the fundamental technique can be used in any regression setting. The idea of this isotonic or antitonic regression approach is developed in the univariate case by Dette et al. (2006) and extended to higher dimensions in this thesis.
In the second part of this thesis , the focus lies on nonparametric quantile regression models [see Chapter 3]. In the framework of regression models, conditional quantile models can be described as

$$Y = Q_\alpha(X) + \varepsilon,$$

where Q_α is the quantile function and ε an error term, whose α quantile is zero. The quantile regression approach originally proposed by Koenker and Bassett (1978) offers a complete picture of the relationship between variables. Instead of just averaging over certain values of X_j, regression quantiles give new understanding, since several quantiles can be computed. Koenker and Bassett (1978) suggest to estimate the conditional quantile by minimizing the check function [see Definition 3.1], while in regression models the quadratic loss function is used. For each quantile an own model is formulated and the corresponding minimization problem is solved. In Figure 2, the `Prestige` data from the `R` package `effects` is analyzed with the `lprq` function implemented by Koenker available in the `quantreg` package. The resulting estimate is a nonparametric quantile estimator using the check function [see Koenker (2005) as review compendium]. The estimation of conditional quantiles via check function suffers from the fact that the quantile curves may accidently cross each other [see Figure 2], which they are not supposed to do, since the quantile function fixed on a certain covariate is in fact the inverse of the conditional distribution function and therefore monotone increasing with respect to α. Hence,

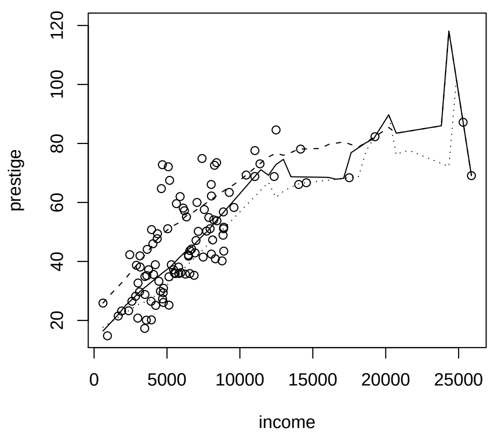

Figure 2: Local linear quantile estimates by check function: 0.5 quantile (straight line), 0.25 quantile (dotted line), and 0.75 quantile (dashed line).

we concentrate on a different approach. Dette and Volgushev (2007) recently developed a nonparametric conditional quantile estimator. Their approach starts with a nonparametric estimate for the conditional distribution function, which is then isotonized and reversed at the same time using the monotonization idea from Chapter 2. For this reason the quantile curves do not cross anymore. Again, this idea is extended to a multivariate setting with the main focus on additive quantile models.
We close the introduction with a quotation by Francis Galton, who formed the notion of regression as well as the concept of quantiles as one of the first.

> *The Charms of Statistics.*
> It is difficult to understand why statisticians commonly limit their inquiries to Averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once.
>
> Sir Francis Galton, *Natural Inheritance* (1889), p. 62

In this spirit, we conclude that the developed estimates in the multivariate monotone nonparametric regression model and in the additive conditional quantile model provide new tools that there is less need to throw mountains in the lake of averages.

Concisely, the outline of this thesis is as follows. In Chapter 1, we start with an overview of nonparametric regression in the context of an univariate covariate. We introduce kernel regression estimates and discuss their properties. Moreover, we present a procedure to monotonize a given function. In the literature, this method is called *measure preserving rearrangement* [see Ryff (1970)]. Dette et al. (2006) transfered this principle to a nonparametric regression setting under monotonicity constraints. Chapter 2 starts with the generalization of nonparametric regression models to higher dimensions. Afterwards, the monotonizing procedure introduced in Chapter 1 is extended to higher dimensional estimation problems. This yields a strictly monotone regression estimate in d dimensions in the framework of nonparametric regression. The asymptotic properties of this estimate is derived. In the last Chapter 3, we utilize the idea of measure preserving rearrangement in a different context. Estimates for the additive conditional quantile model are proposed and analyzed asymptotically. Chapter 2 and Chapter 3 close with finite sample studies of the developed estimates.

Chapter 1

Nonparametric estimation

1.1 Overview

This chapter starts with a summary of regression models in the context of an univariate covariate. Motivated by linear regression, nonparametric regression techniques are introduced as a flexible technique to model nonlinear data. In particular, kernel regression estimators and their properties are regarded to get familiar with this approach in the univariate case. Furthermore, we explain the difference between external and internal kernel-type estimators from a bit philosophical point of view. At the end of this chapter, we present a monotonizing procedure introduced by Dette et al. (2006) in the framework of nonparametric regression. The basic properties of this procedure are discussed and ranked among methods for shape constraints. Most of the ideas and procedures in this chapter are illustrated by data examples.

1.2 Regression models

In statistics, regression analysis is one of the most commonly used technique. The aim of such an analysis is to study the relationship between an explanatory variable X and a response variable Y, to measure the influence or the effect of the independent variable on the dependent variable. The regression function describes this relation between X and Y, where Y is a function of X. In this section, we will focus on a nonparametric regression setup, which will provide the framework of this thesis. There exits a a broad literature on nonparametric regression techniques. We refer to the monographs of Fan and Gijbels (1996), Wand and Jones (1995), or Härdle (1990) to get an overview of the existing methods. But before we start to introduce the key idea of local modeling, we have a brief look at parametric regression models.

1.2.1 A brief survey of linear Regression

Linear models are a classical methodology in statistics, which is still widely used in many applied sciences. This technique is well understood, and there are many tools and implementations available. For a given set of data $\{(\mathbf{X}_j, Y_j)\}_{j=1}^n$, one assumes that Y_j is a linear combination of d unknown quantities $\beta_1, \ldots, \beta_d$ plus error $\varepsilon_1, \ldots, \varepsilon_n$,

$$Y_j = X_{j1}\beta_1 + X_{j2}\beta_2 + \ldots + X_{jd}\beta_d + \varepsilon_j,$$

where $\mathbf{X}_j = (X_{j1}, \ldots, X_{jd})$. The random errors ε_j have expected value zero

$$E[\varepsilon_j] = 0, \quad (j = 1, \ldots, n),$$

are uncorrelated, and have all the same variance called homoscedastic

$$E[\varepsilon_j \varepsilon_i] = \sigma^2 \delta_{ji},$$

where δ_{ji} is 0 or 1, according as $j \neq i$ or $i = j$, respectively. In the classical *analysis of variance*, the values X_{jk} are assumed to be *counter* or *indicator variables* which refers to the presence or absence of the effect β_k. In other words, this means X_{jk} is either 0 or 1. See Scheffé (1999) for a classical reference on the analysis of variance. Our primary interest is another case called *linear regression*, where the X_{jk} are continuous variables. Under the above mentioned assumptions on the random error, the conditional mean $E[Y|\mathbf{X}]$ is a linear function:

$$E[Y|\mathbf{X}] = X_1\beta_1 + \ldots + X_d\beta_d,$$

where the expectation is conditioned on the vector $\mathbf{X} = (X_1, \ldots, X_d)^T$. If the independent variable X is one-dimensional, a straight line is a simple linear regression model,

$$Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j, \ \ (j = 1, \ldots, n).$$

This model often refers to the term *linear regression* in contrast to polynomial regression models where a higher degree polynomial is used as the regression curve. The framework of linear regression includes polynomial regression as well

$$Y_j = \beta_0 + \beta_1 X_j + \beta_2 X_j^2 + \ldots + \beta_p X_j^p + \varepsilon_j \ \ (j = 1, \ldots, n).$$

Polynomial regression offers an easy model framework for regression analysis. Fitting a polynomial regression to data $\{(X_j, Y_j)\}_{j=1}^n$ using the method of least squares means solving

$$\min_{\beta_0, \beta_1, \ldots, \beta_p} \sum_{j=1}^{n} \left(Y_j - \beta_0 + \beta_1 X_j + \beta_2 X_j^2 + \ldots + \beta_p X_j^p \right)^2. \tag{1.1}$$

There is an explicit formula for the solution of the above minimization problem, which is

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 & X_1^2 & \ldots & X_1^p \\ 1 & X_2 & X_2^2 & \ldots & X_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_n & X_n^2 & \ldots & X_n^p \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}.$$

It is not easy to determine the degree p. A widely used approach is to increase the degree of the polynomial step by step and find the best model fit. Still the linearity assumption is often quite restrictive, which will be discussed in the following examples.

Example 1.1 (i) The motorcycle data set `mcycle` in `R` contains a series of measurements of head accelerations in a simulated motorcycle accident, used to test crash helmets. The independent variable X is `time` (in milliseconds after the impact) and the dependent variable Y is `accel`, the acceleration (in g). We fit three parametric models:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{1.2}$$

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon, \tag{1.3}$$

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon. \tag{1.4}$$

From a first look at the scatterplot in Figure 1.1, it is observable that a polynomial fit might be problematic. Accordingly, the three polynomial fits have large biases. Furthermore, polynomial functions have derivatives of all order, which seem to be not the best approach in this example.

(ii) The data set `PublicSchools` shows the US expenditure per capita on public schools Y as a function of the per capita income X by state in 1979. Again, we fit three different polynomials to the data. In parametric regression all observations are treated the same, which can imply that a single data point has a huge influence on the curve. Figure 1.2 shows that `Alaska` is a clear outlier and has large influence on the fitting of higher degree polynomials.

1.2.2 Nonparametric Regression

Having seen the drawbacks of polynomial regression, the nonparametric regression methods fit polynomial curves locally to repair the defects of the global parametric approach. In general, the idea of nonparametric regression is that one is only willing to assume

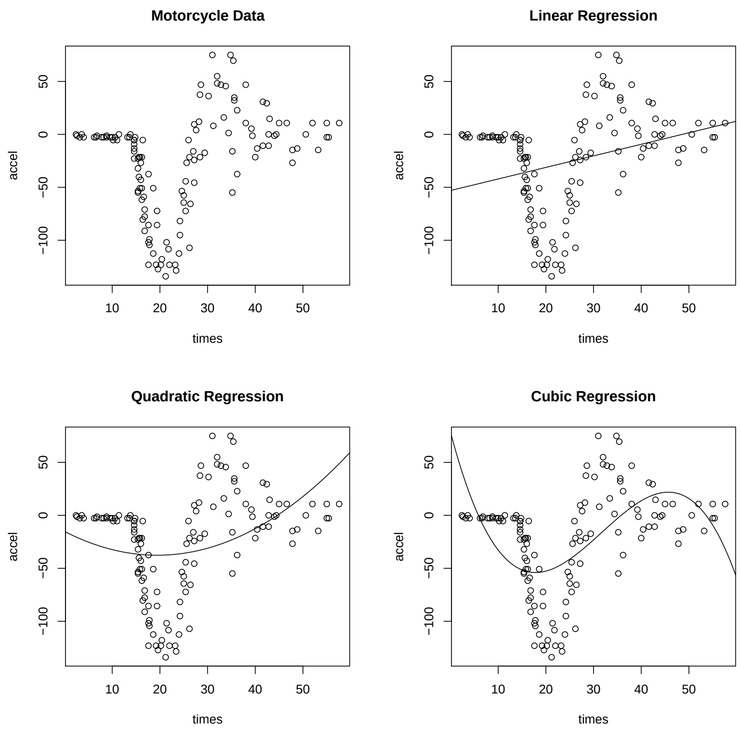

Figure 1.1: Scatterplot of the motorcycle data and three polynomial fits (1.2)- (1.4) (from top right to bottom right).

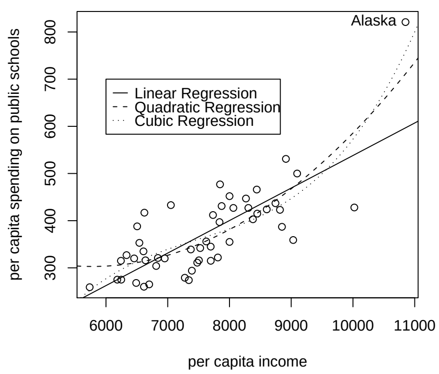

Figure 1.2: Three polynomial regression fits for the US expenditure for public schools data.

some smoothness condition on the regression function $m(\cdot)$, which is modeled as the conditional expectation

$$E[Y|X = x] = m(x).$$

As in the parametric setup, for given data $\{(X_j, Y_j)\}_{j=1}^n$ we consider the nonparametric regression model

$$Y_j = m(X_j) + \sigma(X_j)\varepsilon_j \quad (j = 1, \dots, n),$$

where the regression function m is twice continuously differentiable, the variance function σ is positive and continuous, and the random error has zero expectation and is independent, identically distributed noise. In this model, we allow the variance to be heteroscedastic by introducing $\sigma(\cdot)$ as a function of X. In the polynomial regression, a specific form is used for the regression curve chosen beforehand

$$m(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p.$$

Nonparametric regression techniques can be classified into three main groups: *Splines*, *Orthogonal Series*, and *Kernels*. The spline approach allows discontinuities of derivative curves, which are located in so called *knots*, between these knots, e.g., cubic polynomials are fitted. The knots can be selected by data via smoothing spline method or by a stepwise deletion method. Another approach is to expand the regression function into an orthogonal series, where about an appropriate chosen basis function set the coefficients have to be estimated. The third idea is a local modeling approach, on which we will focus in this thesis. Instead of fitting a line to the whole data, we fit for any given point x a linear regression to a fraction of the data around x. The size of the local neighborhood or local window is called bandwidth h. We are using kernel functions to incorporate local modeling. Let K be a nonnegative function with compact support. As long as the kernel function K decays fast enough it does not have to be compactly supported, but to make things more comprehensible we will assume that K is supported on $[-1, 1]$. For a given point x, we assign the weight $K\left(\frac{x - X_j}{h}\right)$ to the observation (X_j, Y_j). All data with $X_j \in x \pm h$ has positive weight and represent the local window. With these weights the minimization problem (1.1) can be rewritten in

$$\min_{\beta_0, \beta_1} \sum_{j=1}^n \{Y_j - \beta_0 - \beta_1(X_j - x)\}^2 K\left(\frac{x - X_j}{h}\right)$$

for the local linear case. To get a better feeling for this idea, we will illustrate this approach on the data sets from the last subsection.

Example 1.2 For the motorcycle and the `PublicSchools` data, the kernel smoothing approach is applied to fit a linear regression locally. In Figure 1.3, it is abundantly clear that this method to estimate the regression function in both examples works better. The modeling bias is reduced and the relationship between the variables is described more reasonably as in Example 1.1. In both data sets, we use the so called Epanechnikov kernel

$$K(u) = \frac{3}{4}(1-u^2)I_{[-1,1]}(u),$$

where

$$I_{[-1,1]}(u) = \begin{cases} 1 & x \in [-1,1] \\ 0 & \text{else} \end{cases} .$$

For the motorcycle data the local window $x \pm 3.3$ is used. In the next subsection, we will give a precise definition of local polynomial estimates and how to determine the local window size. Figure 1.3 shows the regression estimates.

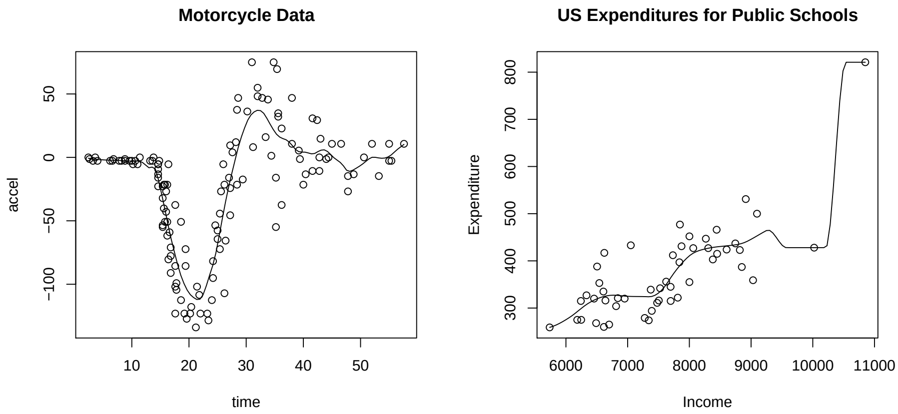

Figure 1.3: Illustrations of the local modeling approach.

1.2.3 Kernel Regression Estimators

In this subsection, we summarize key aspects about kernel estimation in the regression setting. The exact definition of the estimators and their statistical properties are given. We restrict ourselves to the case of one-dimensional predictors to get a better understanding for the methodology, which will be helpful for the extension to higher dimensions. In the subsequent chapters, we strictly focus on the multivariate case.

To start with the statistical framework, we assume that the given data $\{(X_j, Y_j)\}_{j=1}^n$ is independently and identically distributed (i.i.d.) and comes from the nonparametric heteroscedastic regression model

$$Y_j = m(X_j) + \sigma(X_j)\varepsilon_j, \quad (j = 1, \ldots, n),$$

where $E[\varepsilon_j] = 0$ and $\mathrm{Var}(\varepsilon_j) = 1$ that $E[Y|X = x] = m(x)$. Further we define kernels of order $q \in I\!N$. Let K be a real-valued function with

$$\int_{I\!R} u^k K(u)du = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k = 1, \dots, q-1 \\ c & \text{if } k = q \end{cases},$$

$||K||_2^2 = \int_{I\!R} K^2(u)du < \infty$, and $\lim_{|u|\to\infty} |u|K(u) = 0$. Then K is called a kernel of order q. For $q = 2$, one usually assumes that K is a symmetric probability density, which essentially means that above all $K(u) \geq 0$. Often additionally, K has to be compactly supported. If K is symmetric, the order q is an even number. The following table outlines some important kernel functions.

Kernel	$K(u)$
Uniform	$\frac{1}{2} I_{[-1,1]}(u)$
Triangle	$(1-|u|) I_{[-1,1]}(u)$
Epanechnikov	$\frac{3}{4}(1-u^2) I_{[-1,1]}(u)$
Biweight	$\frac{15}{16}(1-u^2)^2 I_{[-1,1]}(u)$
Triweight	$\frac{35}{32}(1-u^2)^3 I_{[-1,1]}(u)$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$
Cosine	$\frac{\pi}{4}\cos(\frac{\pi}{2}u) I_{[-1,1]}(u)$

Table 1.1: Kernel functions

In the previous subsections, we presented kernel smoothing as the idea of fitting a polynomial regression in a small neighborhood around a given point x. Let h be a positive number controlling the size of the neighborhood. We call h a *bandwidth*. Denote $K_h(\cdot) = K(\cdot/h)/h$. Local polynomial estimators of degree p are defined via the minimization problem

$$\min_{\beta_0,\beta_1,\dots,\beta_p} \sum_{j=1}^{n} \left\{ Y_j - \sum_{k=0}^{p} \beta_k (X_j - x)^k \right\}^2 K_h(x - X_j). \tag{1.5}$$

If we denote $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T$ as the minimizer of (1.5), the estimator for $m^{(\nu)}(x)$ is given by

$$\hat{m}^{(\nu)}(x) = \nu! \hat{\beta}_\nu \quad (\nu = 0, \dots, p).$$

This relationship is easily motivated by a taylor expansion of the regression function m since

$$m(z) \approx \sum_{k=0}^{p} \frac{m^{(k)}(x)}{k!}(z-x)^k \equiv \sum_{k=0}^{p} \beta_k (z-x)^k.$$

The minimization problem (1.5) can be solved via least square method like the parametric counterpart. The explicit solution for $\hat{\beta}$ is

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^T = \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y},$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x) & \ldots & (X_1 - x)^p \\ 1 & (X_2 - x) & \ldots & (X_2 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x) & \ldots & (X_n - x)^p \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},$$

and $\mathbf{W} = \mathrm{diag}(K_h(x - X_1), \ldots, K_h(x - X_n))$, which basically is the solution of a weighted least squares problem provided the invertibility of $\mathbf{X}^T\mathbf{W}\mathbf{X}$. In this setup, two estimators are worth mentioning explicitly in the context of the general local polynomial method: the Nadaraya-Watson estimator and the local linear estimator. The Nadaraya-Watson estimate corresponds to the case $p = 0$ and is sometimes called the local constant estimator, i.e.

$$\hat{m}_{NW}(x) = \frac{\sum_{j=1}^n K_h(x - X_j)Y_j}{\sum_{j=1}^n K_h(x - X_j)}. \tag{1.6}$$

See Nadaraya (1964) and Watson (1964) for further reference. This estimator is basically a weighted sum over the response data Y_j.To obtain the local linear estimator, where $p = 1$, observe that $\hat{\beta}$ can be written as

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T = \begin{pmatrix} S_0(x) & S_1(x) \\ S_1(x) & S_2(x) \end{pmatrix}^{-1} \begin{pmatrix} T_0(x) \\ T_1(x) \end{pmatrix},$$

where

$$\begin{aligned} S_k(x) &= \frac{1}{n}\sum_{j=1}^n (X_j - x)^k K_h(x - X_j), \\ T_k(x) &= \frac{1}{n}\sum_{j=1}^n (X_j - x)^k K_h(x - X_j)Y_j. \end{aligned}$$

This notation yields the explicit expression

$$\hat{m}_{LL}(x) = \frac{T_0(x)S_2(x) - T_1(x)S_1(x)}{S_0(x)S_2(x) - S_1^2(x)}.$$

In principle, the local linear estimate can be expressed as a weighted sum over Y_j, but in contrary to the Nadaraya-Watson estimator the weights are not positive inherently provided the kernel is positive.

In the following, we discuss some statistical properties of the Nadaraya and Watson estimator and the local linear estimate. We start and formulate some basic model assumptions.

(A1) $\{(X_j, Y_j)\}_{j=1}^n$ form a sample of independent and identically distributed observations, where X_j has a q times continuously differentiable density p supported on $[0, 1]$.

(A2) The variance function $\sigma : [0, 1] \to I\!R^+$ is continuous.

(A3) The regression function $m : [0, 1] \to I\!R$ is q times continuously differentiable.

(A4) X_j and ε_j are independent, and $E[\varepsilon_j] = 0$, $E[\varepsilon_j^2] = 1$, and $E[\varepsilon_j^4] < \infty$ for $j = 1, \ldots, n$.

(A5) K is a kernel of order q supported on $[-1, 1]$. Define the constant

$$\kappa_s(K) = \frac{(-1)^s}{s!} \int_{-1}^{1} u^s K(u) du. \tag{1.7}$$

In the following theorem, the asymptotic biases and variances are given. Furthermore, the asymptotic distribution of the estimators $\hat{m}_{NW}(x)$ and $\hat{m}_{LL}(x)$ is established.

Theorem 1.3 *Suppose that the assumptions (A1) - (A5) hold, and the bandwidth fulfills $nh \to \infty$ and $h = cn^{-1/(2q+1)}$ for a constant $c \in (0, \infty)$. Then for $x \in (0, 1)$, we have*

(i) for the Nadaraya and Watson estimator

$$\sqrt{nh}(\hat{m}_{NW}(x) - m(x) - b_{NW}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x)),$$

where

$$\begin{aligned} b_{NW}(x) &= h^q \kappa_q(K) \frac{(mp)^{(q)}(x) - mp^{(q)}(x)}{p(x)}, \\ s^2(x) &= ||K||_2^2 \frac{\sigma^2(x)}{p(x)}; \end{aligned}$$

(ii) for the local linear estimate

$$\sqrt{nh}(\hat{m}_{LL}(x) - m(x) - b_{LL}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x)),$$

where $s^2(x)$ is defined above and

$$b_{LL}(x) = h^q \kappa_q(K) m^{(q)}(x).$$

For the expression of the bias and variance term of the estimates $\hat{m}_{NW}$ and $\hat{m}_{LL}$, it is not necessary that the bandwidth is of order $O(n^{-1/5})$. In addition, a byproduct of this result is the asymptotic mean squared error of both, the Nadaraya-Watson and the local linear estimate

$$\begin{aligned}\mathrm{AMSE}(\hat{m}_{NW}(x)|X_1,\ldots,X_n) &= \frac{1}{nh}||K||_2^2\frac{\sigma^2(x)}{p(x)} + h^{2q}\kappa_q^2(K)\left(\frac{(mp)^{(q)}(x) - mp^{(q)}(x)}{p(x)}\right)^2, \\ \mathrm{AMSE}(\hat{m}_{LL}(x)|X_1,\ldots,X_n) &= \frac{1}{nh}||K||_2^2\frac{\sigma^2(x)}{p(x)} + h^{2q}\kappa_q^2(K)(m^{(q)}(x))^2.\end{aligned}$$

Minimizing the asymptotic MSE yields the optimal bandwidth h for the two estimators. In the case of the Nadaraya-Watson estimate, the optimal bandwidth h^*_{NW} is given by

$$h^*_{NW} = \left(\frac{||K||_2^2\sigma^2(x)}{2qn\kappa_q^2(K)((mp)^{(q)}(x) - mp^{(q)}(x))}\right)^{1/(2q+1)}.$$

Since the optimal bandwidth depends on the unknown functions m, σ, and p, for a given estimation problem the bandwidth cannot be determined in this way. A simple rule of thumb is to estimate the variance function σ and use $\hat{h}^* = \left(\frac{\hat{\sigma}^2}{n}\right)^{1/(2q+1)}$, where $\hat{\sigma}^2$ is an estimate of $E[\sigma^2(X)]$. Another approach is using cross validation to select the bandwidth. For this method, one minimizes

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{-i}(X_i)\}^2,$$

where the estimate $\hat{m}_{-i}(X_i)$ is a kernel estimator using the bandwidth h and the data $\{(X_j, Y_j)\}$ but the observation (X_i, Y_i). Typically, one tries to minimize $\mathrm{CV}(h)$ over an appropriate interval. The bandwidth selection is an important issue in kernel smoothing techniques. In the literature, there are different approaches to deal with this problem [see, e.g., Fan and Gijbels (1996)].
It is worth mentioning that the optimal bandwidth h^*_{NW} gives the rate of convergence (similarly in the local linear case), which is of order $O(n^{-2q/(2q+1)})$ for the AMSE slower than in the parametric regression. In fact, a parametric approach yields better results, if the underlying parametric model is true. On the other hand, most data cannot be expected to come from a parametric model. An inaccurate parametric model can create large modeling biases, which then can lead to wrong conclusions. Nonparametric methods try to reduce this modeling bias and fit a curve from a larger class [see Fan (2000) for prospect of nonparametric methods].

1.2.4 External and internal methods

In this subsection, we identify and distinguish between two types of kernel estimators, which will be important in the discussion of Chapter 3. The random variables $Y_1, \ldots, Y_n$ contain not only information about the regression function m but also about the distribution of the $X_1, \ldots, X_n$. Basically, the dependent variable Y corresponds to $r(x) \equiv m(x) \times p(x)$. There are many estimators which use the relationship $m = \frac{mp}{p}$ to estimate the regression function, e.g. recall the Nadaraya-Watson estimator in (1.6). We assume for the moment that the density function p is known. Estimators of the form $\frac{\hat{r}(x)}{p(x)}$ are a natural choice, namely,

$$\hat{m}_E(x) = \frac{1}{np(x)} \sum_{j=1}^{n} K_h(x - X_j) Y_j.$$

This is the prototype estimator of *external methods*. External refers to the the density function p which appears in the external denominator. In contrary, the other type of estimators modifies the data in the first hand. Instead of using $Y_1, \ldots, Y_n$, the internal method applies the adjusted dependent variables $\frac{Y_1}{p(X_1)}, \ldots, \frac{Y_n}{p(X_n)}$. The prototype of the internal methods is

$$\hat{m}_I(x) = \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j) \frac{Y_j}{p(X_j)}.$$

Interestingly, if p is assumed to be an uniform density, $\hat{m}_E(x)$ and $\hat{m}_I(x)$ yield the same estimates. This is not the case if p is a nonuniform density. The expectation of the two estimates differs clearly [see Jones et al. (1994) for more details on this subject]. In the random design case, the density p is unknown and has to be estimated as well. The Nadaraya-Watson estimator $\hat{m}_{NW}(x)$ (1.6), is the most famous example for external estimates if p is unknown. The estimator for p is the kernel density estimator

$$\hat{p}(x) = \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j),$$

where the same bandwidth is used as for estimating $r(x)$.
Mack and Müller (1989) propose the internal version of the Nadaraya-Watson estimator

$$\hat{m}_{INW}(x) = \frac{1}{n} \sum_{j=1}^{n} K_h(X_j - x) \frac{Y_j}{\hat{p}(X_j)},$$

where $\hat{p}$ is the kernel density estimator. This estimator is appealing from a computational point of view, since the kernel density estimator has to be calculated only for $X_1, \ldots, X_n$. In the following theorem, we demonstrate the asymptotic normality of $\hat{m}_{INW}(x)$. We obtain that this estimator is less efficient than the usual Nadaraya and Watson estimator.

Theorem 1.4 *Under the assumptions (A1)-(A5) from Subsection 1.2.3 and the bandwidth conditions $nh \to \infty$ and $h = cn^{-1/(2q+1)}$ for a constant $c \in (0, \infty)$, we have*

$$\sqrt{nh}(\hat{m}_{INW}(x) - m(x) - b_{INW}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_I^2(x)),$$

where

$$\begin{aligned} b_{INW} &= h^q \kappa_q(K) \left(m^{(q)}(x) - \frac{m(x)p^{(q)}(x)}{p(x)} \right), \\ s_I^2(x) &= ||K||_2^2 \frac{\sigma^2(x) + m^2(x)}{p(x)}. \end{aligned}$$

Mack and Müller (1989) use a slightly slower converging bandwidth h_p for the kernel density estimator in the denominator in contrast to just applying h as well. Under this additional condition they get rid of the second term in the bias. This causes the same bias as the local linear estimator.

Remark 1.5 The local linear estimator can be regarded as an estimator of an internal form. Recall

$$\hat{m}_{LL}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{S_2(x) - (X_j - x)S_1(x)}{S_0(x)S_2(x) - S_1^2(x)} K_h(x - X_j)Y_j.$$

If the kernel K is of order $q = 2$, we have $S_0(x) \sim p(x)$, $S_1(x) \sim 2h^2\kappa_2(K)p'(x)$, and $S_2(x) \sim 2h^2\kappa_2(K)p(x)$ in terms of expectation. This yields the representation for the fraction in the above representation of $\hat{m}_{LL}(x)$

$$\begin{aligned} & \left(S_0(x)S_2(x) - S_1^2(x)\right)^{-1} \times (S_2(x) - (X_j - x)S_1(x)) \\ \sim \; & (2h^2\kappa_2(K)p^2(x))^{-1}(2h^2\kappa_2(K)p(x) - (X_j - x)2h^2\kappa_2(K)p'(x)) \\ = \; & \left(\frac{1}{p}\right)(x) - (X_j - x)\left(\frac{p'}{p^2}\right)(x) \\ = \; & \left(\frac{1}{p}\right)(x) + (X_j - x)\left(\frac{1}{p}\right)'(x) \approx \frac{1}{p(X_j)} \end{aligned}$$

To make the result of the above theorem more comprehensible, we give a sketch of the proof.

Sketch of the Proof of Theorem 1.4 We decompose the internal Nadaraya-Watson estimator in the following way

$$\begin{aligned} \hat{m}_{INW}(x) &= \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j) \frac{Y_j}{\hat{p}(X_j)} \left(\frac{\hat{p}(X_j)}{p(X_j)} + (1 - \frac{\hat{p}(X_j)}{p(X_j)}) \right) \\ &= \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j) \frac{Y_j}{p(X_j)} + \frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j) \frac{Y_j}{\hat{p}(X_j)} \left(\frac{p(X_j) - \hat{p}(X_j)}{p(X_j)} \right). \end{aligned}$$

For the first term, we can easily derive

$$\begin{aligned} E\left[\frac{1}{n} \sum_{j=1}^{n} K_h(x - X_j) \frac{Y_j}{p(X_j)} \right] &= \int K_h(x - u)m(u)du \\ &= m(x) + h^q \kappa_q(K) m^{(q)}(x) \end{aligned}$$

using $Y_j = m(X_j) + \sigma(X_j)\varepsilon_j$. For the second term, a similar decomposition as in the first step has to be used, since the term is not negligible, and the kernel density estimator $\hat{p}$ can not be just exchanged by the true density as in the case of the Nadaraya-Watson estimator.

$$\frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j)\frac{Y_j}{\hat{p}(X_j)}\left(\frac{p(X_j) - \hat{p}(X_j)}{p(X_j)}\right) = \Delta_n(x) + o_p\left(h^q + \frac{1}{nh}\right),$$

where

$$\Delta_n(x) = \frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j)\left(\frac{p(X_j) - \hat{p}(X_j)}{p^2(X_j)}\right)Y_j.$$

To calculate the expectation of $\Delta_n(x)$, we first condition on X_j and then use the asymptotic bias of the kernel density estimator $\hat{p}$ [see Härdle et al. (2004) p. 47 ff].

$$\begin{aligned}
E[E[\Delta_n(x)|X_j]] &= -h^q\kappa_q(K)E\left[\frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j)\frac{p^{(q)}(X_j)}{p^2(X_j)}m(X_j)\right] \\
&= -h^q\kappa_q(K)\int K_h(x-u)\frac{p^{(q)}(u)}{p(u)}m(u)du \\
&= -h^q\kappa_q(K)\int K(v)\frac{p^{(q)}(x-hv)}{p(x-hv)}m(x-hv)dv \\
&= -h^q\kappa_q(K)\frac{p^{(q)}(x)m(x)}{p(x)} + o(h^q).
\end{aligned}$$

Note that a slightly slower bandwidth for the kernel density estimator $\hat{p}$ yields $\Delta_n(x)$ to be of order $o(h^q)$. To estimate the variance of $\hat{m}_{INW}(x)$, we detect that only

$$\frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j)\frac{Y_j}{p(X_j)}$$

contributes. We compute the variance and obtain

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j)\frac{m(X_j) + \sigma(X_j)\varepsilon_j}{p(X_j)}\right) &= \frac{1}{n}\int K_h^2(x-u)\frac{m^2(u)+\sigma^2(u)}{p(u)}du \\
&\quad -\frac{1}{n}(m(x) + h^q\kappa_q(K)m^{(q)}(x)) \\
&= \frac{1}{nh}||K||_2^2\frac{m^2(x)+\sigma^2(x)}{p(x)}(1+o(1)).
\end{aligned}$$

1.3 Monotonizing Procedure

In the previous section, the framework of nonparametric regression was briefly summarized. We motivated the kernel smoothing techniques by a local parametric fitting idea. The polynomial regression globally (parametric) and locally (nonparametric), respectively, are in some sense two methods on either side of the spectrum of fitting a regression model. The parametric model may be too smooth, while the unrestricted nonparametric approach is likely to be too erratic. A regression model somewhere in between these two extremes is in most cases more adequate. However, in implementations to fit a polynomial is convenient, even though this is for computational reasons than that one assumes this particular parametric form. Whereas nonparametric regression models assume that the underlying regression curve belongs to a quite general class of smooth function, additional shape constraints downsize this class to a often more reasonable and natural class of functions.

In many situations, researchers strongly believe in a particular form of the regression function. For example, they investigated that the regression curve has to be unimodal, convex, or monotone. Unimodality occurs if a regression curve is monotone increasing up to a certain value and monotone decreasing afterwards. If the covariate is the time and the response variable counts the outbreaks of a disease, it can be reasonable to assume that at the beginning more and more people get infected until it reaches a peak, and later on this number goes slowly back. In this thesis, we will discuss monotonic shape constraints. The vocable *monotonic* can mean either isotonic or antitonic. Isotonic or monotone increasing indicates that two variables have the same tone, whereas antitonic or monotone decreasing says the converse. Monotonic relationships are often important for interpretation. Actually, if an estimate for a monotone regression function does not fulfill the monotonicity condition, researchers fail to draw a meaningful conclusion.

In the literature, there is a broad variety of methods to estimate a monotone regression function. See Barlow et al. (1972) or Robertson et al. (1988) for a comprehensive and detailed introduction into statistical inference under order restriction. Most of these approaches only deal with the univariate regression model. In particular, there is a new method proposed by Dette et al. (2006) in the context of nonparametric regression models, which compared to other more intrinsic methods is an easy applied methodology. Actually, the procedure can also be used in other regression settings. In the following, the basic idea and the application of this monotonizing procedure is described.

Example 1.6 To give an example for a monotone nonparametric regression setup, we analyze the historical `cars` data set in the `datasets` package in `R` containing speed and stopping distances of 50 cars. The more speed a car has the longer is the distance taken to stop. In Figure 1.4, you can compare an unconstrained Nadaraya and Watson estimator in `R` implemented through the function `ksmooth`, the estimate suggested by Dette et al. (2006) and implemented through `monoproc`, which is a smooth and monotonized version

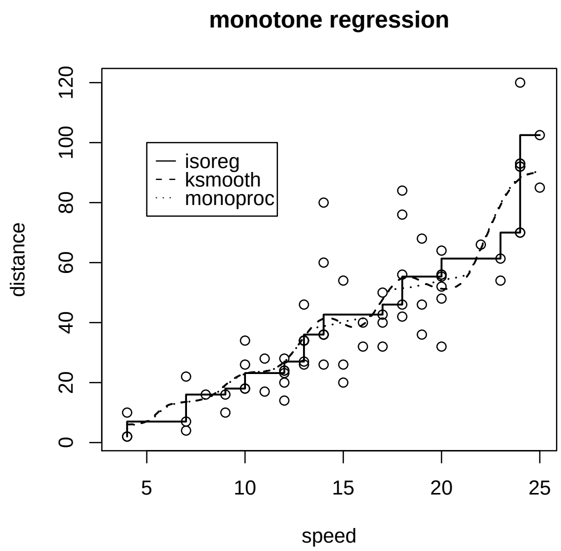

Figure 1.4: Different regression estimates.

of the Nadaraya and Watson estimate, and the isotonic least square regression which is piecewise constant by the function `isoreg`. For many examples, a piecewise constant monotone function may be sufficient for an analysis, but often practitioners disapprove the "flat" spots of this estimator [see Wright (1978)].

1.3.1 Motivation

There are two ways to motivate the general idea of the monotonizing procedure from Dette et al. (2006). In the context of nonparametric estimation, we consider an independent and identically distributed sample of uniformly distributed random variables, say $U_1, \ldots, U_n \sim \mathcal{U}([0,1])$. For an arbitrary strictly increasing and continuous function g on the interval $[0,1]$ with positive derivative, we define the kernel density estimate for the random variable $g(U_1)$

$$\frac{1}{Nh_m} \sum_{i=1}^{N} K_m \left(\frac{g(U_i) - u}{h_m} \right),$$

where K_m is a positive kernel function with the corresponding bandwidth h_m. The index m indicates the monotonization aspect of the kernel and its bandwidth and is used throughout this thesis. The density of $g(U_1)$ is in fact $(g^{-1})'(u) I_{[g(0),g(1)]}(u)$, so that the integrated density estimator

$$\frac{1}{Nh_m}\sum_{i=1}^{N}\int_{-\infty}^{t} K_m\left(\frac{g(U_i)-u}{h_m}\right)du$$

is in consequence an consistent estimate of the function g^{-1} at the point $t \in \mathrm{Im}(g)$ (taken as a distribution function in some sense). An estimate for g is obtained by reversing the above estimate, which is strictly increasing with respect to t for $t \in \mathrm{Im}(g)$. This is the kernel interpretation of the monotonizing method.

On the other hand, for a strictly monotone function g on $[0,1]$ the inverse can be calculated simply by integrating

$$g^{-1}(t) = \int_0^1 I\{g(y) \leq t\}dy.$$

In general for an arbitrary function g measurable on $[0,1]$ with respect to the Lebesgue measure,

$$g_I^{-1}(t) = \int_0^1 I\{g(y) \leq t\}dy$$

is called the distribution of g for $t \in \mathrm{Im}(g)$. The increasing rearrangement

$$g_I(t) = \inf_{g_I^{-1}(y)\geq t} y$$

has the same distribution function as g, which is basically the generalized inverse of the distribution function g_I^{-1}. The function g_I^{-1} is non-decreasing and continuous to the right. Roughly speaking, stretches of constancy of a continuous g correspond to discontinuities of g_I^{-1}. Two functions with the same distribution function are called equidistributed. In the literature this approach is called measure preserving rearrangement [see Ryff (1970), Bennett and Sharpley (1988), or Zygmund (2002)]. This glance at the monotonizing procedure allows to grasp this approach as shifting the measure of a function to the right. In the following, we illustrate the idea by examples.

Example 1.7 (i) In the first example, we analyze the function

$$g(x) = \frac{1}{2}I_{[0,\frac{1}{3})}(x) + \frac{1}{4}I_{[\frac{1}{3},\frac{2}{3})}(x) + \frac{3}{4}I_{[\frac{2}{3},1]}(x),$$

which is a piecewise constant function. Note that this function is not continuous. The distribution function g_I^{-1} is discontinuous in 0.25, 0.5, and 0.75. Outside of $\mathrm{Im}(g) = \{0.25, 0.5, 0.75\}$, the distribution function g_I^{-1} is constant. In the following chapter, we use continuous functions and estimates only in order to obtain a *strictly* monotone estimate. Figure 1.5 shows the function g, its distribution function g_I^{-1}, and the increasing rearrangement g_I. It is easy to see that the blocks are just shifted in the right order to obtain an increasing rearrangement.

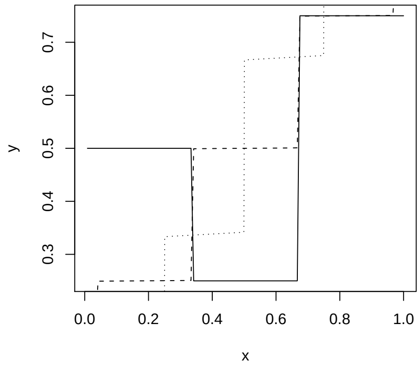

Figure 1.5: The function g, its distribution function g_I^{-1} (dotted line), the increasing rearrangement g_I (dashed line).

(ii) The decreasing rearrangement is constructed as the inverse of the corresponding non-increasing distribution function

$$g_A^{-1}(t) = \int_0^1 I\{g(y) > t\}dy.$$

In Figure 1.6, the function $g(x) = 4(x - \frac{1}{2})^2$ is displayed with its increasing and

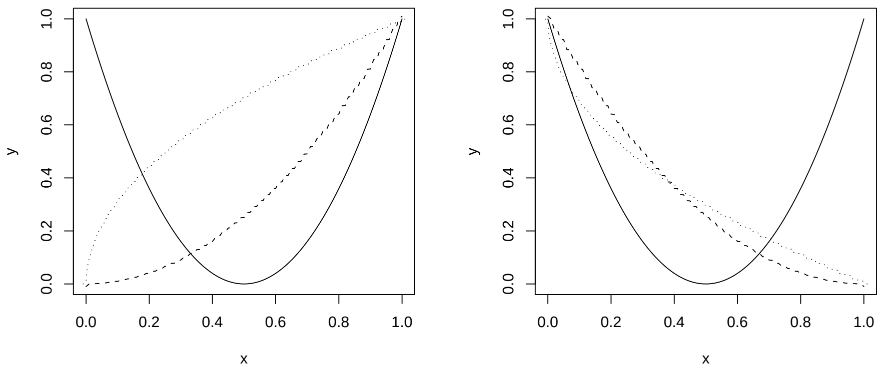

Figure 1.6: Left panel: The function g, its distribution function g_I^{-1} (dotted line), the increasing rearrangement g_I (dashed line). Right panel: The function g, its distribution function g_A^{-1} (dotted line), and the decreasing rearrangement g_A.

decreasing rearrangements.

(iii) The last example shows a more elaborate function

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

with its increasing rearrangement. Figure 1.7 displays the results.

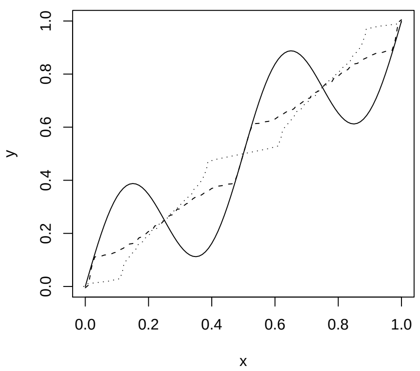

Figure 1.7: The function g, its distribution function g_I^{-1} (dotted line), the increasing rearrangement g_I (dashed line).

Since the indicator function I is not continuous, we approximate it continuously by an integrated kernel function

$$I\{g(y) \le t\} \approx \int_{-\infty}^{t} \frac{1}{h_m} K_m\left(\frac{g(y)-u}{h_m}\right) du,$$

where K_m is a positive kernel with the bandwidth h_m. For $h_m \to 0$, the approximation can be justified. For a continuous function g on the compact interval $[0,1]$, the right hand side of the above expression is continuous for $t \in \mathrm{Im}(g)$. This means that discontinuities of the left side are approximated continuously by the integrated kernel function. Finally, to simplify the expression computationally the outer integral is approximated by a sum. We obtain

$$\int_0^1 I\{g(y) \le t\} dy \approx \frac{1}{N}\sum_{i=1}^{N} \int_{-\infty}^{t} \frac{1}{h_m} K_m\left(\frac{g(i/N)-u}{h_m}\right) du.$$

If $h_m \to 0$ and $N \to \infty$ sufficiently fast, it seems clear that both sides behave similarly. To investigate the differences, we discuss Example 1.7 (i) again. See Figure 1.8, which displays the distribution function g_I^{-1} and the increasing rearrangement of the function g from Example 1.7 (i) using the approximation by an integrated kernel function. We used a uniform kernel and the bandwidth $h_m = 0.05$. Note that the discontinuities in Example 1.7 (i) are approximated by a straight line with a positive slope determined by the bandwidth $h_m > 0$ and the length of the stretches of constancy.

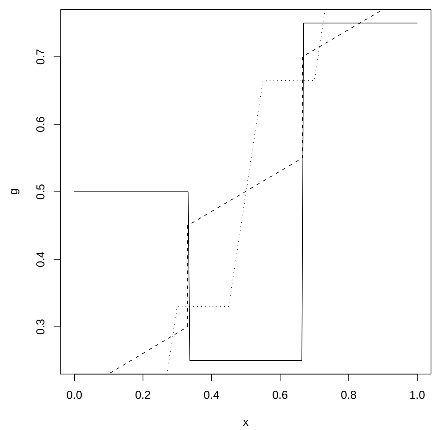

Figure 1.8: The function g, its continuously approximated distribution function g_I^{-1} (dotted line), the increasing rearrangement g_I (dashed line).

1.3.2 Monotonizing Procedure

In order to fix ideas, we illustrate the application of this procedure to construct a strictly monotone estimate for a regression function. Let $g : [0,1] \to I\!R$ denote an arbitrary continuous function or a continuous estimate. We define the smooth monotonized inverse by

$$g_I^{-1}(t, h_m) = \frac{1}{Nh_m} \sum_{i=1}^{N} \int_{-\infty}^{t} K_m \left(\frac{g(\frac{i}{N}) - u}{h_m} \right) du,$$

where K_m is a positive twice continuously differentiable, symmetric kernel supported on $[-1, 1]$ and $t \in \mathrm{Im}(g)$. The bandwidth h_m satisfies $h_m \to 0$ with increasing N. Further, we denote the inverse of $g_I^{-1}(t, h_m)$ as $g_I(t, h_m)$. This inverse is easy obtainable for $t \in \mathrm{Im}(g)$ since $g_I^{-1}(t, h_m)$ is a strictly increasing function as the kernel K_m is positive and continuous. Accordingly, the strictly decreasing transformation is obtained as the inverse of

$$g_A^{-1}(t, h_m) = \frac{1}{Nh_m} \sum_{i=1}^{N} \int_{t}^{\infty} K_m \left(\frac{g(\frac{i}{N}) - u}{h_m} \right) du$$

for $t \in \mathrm{Im}(g)$. To keep things simple, we restrict ourselves to the case of increasing rearrangements, but the results can be carried over accordingly to the case of a strictly decreasing rearrangements. As we have seen in the last subsection, the crucial point to get a *strictly* increasing rearrangement is to assume continuity of the unconstrained function and to apply the monotonzing inversion only for values $t \in \mathrm{Im}(g)$. Recapitulatory, this monotonizing procedure is a two-step procedure applied on a given continuous function or estimate g.

Step 1 *Isotonization*

Assess

$$g_I^{-1}(t, h_m) = \frac{1}{Nh_m} \sum_{i=1}^{N} \int_{-\infty}^{t} K_m \left(\frac{g(\frac{i}{N}) - u}{h_m} \right) du,$$

which is a strictly increasing function with respect to t on $\mathrm{Im}(g)$ provided N is large enough.

Step 2 *Inversion*

The inverse of $g_I^{-1}(t, h_m)$ is calculated and denoted by $g_I(t, h_m)$. This function is strictly increasing.

Next we present an important result of this smoothed measure preserving rearrangement.

Lemma 1.8 *If the function g is strictly increasing and twice continuously differentiable, we have for any $t \in \mathrm{Im}(g)$ with $g'(g^{-1}(t)) > 0$*

$$\begin{aligned} g_I^{-1}(t, h_m) &= g^{-1}(t) + \kappa_2(K_m) h_m^2 (g^{-1})''(t) + o(h_m^2) + O\left(\frac{1}{Nh_m}\right), \\ g_I(t, h_m) &= g(t) + \kappa_2(K_m) h_m^2 \frac{g''(t)}{(g'(t))^2} + o(h_m^2) + O\left(\frac{1}{Nh_m}\right), \end{aligned} \tag{1.8}$$

where the constant $\kappa_2(K_m)$ is defined in (1.7).

The proof of this result can be found in Dette et al. (2006). To prove the second statement (1.8), the operator which maps a non-decreasing function g to its "quantile" $g^{-1}(t)$, has to be examined carefully. We briefly describe this operator to make the proofs in the next chapter more comprehensible. For fixed $t \in \mathbb{R}$ and for some open set $D \subseteq \mathbb{R}$ let $\mathcal{M}$ denote the set of all twice continuously differentiable functions of the form $g : D \to \mathbb{R}$, which contain t in the interior of $g(D)$, and have a positive derivative in a neighborhood of the point $g^{-1}(t)$. Then we consider the functional

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

and define for $g_1, g_2 \in \mathcal{M}$ the function

$$Q : \begin{cases} [0,1] \to I\!R \\ \lambda \to \Phi(g_1 + \lambda(g_2 - g_1)) \end{cases} .$$

If $Q'(0)$ exists, it is the Gâteaux derivative of the functional Φ at g_1 in the direction of $g_2 - g_1$. In particular, Q is twice continuously differentiable with

$$Q'(\lambda) = -\frac{(g_2 - g_1)}{g_1' + \lambda(g_2' - g_1')} \circ (g_1 + \lambda(g_2 - g_1))^{-1}(t) \tag{1.9}$$

$$Q''(\lambda) = Q'(\lambda) \left\{ \frac{-2(g_2' - g_1')}{g_1' + \lambda(g_2' - g_1')} + \frac{(g_2 - g_1)(g_1'' + \lambda(g_2'' - g_1''))}{(g_1' + \lambda(g_2' - g_1'))^2} \right\} \circ Q(\lambda), \tag{1.10}$$

where g_1', g_2', and g_1'', g_2'' are the first and the second derivatives of g_1 and g_2, respectively. The symbol $\circ$ denotes the composition of functions, i.e., $(f \circ g)(x) = f(g(x))$. To apply this result, consider $g_1 = g^{-1}$ and $g_2 = g_I^{-1}$. For $t \in (0,1)$, we define the quantity $\epsilon = \min\left\{ \frac{g(t)-g(0)}{2}, \frac{g(1)-g(t)}{2} \right\} > 0$ and the set $D = (g(t) - \epsilon, g(t) + \epsilon)$. Since the support of g_I^{-1} contains the set D, we have the following relationship through a Taylor expansion of Q

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

for an appropriate $\lambda^* \in [0,1]$ with

$$Q'(\lambda^*) = -\frac{(g_I^{-1} - g^{-1})}{(g^{-1})' + \lambda^*((g_I^{-1})' - (g^{-1})')} \circ (g^{-1} + \lambda^*(g_I^{-1} - g^{-1}))^{-1}(t).$$

With this construction the difference $g_I(t) - g(t)$ can be expressed in terms of g_I^{-1} and g^{-1}. For g_I^{-1} an explicit expression is available and we obtain an functioning term for g_I to work with. This is the essential step for all assertion about properties of the monotonized function g_I.

Chapter 2

Multivariate monotone regression

2.1 Overview

In this chapter, the nonparametric regression setting with a one-dimensional covariate is extended to higher dimensional covariates. In Section 2.2, kernel regression estimators are introduced for a d-dimensional independent variable $\mathbf{X}$. The corresponding asymptotic properties in this context are summarized as well. Section 2.3 deals with the problem of estimating a multivariate regression function under monotonicity constraints. So far, partly due to computational difficulties multivariate predicators have not been considered in regression models under certain shape constraints as much as univariate ones. We propose a strictly monotone regression estimate in d-dimensions which starts with an unconstrained nonparametric regression estimate and uses successively the one-dimensional monotonizing procedure from Chapter 1. In the case of a strictly monotone regression function, it is shown that the new estimate is first order asymptotic equivalent to the unconstrained kernel estimator, and asymptotic normality of an appropriate standardization of the estimate is established. In the last section, the methodology is also illustrated by means of a simulation study, and two data examples are analyzed.

2.2 Multivariate Regression Estimation

In this section, we extend nonparametric regression models to higher dimensions. In the previous chapter, the theory for a one-dimensional predictor was developed. For a finite set $\{(\mathbf{X}_j, Y_j)\}_{j=1}^n$ of i.i.d. observations, we consider the following model

$$Y_j = m(\mathbf{X}_j) + \sigma(\mathbf{X}_j)\varepsilon_j, \;\; j = 1, \ldots, n \tag{2.1}$$

where $\mathbf{X}_j = (X_{j1}, \ldots, X_{jd})^T$ is a d-dimensional vector of independent variables. The functions m and σ are the unknown regression and variance function. ε denotes a random error with $E[\varepsilon_j] = 0$ and $E[\varepsilon_j^2] = 1$, such that the regression function is defined as the conditional expectation

$$E[Y_j|\mathbf{X}_j = \mathbf{x}] = m(\mathbf{x}) = m(x_1, \dots, x_d),$$

where $\mathbf{x} = (x_1, \dots, x_d)$. In the last chapter, nonparametric kernel estimators were introduced. Let K be a d-dimensional kernel function of order 2, this means that $\int K(\mathbf{u})d\mathbf{u} = 1$ and

$$\frac{1}{2}\int \mathbf{u}\mathbf{u}^T K(\mathbf{u})d\mathbf{u} = I_d\kappa_2(K)$$

with I_d as the $d \times d$ identity matrix and $\kappa_2(K) \geq 0$. Further we define

$$K_H(\mathbf{u}) = \frac{1}{\det(H)}K(H^{-1}\mathbf{u}),$$

where H is a nonsingular positive definite $d\times d$ matrix, the so called *bandwidth matrix*. The extension of the Nadaraya-Watson estimator to higher dimensions is straightforward as the weighted average with the d-dimensional kernel function

$$\hat{m}_{NW}(\mathbf{x}) = \frac{\sum_{j=1}^n K_H(\mathbf{x} - \mathbf{X}_j)Y_j}{\sum_{j=1}^n K_H(\mathbf{x} - \mathbf{X}_j)}.$$

Analogously, local linear estimators are defined by solving the minimization problem

$$\min_{\beta_0,\beta_1,\dots,\beta_d} \sum_{j=1}^n \left\{ Y_j - \beta_0 - \sum_{k=1}^d \beta_k(X_{jk} - x_k) \right\}^2 K_H(\mathbf{x} - \mathbf{X}_j),$$

where $\beta_0 = m(\mathbf{x})$ and $\beta_k = \frac{\partial}{\partial x_k}m(\mathbf{x})$ for $k = 1, \dots, d$. The solution to this weighted least square problem can be written as

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_d)^T = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y} \tag{2.2}$$

using the notations

$$\mathbf{X} = \begin{pmatrix} 1 & (X_{11} - x_1) & \dots & (X_{1d} - x_d) \\ 1 & (X_{21} - x_1) & \dots & (X_{2d} - x_d) \\ \vdots & \vdots & & \vdots \\ 1 & (X_{n1} - x_1) & \dots & (X_{nd} - x_d) \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

and $\mathbf{W} = \mathrm{diag}(K_H(\mathbf{x} - \mathbf{X}_1), \dots, K_H(\mathbf{x} - \mathbf{X}_n))$. The multivariate local linear estimator is then given by

$$\hat{m}_{LL}(\mathbf{x}) = \hat{\beta}_0 \tag{2.3}$$

provided the invertibility of $\mathbf{X}^T\mathbf{W}\mathbf{X}$.

Example 2.1 For $d = 2$ and for fixed $\mathbf{x} = (x_1, x_2)^T$ an explicit formula for the local linear estimator can be derived. Denote

$$\begin{aligned} S_{lk}(\mathbf{x}) &= \sum_{j=1}^{n} K_H(\mathbf{x} - \mathbf{X}_j)(X_{j1} - x_1)^l (X_{j2} - x_2)^k, \\ T_{lk}(\mathbf{x}) &= \sum_{j=1}^{n} K_H(\mathbf{x} - \mathbf{X}_j)(X_{j1} - x_1)^l (X_{j2} - x_2)^k Y_j. \end{aligned}$$

The least square solution for the local linear estimation problem (2.2) becomes

$$\hat{\beta} = \begin{pmatrix} S_{00} & S_{10} & S_{01} \\ S_{10} & S_{20} & S_{11} \\ S_{01} & S_{11} & S_{02} \end{pmatrix}^{-1} \begin{pmatrix} T_{00} \\ T_{10} \\ T_{01} \end{pmatrix}.$$

More precisely, the local linear regression estimator is given by

$$\hat{m}_{LL}(\mathbf{x}) = \frac{(S_{20}S_{02} - S_{11}^2)T_{00} + (S_{10}S_{11} - S_{01}S_{20})T_{01} + (S_{01}S_{11} - S_{02}S_{10})T_{10}}{2S_{01}S_{10}S_{11} - S_{02}S_{10}^2 - S_{00}S_{11}^2 - S_{01}^2 S_{20} + S_{00}S_{02}S_{20}}.$$

2.2.1 Asymptotic Properties

In this subsection, we shortly give an overview of the statistical properties of the estimates $\hat{m}_{NW}(\mathbf{x})$ and $\hat{m}_{LL}(\mathbf{x})$ in the multivariate case. Similar results as in the one-dimensional case are stated. A more detailed derivation can be found in Ruppert and Wand (1994). The following model assumptions regarding (2.1) are needed and formulated in analogy to Chapter 1:

(A1) $\mathbf{X}_j$, $j = 1, \ldots, n$ is an i.i.d. sample with a twice continuously differentiable positive density, say p, supported on $[0, 1]^d$. The gradient of p is denoted by ∇p.

(A2) $\sigma : [0, 1]^d \to I\!R^+$ is a continuous variance function.

(A3) The regression function $m : [0, 1]^d \to I\!R$ is twice continuously differentiable with respect to all arguments. We denote the gradient of m as ∇m and the Hessian matrix of m as $\mathcal{H}(m)$.

(A4) The random error ε_j and $\mathbf{X}_j$ are independent, $E[\varepsilon_j] = 0$, and $E[\varepsilon_j^2] = 1$ for $j = 1, \ldots, n$. For an asymptotic normality result, we further assume $E[\varepsilon_j^4] < \infty$ for $j = 1, \ldots, n$.

(A5) K is a symmetric d-dimensional kernel of order 2 (see above) compactly supported on $[-1, 1]^d$.

(A6) All entries of the bandwidth matrix H and $n^{-1}\det(H)$ tend to zero as $n \to \infty$, whereas H remains symmetric and positive definite. In addition, there exists a fixed constant L such that the ratio of its largest to its smallest eigenvalue is at most L for all n.

Theorem 2.2 *Assume that the assumptions (A1)-(A6) are satisfied. Let* $\mathbf{x}$ *be a fixed point in the interior of* $\operatorname{supp}(p)$.

(i) Then for the Nadaraya-Watson estimate $\hat{m}_{NW}(\mathbf{x})$*, we obtain*

$$\begin{aligned} E\left[\hat{m}_{NW}(\mathbf{x}) - m(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] &= \kappa_2(K)\biggl(2\frac{\nabla m(\mathbf{x})^T HH^T \nabla p(\mathbf{x})}{p(\mathbf{x})} \\ &\qquad + \operatorname{tr}(H^T\mathcal{H}(m)H)\biggr) + o_p(H^TH) \\ \operatorname{Var}(\hat{m}_{NW}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n) &= \frac{(1+o_p(1))}{n\det(H)}||K||_2^2\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})}. \end{aligned}$$

(ii) Analogously, for the local linear estimate $\hat{m}_{LL}(\mathbf{x})$ *we have*

$$\begin{aligned} E\left[\hat{m}_{LL}(\mathbf{x}) - m(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] &= \kappa_2(K)\operatorname{tr}(H^T\mathcal{H}(m)H) + o_p(H^TH) \\ \operatorname{Var}(\hat{m}_{LL}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n) &= \frac{(1+o_p(1))}{n\det(H)}||K||_2^2\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})}. \end{aligned}$$

Remark 2.3 The leading terms of the conditional bias and variance of the Nadaraya-Watson estimator and the local linear one in the above theorem do not depend on $\mathbf{X}_1,\ldots,\mathbf{X}_n$. In particular, under some additional conditions $\hat{m}_{NW}(\mathbf{x})$ and $\hat{m}_{LL}(\mathbf{x})$ are asymptotically normal as the univariate estimates [see Theorem 1.3] with asymptotic bias and variance given in Theorem 2.2.

Remark 2.4 In this chapter, we assume that the kernel K is of order 2. It is possible to extend the results for higher order kernels. On the other hand, further smoothness conditions on m are necessary for kernels of order $q > 2$ [see Chapter 1]. In Chapter 3, we apply higher order multivariate kernels in the context of additive quantile models to obtain the optimal convergence rate.

The asymptotic expressions for the conditional bias and variance in Theorem 2.2 for the Nadaraya-Watson estimator and the local linear estimate, respectively, yield the following asymptotic expression for the conditional mean squared error

$$\begin{aligned} \mathrm{AMSE}(\hat{m}_{NW}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n) &= \frac{1}{n\det(H)}||K||_2^2\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})} \\ &\quad +\kappa_2^2(K)\left(2\frac{\nabla m(\mathbf{x})^T HH^T\nabla p(\mathbf{x})}{p(\mathbf{x})}+\mathrm{tr}(H^T\mathcal{H}(m)H)\right)^2 \end{aligned} \tag{2.4}$$

$$\mathrm{AMSE}(\hat{m}_{LL}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n) = \frac{1}{n\det(H)}||K||_2^2\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})}+\kappa_2^2(K)\left(\mathrm{tr}(H^T\mathcal{H}(m)H)\right)^2 \tag{2.5}$$

The optimal choice of the bandwidth matrix H for the estimate $\hat{m}_{NW}(\mathbf{x})$ or $\hat{m}_{LL}(\mathbf{x})$ would be a matrix which minimizes (2.4) and (2.5), respectively. We simplify these problem and consider the bandwidth matrix $H = h\cdot I_d$, where I_d is the $d\times d$ identity matrix. This means that all components are treated with the same bandwidth h. In the case of the Nadaraya-Watson estimate, we obtain

$$\frac{1}{nh^d}||K||_2^2\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})}+h^4\kappa_2^2(K)\left(2\frac{\nabla m(\mathbf{x})^T\nabla p(\mathbf{x})}{p(\mathbf{x})}+\mathrm{tr}(\mathcal{H}(m))\right)^2.$$

Differentiating leads to

$$-\frac{1}{nh^{d+1}}||K||_2^2\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})}+4h^3\kappa_2^2(K)\left(2\frac{\nabla m(\mathbf{x})^T\nabla p(\mathbf{x})}{p(\mathbf{x})}+\mathrm{tr}(\mathcal{H}(m))\right)^2=0,$$

so that the optimal bandwidth h^*_{NW} has the following form

$$h^*_{NW}=\left(\frac{||K||_2^2\sigma^2(\mathbf{x})p(\mathbf{x})}{4n\kappa_2^2(K)(2\nabla m(\mathbf{x})^T\nabla p(\mathbf{x})+p(\mathbf{x})\mathrm{tr}(\mathcal{H}(m)))^2}\right)^{1/(d+4)}.$$

The optimal bandwidth h depends on n, K, σ, p, and m and converges at the rate $n^{-1/(d+4)}$. For the asymptotic mean squared error, this implies a convergence rate of $n^{-4/(d+4)}$. The larger d the slower the speed of convergence, this melts down to the so called *curse of dimensionality*. The idea of averaging locally requires enough observations in each local window which is problematic in higher dimensions due to data sparseness.

Example 2.5 Figure 2.1 shows the Nadaraya-Watson estimate for the two-dimensional regression function

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

on the square $[0,1]^2$. We simulated 200 observations uniformly distributed, with $\sigma(x,y)=0.5$, and $\varepsilon\sim\mathcal{N}(0,1)$. We used two different bandwidths to demonstrate the phenomena of over- and under-smoothing.

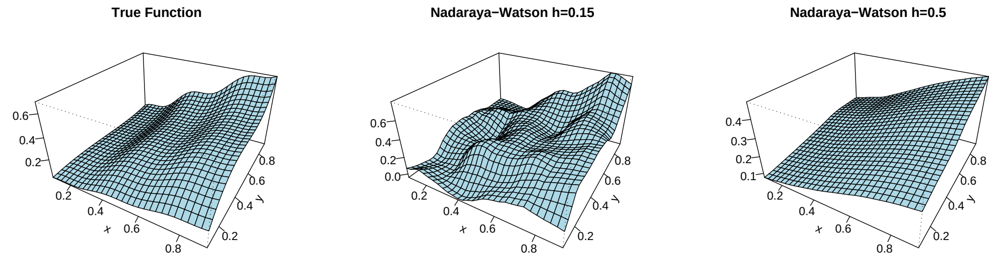

Figure 2.1: Two-dimensional Nadaraya-Watson estimates with different bandwidths.

To close this subsection, a strong uniform convergence result for the Nadaraya-Watson estimator in higher dimensions is presented at the rate

$$\theta_n = \left(\frac{\log n}{n\, h^d}\right)^{1/2},$$

where again $H = h \cdot I_d$, which will be used in the next section. Further details on this result can be found in Collomb and Härdle (1986).

Theorem 2.6 *Suppose $\theta_n^{-1} h^2 \leq C < \infty$ for all $n \in I\!N$, Y is bounded, the second derivatives of p and $m \cdot p$ are uniformly bounded. Then we have for all compact intervals J, where p is bounded below by a strictly positive number,*

$$\theta_n^{-1} \sup_{x \in J} |\hat{m}_{NW}(\mathbf{x}) - m(\mathbf{x})| = O(1) \;\; w.p.1.$$

2.3 Multivariate monotone Regression Estimates

In the previous section, nonparametric methods to estimate a regression function in several variables were introduced. In some settings, there are restrictions that the relationship between a dependent random variable Y and an explanatory variable $\mathbf{X}$ is order-preserving. In contrast to linear regression, where a particular functional form of the relationship is assumed, our aim is to estimate a regression function under monotonicity constraints with no further specifications except for smoothness conditions. Typical examples appear in economics where curvature and monotonicity apply to indirect utility, expenditure, production, profit, and cost function [Gallant and Golub (1984), Matzkin (1986), or Aït-Sahalia and Duarte (2003) among others], or in medicine where the probability of contracting a certain disease, say cancer, depends monotonically on certain factors like smoking frequency, drinking frequency, or weight [see, e.g., Hall et al. (2001)].

In the following, we propose a new strictly monotone estimate for higher dimensions. In univariate settings, there are quite a few monotone estimates available. This is different in multivariate regression setups. Because of the computational difficulties in higher dimensions, several authors propose alternative methodologies such as additive isotonic regression [see Bacchetti (1989)] or hybrid-type estimators [see Mukarjee and Stern (1994)].
Since $[0,1]^d \subset I\!R^d$ cannot be totally ordered, it is worth to give some thoughts to the definition of monotonicity in higher dimensions. For $\mathbf{x}$ and $\mathbf{x}'$ with $\mathbf{x} \le \mathbf{x}'$, where $\mathbf{x} \le \mathbf{x}'$ means that $x_k \le x'_k$ for all $k = 1, \ldots, d$, a function m is called monotone increasing if $m(\mathbf{x}) \le m(\mathbf{x}')$. Equivalently, we can define a strictly monotone function m with d arguments as a function which is strictly monotone in each argument. This definition allows a meaningful interpretation for a monotone function in higher dimensions. In regression models, the variables usually correspond to certain influencing values, e.g. we consider the body fat of a person as the response variable depending on the height and the weight as the independent variables. We suspect that the body fat percentage of a person is higher the fatter and lower the taller a person is. This is an example of a strictly monotone two-dimensional regression model. The regression function m is supposed to be strictly monotone increasing with respect to weight and strictly monotone decreasing with respect to height. This quite simple example describes in a more informal way what we mean by strictly monotone regression in higher dimensions.
In the following, we consider the model (2.1) under additional monotonicity constraints. The i.i.d. sample $\{(\mathbf{X}_j, Y_j)\}_{j=1}^n$ with $\mathbf{X}_j = (X_{j1}, \ldots, X_{jd})^T$ are realizations from the model:

$$\begin{aligned} Y_j &= m(\mathbf{X}_j) + \sigma(\mathbf{X}_j)\varepsilon_j \\ &= m(X_{j1}, \ldots, X_{jd}) + \sigma(X_{j1}, \ldots, X_{jd})\varepsilon_j \qquad (j = 1, \ldots, N), \end{aligned} \tag{2.6}$$

where m is assumed to be strictly monotone with respect to each variable. For simplicity, we suppose that m is strictly increasing for each variable $X_1, \ldots, X_d$. In the next subsection, we introduce a strictly monotone estimate for higher dimensions using the monotonizing procedure presented in Chapter 1.

2.3.1 The strictly monotone estimate

The idea, we will utilize to construct a strictly monotone nonparametric regression estimate, is applying the monotonizing procedure introduced in Chapter 1 stepwise. We start with an unconstrained nonparametric regression estimate $\hat{m}(\mathbf{x})$ for $m(\mathbf{x})$ like the Nadaraya and Watson estimator $\hat{m}_{NW}(\mathbf{x})$ or the local linear estimate $\hat{m}_{LL}(\mathbf{x})$. Then the function is monotonized step by step for each variable. To clarify this approach, let $\hat{m}(x_1, \ldots, x_d)$ be a continuous nonparametric estimate of the regression function. We denote $x_{\underline{k}} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_k)^T$ the vector of all variables but x_k. The kernel K_m is positive and supported on $[-1, 1]$ with corresponding bandwidth h_m. Furthermore, the kernel K_m is twice continuously differentiable and K'_m is Lipschitz continuous. Note that for a strictly increasing function in the kth argument $m_k^{-1}(z|x_{\underline{k}})$ is the inverse of m with respect to x_k where $x_{\underline{k}}$ is fixed. Moreover, $\frac{\partial}{\partial x_k} m(x_k, x_{\underline{k}})$ indicates the partial derivative regarding the kth argument. Using these notations, the monotonizing procedure is applied as follows.

Step 1 *Isotonization with respect to the first coordinate*
For fixed $x_{\underline{1}} \in (0,1)^{d-1}$, define

$$\hat{m}_{I_1}^{-1}(z|x_{\underline{1}}) = \frac{1}{Nh_m} \sum_{i=1}^{N} \int_{-\infty}^{z} K_m \left(\frac{\hat{m}(\frac{i}{N}, x_{\underline{1}}) - u}{h_m} \right) du,$$

where $\hat{m}$ is an unconstrained estimator for m. $\hat{m}_{I_1}^{-1}(z|x_{\underline{1}})$ is a strictly increasing function of z provided N is large enough on $\mathrm{Im}(m(\cdot, x_{\underline{1}}))$.

Step 2 *Inversion with respect to the first coordinate*
For fixed $x_{\underline{1}}$, the inverse of the strictly increasing function $\hat{m}_{I_1}^{-1}(z|x_{\underline{1}})$ is computed and denoted by $\hat{m}_{I_1}(x_1, x_{\underline{1}})$. This function is strictly increasing in x_1.

Step 3 *Isotonization with respect to the second coordinate*
Now we fix $x_{\underline{2}} \in (0,1)^{d-1}$ and define

$$\hat{m}_{I_{1,2}}^{-1}(z|x_{\underline{2}}) = \frac{1}{Nh_m} \sum_{i=1}^{N} \int_{-\infty}^{z} K_m \left(\frac{\hat{m}_{I_1}(\frac{i}{N}, x_{\underline{2}}) - u}{h_m} \right) du,$$

which is again a strictly increasing function in z for fixed $x_{\underline{2}}$.

Step 4 *Inversion with respect to the second coordinate*
For fixed $x_{\underline{2}} \in (0,1)^{d-1}$, the inverse of $\hat{m}_{I_{1,2}}^{-1}(z|x_{\underline{2}})$ is calculated and denoted by $\hat{m}_{I_{1,2}}(x_2, x_{\underline{2}})$.

and so on until the dth coordinate.

Step 2d *Inversion with respect to the last coordinate*
The estimate after the dth isotonization is denoted by $\hat{m}_{I_{1,\ldots,d}}(x_1, \ldots, x_d)$ which is again defined as the inverse of $\hat{m}_{I_{1,\ldots,d}}^{-1}$.

It is possible to monotonize just a few directions, e.g. to stop after k monotonizations. Or it is possible to isotonize some directions and to antitonize others by using

$$\hat{m}_{A_k}^{-1}(z|x_{\underline{k}}) = \frac{1}{Nh_m} \sum_{i=1}^{N} \int_{z}^{\infty} K_m \left(\frac{\hat{m}(\frac{i}{N}, x_{\underline{k}}) - u}{h_m} \right) du$$

as the antitone version of $\hat{m}_{I_k}^{-1}$ to obtain a strictly decreasing estimate in the kth direction. The inverse of $\hat{m}_{A_k}^{-1}(z|x_{\underline{k}})$ denoted by $\hat{m}_{A_k}(x_k, x_{\underline{k}})$ is strictly decreasing with respect

to x_k. To be precise, we indicate the isotonization by the subscript I versus A for the antitone counterpart. The subscripts of I and A, respectively, name the monotonized variables. The crucial point of this stepwise construction is that the monotonicity of the estimate $\hat{m}_{I_1}(x_1, x_{\underline{1}})$ in x_1 is not destroyed in the next monotonizing step. The following two lemmas cover this problem and show that one can use this univariate monotonizing procedure step by step for each variable.
The first Lemma analyzes Step 1 of the monotonizing procedure with respect to the first coordinate. By exchanging the order of the coordinates, this result can be applied with respect to each coordinate.

Lemma 2.7 *Assume that the preliminary continuous estimate $\hat{m}(x_1, x_2, \ldots, x_d)$ is strictly increasing in x_2, then for sufficiently large N the estimate $\hat{m}_{I_1}^{-1}(z|x_2, \ldots, x_d)$ obtained after the first step of the monotonizing procedure is*

(i) strictly increasing in $z \in \mathrm{Im}(\hat{m}(\cdot, x_{\underline{1}}))$ for fixed $x_{\underline{1}}$;

(ii) strictly decreasing in x_2 for fixed $z \in \mathrm{Im}(\hat{m}(\cdot, x_{\underline{1}}))$ and fixed $x_3, \ldots, x_d$.

Proof (i) Fix $z < z'$ and $x_{\underline{1}}$ with $z, z' \in \mathrm{Im}(\hat{m}(\cdot, x_{\underline{1}}))$ and we have

$$\begin{aligned}
&\hat{m}_{I_1}^{-1}(z'|x_{\underline{1}}) - \hat{m}_{I_1}^{-1}(z|x_{\underline{1}}) \\
= &\ \frac{1}{Nh_m}\sum_{i=1}^{N}\left(\int_{-\infty}^{z'} K_m\left(\frac{\hat{m}(\frac{i}{N}, x_{\underline{1}}) - u}{h_m}\right) du - \int_{-\infty}^{z} K_m\left(\frac{\hat{m}(\frac{i}{N}, x_{\underline{1}}) - u}{h_m}\right) du\right) \\
= &\ \frac{1}{Nh_m}\sum_{i=1}^{N}\int_{z}^{z'} K_m\left(\frac{\hat{m}(\frac{i}{N}, x_{\underline{1}}) - u}{h_m}\right) du > 0.
\end{aligned}$$

The last inequality is true, since we can choose N as large as at least one $j \in \{1, \ldots, N\}$ exists with $\left|\frac{\hat{m}(\frac{j}{N}, x_{\underline{1}}) - u}{h_m}\right| < 1$ for $u \in (z, z')$. Recall $z, z' \in \mathrm{Im}(\hat{m}(\cdot, x_{\underline{1}}))$. For j with $\left|\frac{\hat{m}(\frac{j}{N}, x_{\underline{1}}) - u}{h_m}\right| < 1$, we get using the positivity and the continuity of K_m

$$\int_{z}^{z'} K_m\left(\frac{\hat{m}(\frac{j}{N}, x_{\underline{1}}) - u}{h_m}\right) du > 0.$$

This concludes the assertion of (i).

(ii) This time, we fix $z \in \mathrm{Im}(\hat{m}(\cdot, x_{\underline{1}}))$, $x_3, \ldots, x_d$ and $x_2 < x_2'$ and obtain for the difference

$$
\begin{aligned}
& \hat{m}_{I_1}^{-1}(z|x_2', x_3, \ldots, x_d) - \hat{m}_{I_1}^{-1}(z|x_2, x_3, \ldots, x_d) \\
= \; & \frac{1}{Nh_m} \sum_{i=1}^{N} \Bigg(\int_{-\infty}^{z} K_m \left(\frac{\hat{m}(\frac{i}{N}, x_2', x_3, \ldots, x_d) - u}{h_m} \right) du \\
& - \int_{-\infty}^{z} K_m \left(\frac{\hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d) - u}{h_m} \right) du \Bigg) \\
= \; & \frac{1}{Nh_m} \sum_{i=1}^{N} \Bigg(\int_{-\infty}^{z-(\hat{m}(\frac{i}{N}, x_2', x_3, \ldots, x_d) - \hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d))} K_m \left(\frac{\hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d) - u}{h_m} \right) du \\
& - \int_{-\infty}^{z} K_m \left(\frac{\hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d) - u}{h_m} \right) du \Bigg) \\
= \; & -\frac{1}{Nh_m} \sum_{i=1}^{N} \int_{z-(\hat{m}(\frac{i}{N}, x_2', x_3, \ldots, x_d) - \hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d))}^{z} K_m \left(\frac{\hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d) - u}{h_m} \right) du < 0
\end{aligned}
$$

For we second equation, we used the following substitution

$$
u \rightarrow (\hat{m}(\frac{i}{N}, x_2', x_3, \ldots, x_d) - \hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d)) + u
$$

to get the same arguments in the kernel function K_m. Since $\hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d)$ is supposed to be strictly increasing in its second argument, we have

$$
\hat{m}(\frac{i}{N}, x_2', x_3, \ldots, x_d) - \hat{m}(\frac{i}{N}, x_2, x_3, \ldots, x_d) > 0.
$$

Again, we can choose N so large that there exists at least one j with

$$
K_m \left(\frac{\hat{m}(\frac{j}{N}, x_2, x_3, \ldots, x_d) - u}{h_m} \right) > 0.
$$

This completes the proof of the lemma. □

The next lemma finally combines this result and makes sure that the monotonicity of a variable is not destroyed by monotonizing another variable through this approach. Therefore it is possible to apply the procedure stepwise for several covariates.

Lemma 2.8 *For sufficiently large N and a continuos estimate $\hat{m}$, the monotonized estimate $\hat{m}_{I_{1,\ldots,d}}(x_1, \ldots, x_d)$ is strictly increasing with respect to all d arguments.*

Proof To prove this assertion, we apply a mathematical induction on d and use Lemma 2.7. The basis for this induction is somehow clear from the construction of $\hat{m}_{I_1}(x_1)$ with a one-dimensional covariate. We obtain $\hat{m}_{I_1}(x_1)$ as the inverse of a strictly increasing function $\hat{m}_{I_1}^{-1}$ for sufficiently large N and a continuous estimate $\hat{m}$, and therefore $\hat{m}_{I_1}(x_1)$ is strictly increasing in x_1.
For the inductive step, we assume that the assertion of the Lemma is true for the case $d-1$. For fixed x_d, we understand $\hat{m}_{I_{1,\ldots,d-1}}(x_1,\ldots,x_{d-1},x_d)$ as a $d-1$-dimensional function monotonized in $x_1,\ldots,x_{d-1}$. By the induction hypothesis, this function is strictly increasing in all variables for a sufficiently large N but x_d. We apply the monotonizing inversion with respect to the argument x_d. By Lemma 2.7, we conclude that $\hat{m}_{I_{1,\ldots,d}}^{-1}(z|x_{\underline{d}})$ is strictly increasing with respect to $z \in \mathrm{Im}(\hat{m}_{I_{1,\ldots,d-1}}(x_{\underline{d}},\cdot))$ and strictly decreasing with respect to $x_1,\ldots,x_{d-1}$. Through the inversion with respect to z and fixed $x_{\underline{d}}$, we conserve the strict monotonicity of the variable x_d by construction, i.e. $\hat{m}_{I_{1,\ldots,d}}(x_1,\ldots,x_d)$ is strictly increasing in x_d. Suppose that the inverse of the dth monotonizing step, $\hat{m}_{I_{1,\ldots,d}}(x_1,\ldots,x_d)$, is not strictly increasing with respect to at least one variable x_k, $k \neq d$. Hence, we suppose there exists $x_k, x'_k \in (0,1)$ with $x_k < x'_k$ and

$$z_1 = \hat{m}_{I_{1,\ldots,d}}(x_1,\ldots,x_k,\ldots,x_d) \geq \hat{m}_{I_{1,\ldots,d}}(x_1,\ldots,x'_k,\ldots,x_d) = z_2$$

for $x_{\underline{k}} = (x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_d) \in (0,1)^{d-1}$. Reversing this assertion and applying Lemma 2.7 gives

$$\begin{aligned}
\hat{m}_{I_{1,\ldots,d}}^{-1}(z_1|x_1,\ldots,x_k,\ldots,x_{d-1}) &= x_d = \hat{m}_{I_{1,\ldots,d}}^{-1}(z_2|x_1,\ldots,x'_k,\ldots,x_{d-1}) \\
&\leq \hat{m}_{I_{1,\ldots,d}}^{-1}(z_1|x_1,\ldots,x'_k,\ldots,x_{d-1}) \\
&< m_{I_{1,\ldots,d}}^{-1}(z_1|x_1,\ldots,x_k,\ldots,x_{d-1}),
\end{aligned}$$

which yields a contradiction to the assumption

$$\hat{m}_{I_{1,\ldots,d}}(x_1,\ldots,x_k,\ldots,x_d) \geq \hat{m}_{I_{1,\ldots,d}}(x_1,\ldots,x'_k,\ldots,x_d)$$

for $x_k < x'_k$. Therefore $\hat{m}_{I_{1,\ldots,d}}$ is strictly increasing with respect to $x_1,\ldots,x_d$. □

Remark 2.9 We will show in the next subsection that the order of monotonization does not matter asymptotically. Note that for finite sample problems the estimates may differ slightly. In order to avoid this effect caused by the order of monotonization, the average estimator can be used

$$\hat{m}_I(x_1,\ldots,x_d) = \frac{1}{d!} \sum_{\sigma \in \mathrm{perm}(1,\ldots,d)} \hat{m}_{I_\sigma}(x_1,\ \ldots,x_d).$$

Remark 2.10 The behavior of the estimate $\hat{m}_{I_{1,\ldots,d}}$ depends sensitively on the monotonicity properties of the "true" regression function. If the regression function m in (2.6) is strictly increasing with respect to all d arguments, it is heuristically clear as remarked for the one-dimensional case in Section 1.3.1 that the estimate from the first step $\hat{m}_{I_1}^{-1}$ is a continuous approximation of the quantity

$$\begin{aligned} m_{I_1}^{-1}(z|x_{\underline{1}}) &= \frac{1}{Nh_m}\sum_{i=1}^{N}\int_{-\infty}^{z} K_m\left(\frac{m(\frac{i}{N}, x_{\underline{1}}) - u}{h_m}\right) du \\ &\approx \int_0^1 I\{m(x_1, x_{\underline{1}}) \leq z\} dx_1 =: m_1^{-1}(z|x_{\underline{1}}), \end{aligned}$$

where $z \in \mathrm{Im}(m(\cdot, x_{\underline{1}}))$ and the approximation is justified if $N \to \infty$, $h_m \to 0$ sufficiently fast. It follows that $m_1^{-1}(\cdot|x_{\underline{1}})$ is the inverse of $m(\cdot, x_{\underline{1}})$ for fixed $x_{\underline{1}} \in (0,1)^{d-1}$, and the inversion in Step 2 of the algorithm reproduces m. Applying the same arguments to the following $d-1$ isotonization steps, it is intuitively clear that $\hat{m}_{I_{1,\ldots,d}}$ is a consistent estimate of a strictly isotone regression function in d variables utilizing a continuous unconstrained estimate $\hat{m}$. These heuristic arguments will be made precise in the next section, where the asymptotic properties of the new estimate are discussed.

Example 2.11 To recall the example from the beginning of the section, the data set `fat` in the `UsingR` package in `R` contains the body fat percentage of 252 men from 22 to 81 and several physical measurements useful to predict the body fat of a person. Roughly motivated by the body mass index (BMI), we consider the nonparametric regression model where Y corresponds to the body fat percentage depending on weight (X_1) and height (X_2) as independent variables. It is reasonable that the regression function is strictly monotone increasing in weight and strictly monotone decreasing in height. Two outliers in the data set are removed since they blow up the estimate unreasonably. In Figure 2.11, the unconstrained local linear estimate and the monotonized estimate are displayed as perspective and contour plot. The local linear estimate is implemented in the `locfit` package and the strictly monotone estimate can be found in the `monoProc` package.

2.3.2 Asymptotic Properties

In the following, we discuss the asymptotic behavior of our estimator $\hat{m}_{I_{1,\ldots,d}}(x_1, \ldots, x_d)$. Therefore we have to specify our preliminary estimator. For the sake of simplicity, we use the multivariate Nadaraya and Watson estimate

$$\hat{m}_{NW}(\mathbf{x}) = \hat{m}_{NW}(x_1, \ldots, x_d) = \frac{\sum_{j=1}^n K_H(\mathbf{x} - \mathbf{X}_j)Y_j}{\sum_{j=1}^n K_H(\mathbf{x} - \mathbf{X}_j)}$$

as unconstrained estimator, where we set $H = \mathrm{diag}(h_1, \ldots, h_d)$. In particular, h_k is the corresponding bandwidth of the variable x_k for $k = 1, \ldots, d$. We will establish asymptotic normality under an appropriate standardization of the estimate.
Recall $m_k^{-1}(z|x_{\underline{k}})$ as the inverse of m with respect to the variable x_k, where the arguments

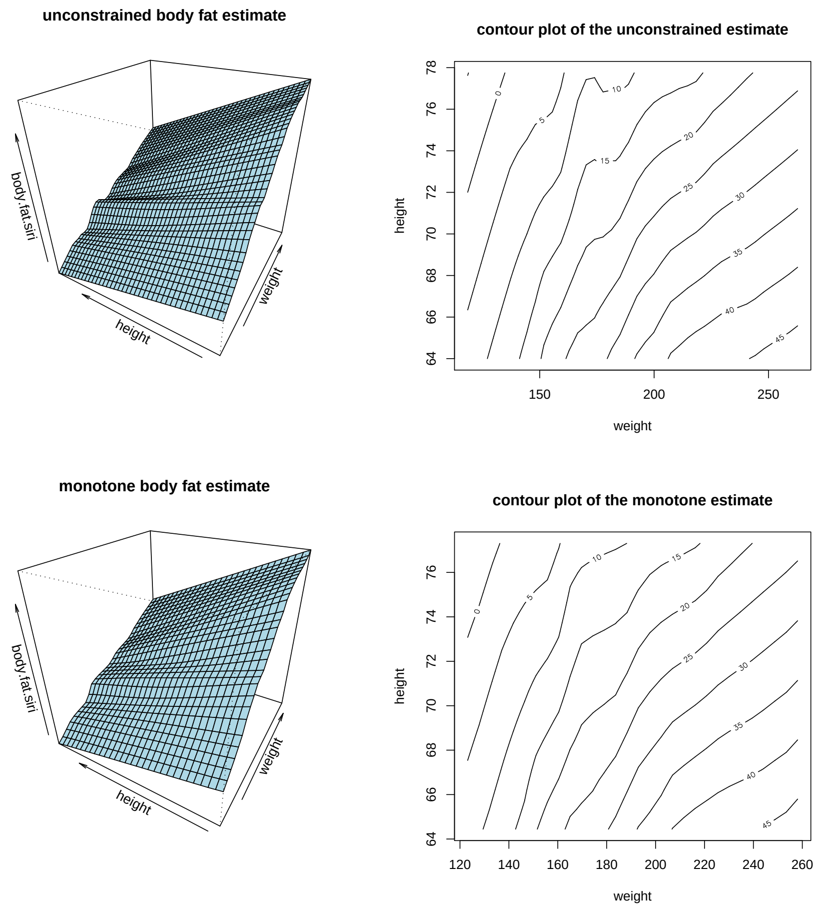

Figure 2.2: Body fat example: perspective and contour plots. Upper panel: the unconstrained regression estimate. Lower panel: the monotone estimate.

$x_{\underline{k}} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_d)$ are fixed.
Throughout this chapter, we assume that K_m is a positive kernel of order 2 on $[-1, 1]$, twice continuously differentiable, and K'_m is Lipschitz continuous. In addition to the model assumptions (A1)-(A6) of Section 2.2, we need several bandwidth conditions, that is

(B1) $h_k \to 0, \quad (k = 1, \ldots, d), \quad nh_1 \ldots h_d \to \infty$, $h_m \to 0$, and $Nh_m \to \infty$,

(B2) $n(h_1 + \ldots + h_d)^{d+4} = O(1), \quad n = O(N)$,

(B3) $\lim_{n\to\infty} \frac{h_k}{h_m} = \infty$ for $k = 1, \ldots, d$

(B4) $\frac{1}{nh_1 \ldots h_d h_m^2} = o(1)$

(B5) $\frac{\log(h_1+\ldots+h_d)^{-1}}{nh_1 \ldots h_d h_m^2} = o(1)$

The bandwidth condition (B4) is redundant since it is included in (B5). We mention this condition separately because for the first result condition (B4) is sufficient. Furthermore it is a slight improvement to the assumptions made in Dette et al. (2006). Note that $\kappa_2(K)$ is defined in condition (A5) of Chapter 1 and $||K||_2$ is the L_2 norm of the kernel function K.
In the following, we present several theorems concerning the asymptotic properties of the estimates of each monotonization step. The first result addresses the estimate $\hat{m}_{I_1}^{-1}(z|x_{\underline{1}})$ after the Step 1 of the isotonization with respect to the first argument.

Theorem 2.12 *Assume that the assumptions (A1)-(A6) and (B1)-(B4) are satisfied. If the regression function m is strictly increasing with respect to x_1, then it follows for any fixed $x_{\underline{1}}$ and for all $z \in (m(0, x_{\underline{1}}), m(1, x_{\underline{1}}))$ with $\frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}})|x_{\underline{1}}) > 0$*

$$\sqrt{nh_1 \ldots h_d}(\hat{m}_{I_1}^{-1}(z|x_{\underline{1}}) - m_1^{-1}(z|x_{\underline{1}}) + b_1(z, x_{\underline{1}})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_1^2(z, x_{\underline{1}})),$$

where the asymptotic bias and variance are given by

$$\begin{aligned} b_1(z, x_{\underline{1}}) &= \frac{\kappa_2(K)}{\frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}})|x_{\underline{1}})} \sum_{l=1}^{d} h_l^2 \Bigg[\frac{\partial^2}{\partial x_l^2} m(m_1^{-1}(z|x_{\underline{1}})|x_{\underline{1}}) \\ &\quad + 2 \frac{\frac{\partial}{\partial x_l} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) \frac{\partial}{\partial x_1} p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})}{p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} \Bigg], \\ s_1^2(z, x_{\underline{1}}) &= \frac{\sigma^2(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})}{p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) \left(\frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) \right)^2} ||K||_2^2. \end{aligned}$$

The next result gives a corresponding statement of asymptotic normality for the monotonized estimate $\hat{m}_{I_1}(x_1, x_{\underline{1}})$ with respect to the first coordinate.

Theorem 2.13 *Suppose the assumptions of Theorem 2.12 and condition (B5) are satisfied, then it follows for any* $\mathbf{x} = (x_1, \ldots, x_d)^T \in (0,1)^d$ *with* $\frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}}) > 0$ *that*

$$\sqrt{nh_1 \ldots h_d}(\hat{m}_{I_1}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}}) - b(x_1, x_{\underline{1}})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x_1, x_{\underline{1}}),$$

where the asymptotic bias and variance are given by

$$\begin{aligned} b(x_1, x_{\underline{1}}) &= \kappa_2(K) \sum_{l=1}^{d} h_l^2 \left[\frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + 2 \frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}}) \frac{\partial}{\partial x_l} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} \right] \\ s^2(x_1, x_{\underline{1}}) &= \frac{\sigma^2(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} ||K||_2^2. \end{aligned}$$

For each monotonization step a corresponding asymptotic normality result can be established. The proofs become a bit more complicated since the estimates are only defined implicitly as the inverse of an operator whereas in the first step the Nadaraya-Watson estimate is defined explicitly. The following Theorems affirm this conclusion and give the asymptotic normality for the second and the last monotonization step.

Theorem 2.14 *If the assumptions of Theorem 2.13 are satisfied, and the regression function* m *is strictly increasing with respect to the first and the second argument, then it follows for all fixed* $x_{\underline{2}} \in (0,1)^{d-1}$ *and for all* $z \in (m(x_{\underline{2}}, 0), m(x_{\underline{2}}, 1))$ *with* $\frac{\partial}{\partial x_k} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})) > 0$ *for all* $k = 1, 2$

$$\sqrt{nh_1 \ldots h_d}(\hat{m}_{I_{1,2}}^{-1}(z|x_{\underline{2}}) - m_2^{-1}(z|x_{\underline{2}}) + b_2(x_{\underline{2}}, z)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_2^2(x_{\underline{2}}, z)),$$

where asymptotic bias and variance are given by

$$\begin{aligned} b_2(x_{\underline{2}}, z) &= \frac{\kappa_2(K)}{\frac{\partial}{\partial x_2} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))} \sum_{l=1}^{d} h_l^2 \left[\frac{\partial^2}{\partial x_l^2} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}) \right. \\ &\quad \left. +2 \frac{\frac{\partial}{\partial x_l} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})) \frac{\partial}{\partial x_l} p(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})}{p(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))} \right], \\ s_2^2(x_{\underline{2}}, z) &= \frac{\sigma^2(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))}{p(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})) \left(\frac{\partial}{\partial x_2} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})) \right)^2} ||K||_2^2 \end{aligned}$$

Corresponding to the second step of the monotonization with respect to the second argument, we obtain the following result.

Theorem 2.15 *If the assumptions of Theorem 2.14 are satisfied, then we have for any* $\mathbf{x} \in (0,1)^d$ *with* $\frac{\partial}{\partial x_k} m(\mathbf{x}) > 0$ *for* $k = 1, 2$

$$\sqrt{nh_1 \ldots h_d}(\hat{m}_{I_{1,2}}(\mathbf{x}) - m(\mathbf{x}) - b(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(\mathbf{x})),$$

where the asymptotic bias and variance are given in Theorem 2.13.

The proof of Theorem 2.15 works with more or less the same arguments as the proof of Theorem 2.13 and is therefore omitted. The last monotonization gives the final asymptotic normality results.

Theorem 2.16 *If the assumptions of Theorem 2.13 are satisfied, and the regression function m is strictly increasing with respect to all d arguments, then it follows for all fixed $x_{\underline{d}} \in (0,1)^{d-1}$ and for all $z \in (m(x_{\underline{d}}, 0), m(x_{\underline{d}}, 1))$ with $\frac{\partial}{\partial x_l} m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}})) > 0$ for all $l = 1, \ldots, d$*

$$\sqrt{nh_1 \ldots h_d}(\hat{m}_{I_{1,\ldots,d}}^{-1}(z|x_{\underline{d}}) - m_d^{-1}(z|x_{\underline{d}}) + b_d(x_{\underline{d}}, z)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_d^2(x_{\underline{d}}, z)),$$

where asymptotic bias and variance are given by

$$\begin{aligned} b_d(x_{\underline{d}}, z) &= \frac{\kappa_2(K)}{\frac{\partial}{\partial x_d} m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}}))} \sum_{l=1}^{d} h_l^2 \Bigg[\frac{\partial^2}{\partial x_l^2} m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}}) \\ &\quad +2 \frac{\frac{\partial}{\partial x_l} m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}})) \frac{\partial}{\partial x_l} p(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}})}{p(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}}))} \Bigg], \\ s_d^2(x_{\underline{d}}, z) &= \frac{\sigma^2(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}}))}{p(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}})) \left(\frac{\partial}{\partial x_d} m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}})) \right)^2} ||K||_2^2 \end{aligned}$$

The final result refers to the asymptotic properties of the estimate $\hat{m}_{I_{1,\ldots,d}}$ obtained after applying the monotonizing procedure for each argument.

Theorem 2.17 *If the assumptions of Theorem 2.16 are satisfied, then we have for any $\mathbf{x} \in (0,1)^d$ with $\frac{\partial}{\partial x_l} m(\mathbf{x}) > 0$ for $l = 1, \ldots, d$*

$$\sqrt{nh_1 \ldots h_d}(\hat{m}_{I_{1,\ldots,d}}(\mathbf{x}) - m(\mathbf{x}) - b(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(\mathbf{x})),$$

where the asymptotic bias and variance are given in Theorem 2.13.

The proof of this Theorem is omitted since it uses the same arguments as for Theorem 2.13.
A few remarks may be appropriate at this point for a better understanding of these results.

Remark 2.18 It follows from Theorem 2.17 that the estimate $\hat{m}_{I_{1,\ldots,d}}$ is first order asymptotic equivalent to the unconstrained Nadaraya-Watson estimate $\hat{m}_{NW}$. Similarly by Theorem 2.13 or Theorem 2.15, the estimates $\hat{m}_{I_1}$ and $\hat{m}_{I_{1,2}}$ are first order asymptotic equivalent to $\hat{m}_{NW}$ and $\hat{m}_{I_{1,\ldots,d}}$. Thus each pair of steps in the algorithm produces a first order asymptotic equivalent estimate with one additional isotonized coordinate. Moreover, Theorem 2.17 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 $\hat{m}_{I_\sigma}$, where σ is a permutation of $\{1, \ldots, d\}$, exhibits the same asymptotic behavior as described for the estimate $\hat{m}_{I_{1,\ldots,d}}$ in Theorem 2.17.

Remark 2.19 It is worthy of mention that similar asymptotic results can be obtained if alternative unconstrained and continuous nonparametric regression estimates are used as preliminary estimate $\hat{m}$. In the last section, we introduced the multivariate local linear estimate $\hat{m}_{LL}$ [see (2.3)]. A similar analysis as for $\hat{m}_{NW}$ shows that the statement in Theorem 2.17 is still valid if the bias $b(\mathbf{x})$ is replaced by

$$\tilde{b}(\mathbf{x}) = \kappa_2(K) \sum_{l=1}^{d} h_l^2 \frac{\partial^2}{\partial x_l^2} m(\mathbf{x}).$$

Again the monotonized estimate $\hat{m}_{I_{1,\ldots,d}}$ is first order equivalent to the unconstrained local linear estimate $\hat{m}_{LL}$. Because of the better performance at the boundaries, the local linear estimate is used in the numerical examples [see Section 2.4].

Remark 2.20 It follows from the proofs of the above Theorems that under the slightly stronger assumption $n(h_1 + \ldots + h_d)^{d+4} = o(1)$ [compare condition (B2)] the estimate $\hat{m}_{I_{1,\ldots,d}}$ has no asymptotic bias. That means that Theorem 2.17 becomes

$$\sqrt{nh_1 \ldots h_d}(\hat{m}_{I_{1,\ldots,d}}(\mathbf{x}) - m(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(\mathbf{x}))$$

under the miscellaneous assumptions given in the Theorem. Similar statements hold for Theorem 2.12-2.16.

2.3.3 Proof of Theorem 2.12

For the sake of simplicity, we assume $h_1 = h_2 = \ldots = h_d$ and $n = N$. Recall that we specified as preliminary estimator

$$\hat{m}(\mathbf{x}) = \hat{m}_{NW}(\mathbf{x}) = \frac{\sum_{j=1}^{n} K_H(\mathbf{x} - \mathbf{X}_j) Y_j}{\sum_{j=1}^{n} K_H(\mathbf{x} - \mathbf{X}_j)},$$

where $H = \operatorname{diag}(h_1, \ldots, h_1)$. The proof follows by adapting the arguments in Dette et al. (2006) to the d-dimensional case. These authors discussed the case where h_1 and h_m are of the same order, i.e. $\lim_{n\to\infty} \frac{h_1}{h_m} = c > 0$. Hence, we present the proof with full details. Furthermore, the proof of Theorem 2.12 helps to understand the more intrinsic proofs of the theorems where the next isotonization steps are applied. We define

$$m_{I_1}^{-1}(z|x_{\underline{1}}) = \frac{1}{nh_m} \sum_{i=1}^{n} \int_{-\infty}^{z} K_m\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - u}{h_m}\right) du$$

as the approximation of the inverse $m_1^{-1}(z|x_{\underline{1}})$ with respect to the first argument. First of all, we analyze the function

$$g\left(\frac{\hat{m}(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right) = \int_{\frac{\hat{m}(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}}^{\infty} K_m(v) dv = \frac{1}{h_m} \int_{-\infty}^{z} K_m\left(\frac{\hat{m}(\frac{i}{n}, x_{\underline{1}}) - u}{h_m}\right) du.$$

By Taylor's theorem for $\hat{m}(\frac{i}{n}, x_{\underline{1}})$, we obtain the following expansion using $\xi_i(x_{\underline{1}}, z)$ between $m(\frac{i}{n}, x_{\underline{1}})$ and $\hat{m}(\frac{i}{n}, x_{\underline{1}})$

$$\begin{aligned} g\left(\frac{\hat{m}(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right) &= \int_{\frac{\hat{m}(\frac{i}{n}, x_{\underline{1}})-z}{h_m}}^{\infty} K_m(v)dv \\ &= \int_{\frac{m(\frac{i}{n}, x_{\underline{1}})-z}{h_m}}^{\infty} K_m(v)dv \\ &\quad -\frac{1}{h_m} K_m\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right)\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right) \\ &\quad -\frac{1}{2}\frac{1}{h_m^2} K_m'\left(\frac{\xi_i(x_{\underline{1}}, z) - z}{h_m}\right)\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right)^2, \end{aligned}$$

where $\xi_i(x_{\underline{1}}, z)$ satisfies for any $z \in (m(0, x_{\underline{1}}), m(1, x_{\underline{1}}))$ the inequalities

$$|\xi_i(x_{\underline{1}}, z) - m(\frac{i}{n}, x_{\underline{1}})| \le |\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})|$$

for $i = 1, \ldots, n$. We will use similar Taylor expansions for the following proofs.
Now for fixed $x_{\underline{1}}$, we derive a decomposition using Lemma 1.8 with $h_m = o(h_1)$ and the above expansion

$$\begin{aligned} \Delta_n(z|x_{\underline{1}}) &= (\hat{m}_{I_1}^{-1}(z|x_{\underline{1}}) - m_1^{-1}(z|x_{\underline{1}})) \\ &= (\hat{m}_{I_1}^{-1}(z|x_{\underline{1}}) - m_{I_1}^{-1}(z|x_{\underline{1}})) + O\left(h_m^2 + \frac{1}{nh_m}\right) \\ &= (\hat{m}_{I_1}^{-1}(z|x_{\underline{1}}) - m_{I_1}^{-1}(z|x_{\underline{1}})) + o\left(\frac{1}{\sqrt{nh_1^d}}\right) \\ &= \frac{1}{nh_m}\sum_{i=1}^{n}\left(\int_{-\infty}^{z} K_m\left(\frac{\hat{m}(\frac{i}{n}, x_{\underline{1}}) - u}{h_m}\right)du - \int_{-\infty}^{z} K_m\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - u}{h_m}\right)du\right) \\ &\quad +o\left(\frac{1}{\sqrt{nh_1^d}}\right) \\ &= \Delta_n^{(1)}(z|x_{\underline{1}}) + \frac{1}{2}\Delta_n^{(2)}(z|x_{\underline{1}}) + o\left(\frac{1}{\sqrt{nh_1^d}}\right). \end{aligned} \tag{2.7}$$

The third equality uses the bandwidth conditions (B1)-(B4) and the last one defines

$$\begin{aligned} \Delta_n^{(1)}(z|x_{\underline{1}}) &= -\frac{1}{nh_m}\sum_{i=1}^{n} K_m\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right)\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right), \\ \Delta_n^{(2)}(z|x_{\underline{1}}) &= -\frac{1}{nh_m^2}\sum_{i=1}^{n} K_m'\left(\frac{\xi_i(x_{\underline{1}}, z) - z}{h_m}\right)\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right)^2. \end{aligned}$$

To conclude the assertion of the theorem, we analyze the terms $\Delta_n^{(1)}(z|x_{\underline{1}})$ and $\Delta_n^{(2)}(z|x_{\underline{1}})$. In the first step, we show that

$$\sqrt{nh_1^d}\Delta_n^{(2)}(z|x_{\underline{1}}) = o_p(1). \tag{2.8}$$

We apply the following estimation using the asymptotic MSE of the Nadaraya-Watson estimate and the hölder's inequality

$$\begin{aligned} \frac{|\xi_i(x_{\underline{1}}, z) - m(\frac{i}{n}, x_{\underline{1}})|}{h_m} &\leq \frac{|\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})|}{h_m} \\ &= O_p\left(\frac{1}{h_m}\left(h_1^4 + \frac{1}{nh_1^d}\right)^{1/2}\right) = o_p(1), \end{aligned}$$

where the last equation follows from the bandwidth conditions (B1)-(B4). From the Lipschitz continuity of K_m', we obtain finally

$$\left|K_m'\left(\frac{\xi_i(x_{\underline{1}}, z) - z)}{h_m}\right) - K_m'\left(\frac{m(\frac{i}{n}, x_{\underline{1}} - z}{h_m}\right)\right| \leq L\frac{|\xi_i(x_{\underline{1}}, z) - m(\frac{i}{n}, x_{\underline{1}})|}{h_m} = o_p(1). \tag{2.9}$$

Now we examine $\Delta_n^{(2)}(z|x_{\underline{1}})$

$$\begin{aligned} |\Delta_n^{(2)}(z|x_{\underline{1}})| &= \frac{1}{nh_m^2}\left|\sum_{i=1}^n K_m'\left(\frac{\xi_i(x_{\underline{1}}, z) - z}{h_m}\right)\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right)^2\right| \\ &= \frac{1}{nh_m^2}\left|\sum_{i=1}^n K_m'\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right)\left[1 + \left(K_m'\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right)\right)^{-1} \times \right.\right. \\ &\left.\left.\left(K_m'\left(\frac{\xi_i(x_{\underline{1}}, z) - z)}{h_m}\right) - K_m'\left(\frac{m(\frac{i}{n}, x_{\underline{1}} - z}{h_m}\right)\right)\right]\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right)^2\right| \\ &= \frac{(1 + o_p(1))}{nh_m^2}\left|\sum_{i=1}^n K_m'\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right)\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right)^2\right| \\ &\leq \frac{(1 + o_p(1))}{nh_m^2}\sum_{i=1}^n \left|K_m'\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right)\right|\left(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})\right)^2 \\ &= \frac{(1 + o_p(1))}{h_m^2}\int_0^1 \left|K_m'\left(\frac{m(x_1, x_{\underline{1}}) - z}{h_m}\right)\right|\left(\hat{m}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}})\right)^2 dx_1 \\ &= \frac{(1 + o_p(1))}{h_m}\int_{-1}^1 |K_m'(s)|ds\frac{\left(\hat{m}(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) - m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2}{\frac{\partial}{\partial x_1}m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})}. \end{aligned}$$

Taking the expectation of the last expression, we obtain

$$\begin{aligned} E[|\Delta_n^{(2)}(z|x_{\underline{1}})|] &\leq \frac{1}{h_m}\int_{-1}^{1}|K'_m(s)|ds\frac{E\left[\left(\hat{m}(m_1^{-1}(z|x_{\underline{1}}),x_{\underline{1}})-m(m_1^{-1}(z|x_{\underline{1}}),x_{\underline{1}})\right)^2\right]}{\frac{\partial}{\partial x_1}m(m_1^{-1}(z|x_{\underline{1}}),x_{\underline{1}})} \\ &= O\left(\frac{1}{h_m}\left(h_1^4+\frac{1}{nh_1^d}\right)\right)=o\left(\frac{1}{\sqrt{nh_1^d}}\right), \end{aligned}$$

where we used again the bandwidth conditions (B1)-(B4).
The remaining term $\Delta_n^{(1)}(z|x_{\underline{1}})$ is decomposed as follows using the specific structure of the Nadaraya-Watson estimator [for the local linear estimator we obtain a different decomposition], model (2.6), and $|\hat{p}(x_1,x_{\underline{1}})-p(x_1,x_{\underline{1}})|=o_p(1)$

$$\begin{aligned} \Delta_n^{(1)}(z|x_{\underline{1}}) &= -\frac{(1+o_p(1))}{n^2h_m}\sum_{i,j=1}^{n}K_m\left(\frac{m(\frac{i}{n},x_{\underline{1}})-z}{h_m}\right)K_H\left(\left(\frac{i}{n},x_{\underline{1}}\right)^T-\mathbf{X}_j\right)\frac{Y_j-m(\frac{i}{n},x_{\underline{1}})}{p(\frac{i}{n},x_{\underline{1}})} \\ &= (1+o_p(1))(\Delta_n^{(1.1)}(z|x_{\underline{1}})+\Delta_n^{(1.2)}(z|x_{\underline{1}})) \end{aligned} \tag{2.10}$$

with

$$\begin{aligned} \Delta_n^{(1.1)}(z|x_{\underline{1}}) &= -\frac{1}{n^2h_m}\sum_{j,i=1}^{n}K_m\left(\frac{m(\frac{i}{n},x_{\underline{1}})-z}{h_m}\right)K_H\left(\left(\frac{i}{n},x_{\underline{1}}\right)^T-\mathbf{X}_j\right)\times \\ &\qquad \frac{m(X_{j1},X_{j\underline{1}})-m(\frac{i}{n},x_{\underline{1}})}{p(\frac{i}{n},x_{\underline{1}})}, \\ \Delta_n^{(1.2)}(z|x_{\underline{1}}) &= -\frac{1}{n^2h_m}\sum_{j,i=1}^{n}K_m\left(\frac{m(\frac{i}{n},x_{\underline{1}})-z}{h_m}\right)K_H\left(\left(\frac{i}{n},x_{\underline{1}}\right)^T-\mathbf{X}_j\right)\frac{\sigma(\mathbf{X}_j)\varepsilon_j}{p(\frac{i}{n},x_{\underline{1}})}. \end{aligned}$$

These terms correspond to bias and variance in Theorem 2.12. To see this, we calculate the expectation of $\Delta_n^{(1.1)}(z|x_{\underline{1}})$. The sum over i is interpreted as the approximation of the corresponding integral which exists. Therefore the remainder term converges to zero.

$$\begin{aligned} E[\Delta_n^{(1.1)}(z|x_{\underline{1}})] &= -\frac{(1+o(1))}{h_m}\int K_m\left(\frac{m(x_1,x_{\underline{1}})-z}{h_m}\right)K_H(\mathbf{x}-\mathbf{u})\times \\ &\quad \frac{m(u_1,u_{\underline{1}})-m(x_1,x_{\underline{1}})}{p(x_1,x_{\underline{1}})}p(u_1,u_{\underline{1}})dx_1du_1du_{\underline{1}} \\ &= -\frac{(1+o(1))}{h_m}\int K_m\left(\frac{m(x_1,x_{\underline{1}})-z}{h_m}\right)K(v_1,v_{\underline{1}})\times \\ &\quad \frac{m(x_1-h_1v_1,x_{\underline{1}}-h_1v_{\underline{1}})-m(x_1,x_{\underline{1}})}{p(x_1,x_{\underline{1}})}p(x_1-h_1v_1,x_{\underline{1}}-h_1v_{\underline{1}})dv_1dv_{\underline{1}}dx_1 \\ &= -\frac{(1+o(1))}{h_m}\int K_m\left(\frac{m(x_1,x_{\underline{1}})-z}{h_m}\right)K(v_1,v_{\underline{1}})\times \end{aligned}$$

$$
\begin{aligned}
& \sum_{l=1}^{d} h_1^2 \left[\frac{1}{2} v_l^2 \frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + v_l^2 \frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}}) \frac{\partial}{\partial x_l} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} \right] dx_1 dv_1 dv_{\underline{1}} \\
= \; & -\frac{(1+o(1))}{h_m} h_1^2 \kappa_2(K) \int_0^1 K_m\left(\frac{m(x_1, x_{\underline{1}}) - z}{h_m}\right) \times \\
& \sum_{l=1}^{d} \left[\frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + 2 \frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}}) \frac{\partial}{\partial x_1} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} \right] dx_1 \\
= \; & -(1+o(1)) h_1^2 \kappa_2(K) \int_{-1}^{1} K_m(s) \frac{1}{\frac{\partial}{\partial x_1} m(m_1^{-1}(z + h_m s | x_{\underline{1}}), x_{\underline{1}})} \times \qquad (2.11) \\
& \sum_{l=1}^{d} \left[\frac{\partial^2}{\partial x_l^2} m(m_1^{-1}(z + h_m s | x_{\underline{1}}), x_{\underline{1}}) \right. \\
& \left. + 2 \frac{\frac{\partial}{\partial x_l} m(m_1^{-1}(z + h_m s | x_{\underline{1}}), x_{\underline{1}}) \frac{\partial}{\partial x_1} p(m_1^{-1}(z + h_m s | x_{\underline{1}}), x_{\underline{1}})}{p(m_1^{-1}(z + h_m s | x_{\underline{1}}), x_{\underline{1}})} \right] ds \\
= \; & -(1+o(1)) h_1^2 \kappa_2(K) \frac{1}{\frac{\partial}{\partial x_1} m(m_1^{-1}(z | x_{\underline{1}}), x_{\underline{1}})} \times \\
& \sum_{l=1}^{d} \left[\frac{\partial^2}{\partial x_l^2} m(m_1^{-1}(z | x_{\underline{1}}), x_{\underline{1}}) + 2 \frac{\frac{\partial}{\partial x_l} m(m_1^{-1}(z | x_{\underline{1}}), x_{\underline{1}}) \frac{\partial}{\partial x_1} p(m_1^{-1}(z | x_{\underline{1}}), x_{\underline{1}})}{p(m_1^{-1}(z | x_{\underline{1}}), x_{\underline{1}})} \right] \\
= \; & -(1+o(1)) b_1(z, x_{\underline{1}}). \qquad (2.12)
\end{aligned}
$$

For the second equation, we apply the substitution $v_k = \frac{u_k - x_k}{h_1}$ for each component $k = 1, \ldots, d$. Equation (2.11) follows by the substitution $s = \frac{m(x_1, x_{\underline{1}}) - z}{h_m}$ with respect to x_1. The last identity is true, since we set $h_1 = h_2 = \ldots = h_d$. Now we compute the variance of $\Delta_n^{(1.1)}(z|x_{\underline{1}})$ using similar substitutions as for the expectation.

$$
\begin{aligned}
\mathrm{Var}(\Delta_n^{(1.1)}(z|x_{\underline{1}}) \;\leq\; & \frac{1}{n^3 h_m^2} E\left[\left(\sum_{i=1}^{n} K_m\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right) K_H\left((\frac{i}{n}, x_{\underline{1}})^T - \mathbf{X}_j \right) \times \right.\right. \\
& \left.\left. \frac{m(X_{j1}, X_{j\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})}{p(\frac{i}{n}, x_{\underline{1}})} \right)^2 \right] \\
= \; & \frac{(1+o(1))}{n h_m^2} \int \left(\int_0^1 K_m\left(\frac{m(x_1, x_{\underline{1}}) - z}{h_m}\right) K_H(\mathbf{x} - \mathbf{u}) \times \right. \\
& \left. \frac{m(u_1, u_{\underline{1}}) - m(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} dx_1 \right)^2 p(u_1, u_{\underline{1}}) du_1 du_{\underline{1}} \\
= \; & \frac{(1+o(1))}{n} \int \left(\int_{-1}^{1} K_m(s) ds \frac{K_H((m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})^T - \mathbf{u})}{\frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} \times \right.
\end{aligned}
$$

$$
\begin{aligned}
& \left. \frac{m(u_1, u_{\underline{1}}) - m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})}{p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} \right)^2 p(u_1, u_{\underline{1}}) du_1 du_{\underline{1}} \\
= \;& \frac{(1+o(1))}{n} \int \frac{K_H^2((m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})^T - \mathbf{u})}{\left(p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) \frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2} \times \\
& \left(m(u_1, u_{\underline{1}}) - m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2 p(u_1, u_{\underline{1}}) du_1 du_{\underline{1}} \\
= \;& \frac{(1+o(1))}{nh_1^d} \frac{1}{\left(p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) \frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2} \times \\
& \int K^2(\mathbf{v}) \left(m(m_1^{-1}(z|x_{\underline{1}}) - h_1 v_1, x_{\underline{1}} - h_1 v_{\underline{1}}) - m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2 \\
& p(m_1^{-1}(z|x_{\underline{1}}) - h_1 v_1, x_{\underline{1}} - h_1 v_{\underline{1}}) dv_1 dv_{\underline{1}} \\
= \;& \frac{(1+o(1))}{nh_1^d} \frac{1}{\left(p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) \frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2} \times \\
& \sum_{l=1}^d \int v_l^2 K^2(\mathbf{v}) d\mathbf{v} (\frac{\partial}{\partial x_l} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}))^2 p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}}) \\
= \;& O\left(\frac{h_1^2}{nh_1^d}\right) = o\left(\frac{1}{nh_1^d}\right). \qquad (2.13)
\end{aligned}
$$

Therefore the variance of $\Delta_n^{(1)}(z|x_{\underline{1}})$ is negligible. We can go ahead and analyze the last term $\Delta_n^{(1.2)}(z|x_{\underline{1}})$. This term has expectation 0 since $E[\varepsilon_j] = 0$ and ε_j and $\mathbf{X}_j$ are independent. For the standardized variance of $\Delta_n^{(1.2)}(z|x_{\underline{1}})$, we obtain

$$
\begin{aligned}
& \operatorname{Var}(\sqrt{nh_1^d} \Delta_n^{(1.2)}(z|x_{\underline{1}})) \\
= \;& \frac{h_1^d}{n^3 h_m^2} \sum_{j=1}^n \operatorname{Var}\left(\sum_{i=1}^n K_m \left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m} \right) K_H((\frac{i}{n}, x_{\underline{1}})^T - \mathbf{X}_j) \frac{\sigma(\mathbf{X}_j)\varepsilon_j}{p(\frac{i}{n}, x_{\underline{1}})} \right) \\
= \;& \frac{(1+o(1))h_1^d}{h_m^2} \int \sigma^2(u_1, u_{\underline{1}}) \left(\int_0^1 K_m \left(\frac{m(x_1, x_{\underline{1}}) - z}{h_m} \right) \frac{K_H(\mathbf{x} - \mathbf{u})}{p(x_1, x_{\underline{1}})} dx_1 \right)^2 p(u_1, u_{\underline{1}}) du_1 du_{\underline{1}} \\
= \;& \frac{(1+o(1))h_1^d}{h_m^2} \int \sigma^2(u_1, u_{\underline{1}}) \left(\int_{-1}^1 \frac{K_m(s) K_H(\mathbf{u} - (m_1^{-1}(z + h_m s|x_{\underline{1}}), x_{\underline{1}})^T}{p(m_1^{-1}(z + h_m s|x_{\underline{1}}), x_{\underline{1}})} ds \right)^2 p(u_1, u_{\underline{1}}) du_1 du_{\underline{1}} \\
= \;& (1+o(1)) h_1^d \int \frac{\sigma^2(u_1, u_{\underline{1}}) K_H^2((m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})^T - \mathbf{u})}{\left(\frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2 p^2(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} p(u_1, u_{\underline{1}}) du_1 du_{\underline{1}} \qquad (2.14) \\
= \;& (1+o(1)) \int K^2(\mathbf{v}) \frac{\sigma^2(m_1^{-1}(z|x_{\underline{1}}) - h_1 v_1, x_{\underline{1}} - h_1 v_{\underline{1}}) p(m_1^{-1}(z|x_{\underline{1}}) - h_1 v_1, x_{\underline{1}} - h_1 v_{\underline{1}})}{\left(\frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2 p^2(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} dv_1 dv_{\underline{1}} \\
= \;& (1+o(1)) \frac{\sigma^2(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})}{\left(\frac{\partial}{\partial x_1} m(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})\right)^2 p(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} ||K||_2^2 = (1+o(1)) s_1^2(z, x_{\underline{1}}). \qquad (2.15)
\end{aligned}
$$

The identity in (2.14) of the above equations is true, since we have $\frac{h_m}{h_1} \to 0$ from condition (B3). This bandwidth condition is crucial to get the right rate of convergence.
As final step, we show the Lyapunov condition for $\delta = 2$ to apply the central limit theorem on $\Delta_n^{(1.2)}(z|x_{\underline{1}})$.

$$\begin{aligned}
&\sum_{j=1}^{n} E\left[\frac{\sqrt{nh_1^d}}{n^2 h_m}\sum_{i=1}^{n} K_m\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - z}{h_m}\right) K_H((\frac{i}{n}, x_{\underline{1}})^T - \mathbf{X}_j)\frac{\sigma(\mathbf{X}_j)\varepsilon_j}{p(\frac{i}{n}, x_{\underline{1}})}\right]^4 \\
= \; & \frac{(1+o(1))E[\varepsilon_1^4]h_1^{2d}}{nh_m^4}\int \sigma^4(u_1, u_{\underline{1}}) \times \\
&\left(\int_0^1 K_m\left(\frac{m(x_1, x_{\underline{1}}) - z}{h_m}\right)\frac{K_H(\mathbf{x}-\mathbf{u})}{p(x_1, x_{\underline{1}})}dx_1\right)^4 p(u_1, u_{\underline{1}})du_1 du_{\underline{1}} \\
= \; & \frac{(1+o(1))E[\varepsilon_1^4]h_1^{2d}}{n}\int \sigma^4(u_1, u_{\underline{1}})\left(\int_{-1}^1 K_m(s)\frac{K_H((m_1^{-1}(z+h_m s|x_{\underline{1}}), x_{\underline{1}})^T - \mathbf{u})}{p(m_1^{-1}(z+h_m s|x_{\underline{1}}), x_{\underline{1}})}ds\right)^4 \\
& p(u_1, u_{\underline{1}})du_1 du_{\underline{1}} \\
= \; & \frac{(1+o(1))E[\varepsilon_1^4]h_1^{2d}}{n}\int \sigma^4(u_1, u_{\underline{1}})\frac{K_H^4((m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})^T - \mathbf{u})}{p^4(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})}p(u_1, u_{\underline{1}})du_1 du_{\underline{1}} \\
= \; & \frac{(1+o(1))E[\varepsilon_1^4]}{nh_1^d}\int K^4(\mathbf{v})\frac{\sigma^4(m_1^{-1}(z|x_{\underline{1}}) - h_1 v_1, x_{\underline{1}} - h_1 v_{\underline{1}})}{p^4(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} \times \\
& p(m_1^{-1}(z|x_{\underline{1}}) - h_1 v_1, x_{\underline{1}} - h_1 v_{\underline{1}})dv_1 dv_{\underline{1}} \\
= \; & \frac{(1+o(1))E[\varepsilon_1^4]}{nh_1^d}||K||_4^4\frac{\sigma^4(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})}{p^3(m_1^{-1}(z|x_{\underline{1}}), x_{\underline{1}})} = O\left(\frac{1}{nh_1^d}\right) = o(1). \qquad (2.16)
\end{aligned}$$

In line 4 of these equations, we again used the assumption (B3), namely $\lim_{n\to\infty}\frac{h_m}{h_1} = 0$. The assertion of Theorem 2.12 follows now from a combination of (2.7), (2.8), (2.10), (2.12), (2.13), (2.15), and (2.16). □

2.3.4 Proof of Theorem 2.13

We assume as in the foregoing proof $h_1 = h_2 = \ldots = h_d$ and $n = N$. To prove this Theorem, we apply the special Taylor expansion for functionals we shortly discussed at the end of Chapter 1. For fixed $x_1 \in (0,1)$ and $x_{\underline{1}} \in (0,1)^{d-1}$, we define

$$\epsilon = \min\left\{\frac{m(x_1, x_{\underline{1}}) - m(0, x_{\underline{1}})}{2}, \frac{m(1, x_{\underline{1}}) - m(x_1, x_{\underline{1}})}{2}\right\}$$

and set $D = (m(x_1, x_{\underline{1}}) - \epsilon, m(x_1, x_{\underline{1}}) + \epsilon)$. By Theorem 2.6, the estimate $\hat{m}$ converges a.s. uniformly to m, so that the support of $\hat{m}_I^{-1}(\cdot|x_{\underline{1}})$, which is essentially given by

$$[\min_{x_1\in[0,1]}\hat{m}(x_1, x_{\underline{1}}), \max_{x_1\in[0,1]}\hat{m}(x_1, x_{\underline{1}})],$$

contains the set D for all sufficiently large n. Thus, we basically examine the following decomposition

$$\hat{m}_{I_1}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}}) = Q'(0) + Q''(\lambda^*),$$

where Q' and Q'' are defined in (1.9) and (1.10), respectively, with $g_2 = \hat{m}_{I_1}^{-1}(\cdot|x_{\underline{1}})$ and $g_1 = m_1^{-1}(\cdot|x_{\underline{1}})$. Q' and Q'' are the Gâteaux derivatives of the operator Φ, which maps a function to its inverse. To be precise, these expressions are given by

$$\begin{aligned} Q'(0) &= -\frac{(\hat{m}_{I_1}^{-1} - m_1^{-1})}{\frac{\partial}{\partial x_1}(m_1^{-1})} \circ (m(x_1, x_{\underline{1}})|x_{\underline{1}}) = -\frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}})(\hat{m}_{I_1}^{-1} - m_1^{-1}) \circ (m(x_1, x_{\underline{1}})|x_{\underline{1}}), \\ Q''(\lambda^*) &= 2\frac{(\hat{m}_{I_1}^{-1} - m_1^{-1})\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_1^{-1})}{\left[\frac{\partial}{\partial x_1}(m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))\right]^2} \circ (m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(x_1, x_{\underline{1}}) \\ &\quad - \frac{(\hat{m}_{I_1}^{-1} - m_1^{-1})^2 \frac{\partial^2}{\partial x_1^2}(m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))}{\left[\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))\right]^3} \circ (m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(x_1, x_{\underline{1}}). \end{aligned}$$

Later we will show that

$$Q''(\lambda^*) = o_p\left(\frac{1}{\sqrt{nh_1^d}}\right). \tag{2.17}$$

The assertion of the Theorem can be transformed by this decomposition into something we already studied.

$$\begin{aligned} &\sqrt{nh_1^d}\Bigg(\hat{m}_{I_1}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}}) \\ &\qquad - \kappa_2(K)h_1^2 \sum_{l=1}^d \left(\frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + 2\frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}})\frac{\partial}{\partial x_l} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})}\right)\Bigg) \\ &= -\sqrt{nh_1^d}\Bigg(\frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}})(\hat{m}_{I_1}^{-1}(m(x_1, x_{\underline{1}})|x_{\underline{1}}) - m_1^{-1}(m(x_1, x_{\underline{1}})|x_{\underline{1}})) \\ &\qquad + \kappa_2(K)h_1^2 \sum_{l=1}^d \left(\frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + 2\frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}})\frac{\partial}{\partial x_l} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})}\right)\Bigg) + o_p(1) \\ &= -\frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}})\sqrt{nh_1^d}\Bigg(\hat{m}_{I_1}^{-1}(m(x_1, x_{\underline{1}})|x_{\underline{1}}) - m_1^{-1}(m(x_1, x_{\underline{1}})|x_{\underline{1}}) \\ &\qquad + \frac{\kappa_2(K)h_1^2}{\frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}})} \sum_{l=1}^d \left(\frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + 2\frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}})\frac{\partial}{\partial x_l} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})}\right)\Bigg) + o_p(1) \end{aligned}$$

We know from Theorem 2.12 with $z = m(x_1, x_{\underline{1}})$ that the above expression is asymptotically normal, i.e.

$$-\frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}}) \sqrt{nh_1^d} \Bigg(\hat{m}_{I_1}^{-1}(m(x_1, x_{\underline{1}})|x_{\underline{1}}) - m_1^{-1}(m(x_1, x_{\underline{1}})|x_{\underline{1}}) + \frac{\kappa_2(K) h_1^2}{\frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}})} \sum_{l=1}^{d} \left(\frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + 2 \frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}}) \frac{\partial}{\partial x_l} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} \right) \Bigg) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\partial}{\partial x_1} m(x_1, x_{\underline{1}})^2 s_1^2(z, x_{\underline{1}})\right).$$

This yields the assertion of the Theorem:

$$\sqrt{nh_1^d} \left(\hat{m}_{I_1}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}}) - \kappa_2(K) h_1^2 \sum_{l=1}^{d} \left(\frac{\partial^2}{\partial x_l^2} m(x_1, x_{\underline{1}}) + 2 \frac{\frac{\partial}{\partial x_l} m(x_1, x_{\underline{1}}) \frac{\partial}{\partial x_l} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} \right) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(z, x_{\underline{1}})).$$

Finally we have to show that the reminder term $Q''(\lambda^*)$ is negligible as claimed in (2.17). We break $Q''(\lambda^*)$ into two parts.

$$Q''(\lambda^*) = 2B_{n1}(x_1, x_{\underline{1}}) - B_{n2}(x_1, x_{\underline{1}}),$$

where

$$\begin{aligned} B_{n1}(x_1, x_{\underline{1}}) &= \frac{(\hat{m}_{I_1}^{-1} - m_1^{-1}) \frac{\partial}{\partial x_1} (\hat{m}_{I_1}^{-1} - m_1^{-1})}{\left[\frac{\partial}{\partial x_1} (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))\right]^2} \circ (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(x_1, x_{\underline{1}}) \\ B_{n2}(x_1, x_{\underline{1}}) &= \frac{(\hat{m}_{I_1}^{-1} - m_1^{-1})^2 \frac{\partial^2}{\partial x_1^2} (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))}{\left[\frac{\partial}{\partial x_1} (\hat{m}_{I_1}^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))\right]^3} \circ (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(x_1, x_{\underline{1}}). \end{aligned}$$

Observe that for fixed z

$$(m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))(z|x_{\underline{1}}) \xrightarrow{P} m_1^{-1}(z|x_{\underline{1}}),$$

which follows from Theorem 2.12 and the Chebyshev's inequality. Furthermore, we have for fixed x_1

$$t_n(x_1, x_{\underline{1}}) = (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(x_1, x_{\underline{1}}) \xrightarrow{P} m(x_1, x_{\underline{1}}),$$

because of the following

$$\begin{aligned} |t_n(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}})| &= |(m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}})| \\ &= |(m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(m_1^{-1}(u|x_{\underline{1}}), x_{\underline{1}}) - u| \qquad (2.18) \\ &= |(m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}(m_1^{-1}(u|x_{\underline{1}}), x_{\underline{1}}) \\ &\quad - (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1})^{-1} \circ (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))(u|x_{\underline{1}})| \\ &= |((m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1})'(\xi_n) \times \qquad (2.19) \\ &\quad (m_1^{-1}(u|x_{\underline{1}}) - (m_1^{-1} + \lambda^* (\hat{m}_{I_1}^{-1} - m_1^{-1}))(u|x_{\underline{1}}))|. \end{aligned}$$

In line (2.18) of the above equations, we set $u = m(x_1, x_{\underline{1}})$ and in line (2.19), we applied a first order Taylor expansion. Note that the derivative of $(m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))^{-1}$ is bounded and $(m_1^{-1} + \lambda^*(\hat{m}_{I_1}^{-1} - m_1^{-1}))(u|x_{\underline{1}}) \xrightarrow{P} m_1^{-1}(u|x_{\underline{1}})$. To deal with the term $\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_1^{-1})(\xi_n|x_{\underline{1}})$, which occurs in the enumerator and the denominator of $B_{n1}(x_1, x_{\underline{1}})$ and $B_{n2}(x_1, x_{\underline{1}})$, we note first that

$$|\hat{m}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}})| = O_p\left(\frac{\log n}{nh_1^d}\right)^{1/2}$$

[see Theorem 2.6 and Collomb and Härdle (1986) for more details on that strong uniform convergence rate]. In total, we have using Lemma 1.8

$$\begin{aligned}
\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_1^{-1})(\xi_n|x_{\underline{1}}) &= \frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_{I_1}^{-1})(\xi_n|x_{\underline{1}}) + \frac{\partial}{\partial x_1}(m_{I_1}^{-1} - m_1^{-1})(\xi_n|x_{\underline{1}}) \\
&= \frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_{I_1}^{-1})(\xi_n|x_{\underline{1}}) + O\left(h_m^2 + \frac{1}{nh_m}\right) \\
&= \frac{1}{nh_m^2}\sum_{i=1}^n\left[K_m\left(\frac{\hat{m}(\frac{i}{n}, x_{\underline{1}}) - \xi_n}{h_m}\right) - K_m\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - \xi_n}{h_m}\right)\right] \\
&\quad +O\left(h_m^2 + \frac{1}{nh_m}\right) \\
&= \frac{1}{nh_m^3}\sum_{i=1}^n K_m'\left(\frac{\eta_i(\xi_n, x_{\underline{1}}) - \xi_n}{h_m}\right)(\hat{m}(\frac{i}{n}, x_{\underline{1}}) - m(\frac{i}{n}, x_{\underline{1}})) \\
&\quad +O\left(h_m^2 + \frac{1}{nh_m}\right),
\end{aligned}$$

where $\eta_i(\xi_n, x_{\underline{1}})$ has the following property

$$\left|\eta_i(\xi_n, x_{\underline{1}}) - m\left(\frac{i}{n}, x_{\underline{1}}\right)\right| < \left|\hat{m}\left(\frac{i}{n}, x_{\underline{1}}\right) - m\left(\frac{i}{n}, x_{\underline{1}}\right)\right| = O_p\left(\frac{\log n}{nh_1^d}\right)^{1/2}.$$

By the bandwidth condition (B5) and the Lipschitz continuity of K_m', we can exchange $K_m'\left(\frac{\eta_i(\xi_n, x_{\underline{1}}) - \xi_n}{h_m}\right)$ by $K_m'\left(\frac{m(\frac{i}{n}, x_{\underline{1}}) - \xi_n}{h_m}\right)$ as in the proof of Theorem 2.12 [see (2.9)] and get the following

$$\begin{aligned}
\frac{\partial}{\partial x_1}(\hat{m}_{I_1}^{-1} - m_1^{-1})(\xi_n|x_{\underline{1}}) &= \frac{(1+o(1))}{h_m^3}\int_0^1 K_m'\left(\frac{m(x_1, x_{\underline{1}}) - \xi_n}{h_m}\right)(\hat{m}(x_1, x_{\underline{1}}) - m(x_1, x_{\underline{1}}))dx_1 \\
&\quad +O\left(h_m^2 + \frac{1}{nh_m}\right) \\
&= \frac{(1+o(1))}{h_m^2}\frac{\hat{m}(m_1^{-1}(\xi_n|x_{\underline{1}}), x_{\underline{1}}) - m(m_1^{-1}(\xi_n|x_{\underline{1}}), x_{\underline{1}})}{\frac{\partial}{\partial x_1}m(m_1^{-1}(\xi_n|x_{\underline{1}}), x_{\underline{1}})}
\end{aligned}$$

$$
\begin{aligned}
& +O\left(h_m^2+\frac{1}{nh_m}\right) \\
= \quad & O_p\left(\left(\frac{\log n}{nh_1^d h_m^2}\right)^{1/2}+h_m^2+\frac{1}{nh_m}\right)=o_p(1),
\end{aligned}
$$

where the last inequality follows again by bandwidth condition (B5). From Theorem 2.12, we derive

$$
\left(\hat{m}_{I_1}^{-1}(m(x_1,x_{\underline{1}})|x_{\underline{1}})-m_1^{-1}(m(x_1,x_{\underline{1}})|x_{\underline{1}})\right)=O_p\left(\frac{1}{\sqrt{nh_1^d}}\right).
$$

So basically, we have

$$
\begin{aligned}
B_{n1}(x_1,x_{\underline{1}}) \quad &= \quad O_p\left(\frac{1}{\sqrt{nh_1^d}}\left(\left(\frac{\log n}{nh_1^d h_m^2}\right)^{1/2}+h_m^2+\frac{1}{nh_m}\right)\right)=o_p\left(\frac{1}{\sqrt{nh_1^d}}\right) \\
B_{n2}(x_1,x_{\underline{1}}) \quad &= \quad O_p\left(\frac{1}{nh_1^d}\right)=o_p\left(\frac{1}{\sqrt{nh_1^d}}\right),
\end{aligned}
$$

where the last equation is true since the second derivative $\frac{\partial^2}{\partial x_1^2}m(x_1,x_{\underline{1}})$ is bounded. □

2.3.5 Proof of Theorem 2.14

Again we assume $h_1=h_2=\ldots,h_d$ and $n=N$. Since the estimate $\hat{m}_{I_1}(x_1,\ldots,x_d)$ is defined only implicitly, the decomposition is more extensive than in the proof of Theorem 2.12. We analyze the following difference and apply Lemma 1.8 with $h_m=o(h_1)$ and $x_{\underline{2}}=(x_1,x_3,\ldots,x_d)^T$, the vector without the component x_2:

$$
\begin{aligned}
\Delta_n(z|x_{\underline{2}}) \quad &= \quad (\hat{m}_{I_{1,2}}^{-1}(z|x_{\underline{2}})-m_2^{-1}(z|x_{\underline{2}})) \\
&= \quad (\hat{m}_{I_{1,2}}^{-1}(z|x_{\underline{2}})-m_{I_2}^{-1}(z|x_{\underline{2}}))+O\left(h_m^2+\frac{1}{nh_m}\right),
\end{aligned}
$$

where

$$
m_{I_2}^{-1}(z|x_{\underline{2}})=\frac{1}{nh_m}\sum_{i=1}^{n}\int_{-\infty}^{z}K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-u}{h_m}\right)du
$$

is an approximation of $m_2^{-1}(z|x_{\underline{2}})$. The approach is similar to the proof of Theorem 2.12, but it is more elaborate to unravel the bias and variance part of this statistic. We obtain none the less the following decomposition using the bandwidth conditions (B1)-(B4)

$$\begin{aligned}
\Delta_n(z|x_{\underline{2}}) &= (\hat{m}_{I_{1,2}}^{-1}(z|x_{\underline{2}}) - m_{I_2}^{-1}(z|x_{\underline{2}})) + o\left(\frac{1}{\sqrt{nh_1^d}}\right) \\
&= \frac{1}{nh_m}\sum_{i=1}^{n}\left(\int_{-\infty}^{z} K_m\left(\frac{\hat{m}_{I_1}(x_{\underline{2}}, \frac{i}{n}) - u}{h_m}\right) du - \int_{-\infty}^{z} K_m\left(\frac{m(x_{\underline{2}}, \frac{i}{n}) - u}{h_m}\right) du\right) \\
&\quad + o\left(\frac{1}{\sqrt{nh_1^d}}\right) \\
&= \Delta_n^{(1)}(z|x_{\underline{2}}) + \frac{1}{2}\Delta_n^{(2)}(z|x_{\underline{2}}) + o\left(\frac{1}{\sqrt{nh_1^d}}\right), \qquad (2.20)
\end{aligned}$$

where the terms $\Delta_n^{(1)}(z|x_{\underline{2}})$ and $\Delta_n^{(2)}(z|x_{\underline{2}})$ are derived by a Taylor expansion and defined as follows

$$\begin{aligned}
\Delta_n^{(1)}(z|x_{\underline{2}}) &= -\frac{1}{nh_m}\sum_{i=1}^{n} K_m\left(\frac{m(x_{\underline{2}}, \frac{i}{n}) - z}{h_m}\right)\left(\hat{m}_{I_1}\left(x_{\underline{2}}, \frac{i}{n}\right) - m\left(x_{\underline{2}}, \frac{i}{n}\right)\right), \\
\Delta_n^{(2)}(z|x_{\underline{2}}) &= -\frac{1}{nh_m^2}\sum_{i=1}^{n} K_m'\left(\frac{\xi_i(x_{\underline{2}}, z) - z}{h_m}\right)\left(\hat{m}_{I_1}\left(x_{\underline{2}}, \frac{i}{n}\right) - m\left(x_{\underline{2}}, \frac{i}{n}\right)\right)^2.
\end{aligned}$$

Note that these terms correspond to the terms $\Delta_n^{(1)}(z|x_{\underline{1}})$ and $\Delta_n^{(2)}(z|x_{\underline{1}})$ in the proof of Theorem 2.12, but $\hat{m}$ is exchanged by $\hat{m}_{I_1}$. The number $\xi_i(x_{\underline{2}}, z)$ fulfills the estimation

$$\left|\xi_i(x_{\underline{2}}, z) - m(x_{\underline{2}}, \frac{i}{n})\right| \leq \left|\hat{m}_{I_1}(x_{\underline{2}}, \frac{i}{n}) - m(x_{\underline{2}}, \frac{i}{n})\right|$$

for all $i = 1, \ldots, n$ and all $z \in (m(x_{\underline{2}}, 0), m(x_{\underline{2}}, 1))$. Since $\hat{m}_{I_1}(x_1, \ldots, x_d)$ has the same asymptotic properties as the Nadaraya-Watson estimate $\hat{m}(x_1, \ldots, x_d)$, we have a similar behavior in terms of stochastic order symbols

$$(\hat{m}_{I_1}(\mathbf{x}) - m(\mathbf{x}))^2 = O_p\left(h_1^4 + \frac{1}{nh_1^d}\right).$$

Therefore the remainder term $\Delta_n^{(2)}(z|x_{\underline{2}})$ can be handled as the corresponding term in the proof of Theorem 2.12 using the Lipschitz continuity of K_m' and the bandwidth conditions (B1)-(B4) [see (2.9)].

$$\begin{aligned}
|\Delta_n^{(2)}(z|x_{\underline{2}})| &= \frac{1}{nh_m^2}\left|\sum_{i=1}^{n} K_m'\left(\frac{\xi_i(x_{\underline{2}}, z) - z}{h_m}\right)\left(\hat{m}_{I_1}(x_{\underline{2}}, \frac{i}{n}) - m(x_{\underline{2}}, \frac{i}{n})\right)^2\right| \\
&\leq \frac{(1 + o_p(1))}{h_m^2}\int_0^1 \left|K_m'\left(\frac{m(x_{\underline{2}}, x_2) - z}{h_m}\right)\right| \left(\hat{m}_{I_1}(x_{\underline{2}}, x_2) - m(x_{\underline{2}}, x_2)\right)^2 dx_2
\end{aligned}$$

$$\begin{aligned}
&= \frac{(1+o_p(1))}{h_m}\int_{-1}^{1}|K_m'(s)|ds\frac{(\hat{m}_{I_1}(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))-m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}})))^2}{\frac{\partial}{\partial x_2}m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}})} \\
&= O_p\left(\frac{1}{h_m}\left(h_1^4+\frac{1}{nh_1^d}\right)\right)=o_p\left(\frac{1}{\sqrt{nh_1^d}}\right). \qquad (2.21)
\end{aligned}$$

Now we turn to the term $\Delta_n^{(1)}(z|x_{\underline{2}})$. In the following, we will fractionalize this term gradually to get a feeling how it is composed. Using the decomposition from the proof of Theorem 2.13, we have

$$\hat{m}_{I_1}(\mathbf{x})-m(\mathbf{x})=-\frac{\partial}{\partial x_1}m(\mathbf{x})\left(\hat{m}_{I_1}^{-1}(m(\mathbf{x})|x_{\underline{1}})-m_1^{-1}(m(\mathbf{x})|x_{\underline{1}})\right)+o_p\left(\frac{1}{\sqrt{nh_1^d}}\right).$$

This yields

$$\Delta_n^{(1)}(z|x_{\underline{2}})=(1+o_p(1))\Delta_n^{(1.1)}(z|x_{\underline{2}}), \qquad (2.22)$$

where

$$\begin{aligned}
\Delta_n^{(1.1)}(z|x_{\underline{2}}) &= \frac{1}{nh_m}\sum_{i=1}^{n}K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})\times \\
&\quad \left(\hat{m}_{I_1}^{-1}\left(m\left(x_{\underline{2}},\frac{i}{n}\right)|\frac{i}{n},x_3,\ldots,x_d\right)-m_1^{-1}\left(m\left(x_{\underline{2}},\frac{i}{n}\right)|\frac{i}{n},x_3,\ldots,x_d\right)\right) \\
&= \Delta_n^{(1.1.1)}(z|x_{\underline{2}})+\Delta_n^{(1.1.2)}(z|x_{\underline{2}}) \qquad (2.23)
\end{aligned}$$

with

$$\begin{aligned}
\Delta_n^{(1.1.1)}(z|x_{\underline{2}}) &= \frac{1}{nh_m}\sum_{i=1}^{n}K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})\times \\
&\quad \left(\hat{m}_{I_1}^{-1}\left(m\left(x_{\underline{2}},\frac{i}{n}\right)|\frac{i}{n},x_3,\ldots,x_d\right)-m_{I_1}^{-1}\left(m\left(x_{\underline{2}},\frac{i}{n}\right)|\frac{i}{n},x_3,\ldots,x_d\right)\right) \\
\Delta_n^{(1.1.2)}(z|x_{\underline{2}}) &= \frac{1}{nh_m}\sum_{i=1}^{n}K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})\times \\
&\quad \left(m_{I_1}^{-1}\left(m\left(x_{\underline{2}},\frac{i}{n}\right)|\frac{i}{n},x_3,\ldots,x_d\right)-m_1^{-1}\left(m\left(x_{\underline{2}},\frac{i}{n}\right)|\frac{i}{n},x_3,\ldots,x_d\right)\right).
\end{aligned}$$

The second term $\Delta_n^{(1.1.2)}(z|x_{\underline{2}})$ is basically of order $h_m^2+\frac{1}{nh_m}$ by applying Lemma 1.8 and therefore asymptotically negligible. The first term has an explicit representation

$$\begin{aligned}\Delta_n^{(1.1.1)}(z|x_{\underline{2}}) &= \frac{1}{n^2h_m^2}\sum_{i,j=1}^n K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})\times\\ &\quad\left[\int_{-\infty}^{m(x_{\underline{2}},\frac{i}{n})} K_m\left(\frac{\hat{m}(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)-u}{h_m}\right)du\right.\\ &\quad\left.-\int_{-\infty}^{m(x_{\underline{2}},\frac{i}{n})} K_m\left(\frac{m(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)-u}{h_m}\right)du\right]\\ &= \Delta_n^{(1.1.1.a)}(z|x_{\underline{2}})+\frac{1}{2}\Delta_n^{(1.1.1.b)}(z|x_{\underline{2}}). \qquad (2.24)\end{aligned}$$

The terms $\Delta_n^{(1.1.1.a)}(z|x_{\underline{2}})$ and $\Delta_n^{(1.1.1.b)}(z|x_{\underline{2}})$ are defined by a Taylor expansion as the terms (2.7) in the proof of Theorem 2.12 or the terms $\Delta_n^{(1)}(z|x_{\underline{2}})$ and $\Delta_n^{(2)}(z|x_{\underline{2}})$ in (2.20) at the beginning of this proof. Since we applied two monotonization steps, the kernel K_m appears two times in these expressions. To be precise, we have

$$\begin{aligned}\Delta_n^{(1.1.1.a)}(z|x_{\underline{2}}) &= -\frac{1}{n^2h_m^2}\sum_{i,j=1}^n\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)\times\\ &\quad K_m\left(\frac{m(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)-m(x_{\underline{2}},\frac{i}{n})}{h_m}\right)\times\\ &\quad\left(\hat{m}\left(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d\right)-m\left(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d\right)\right),\\ \Delta_n^{(1.1.1.b)}(z|x_{\underline{2}}) &= -\frac{1}{n^2h_m^3}\sum_{i,j=1}^n\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)\times\\ &\quad K_m'\left(\frac{\xi_{ij}(x_3,\ldots,x_d)-m(x_{\underline{2}},\frac{i}{n})}{h_m}\right)\times\\ &\quad\left(\hat{m}\left(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d\right)-m\left(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d\right)\right)^2.\end{aligned}$$

The number $\xi_{ij}(x_3,\ldots,x_d)$ fulfills

$$\begin{aligned}\left|\xi_{ij}(x_3,\ldots,x_d)-m\left(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d\right)\right| &\le \left|\hat{m}\left(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d\right)-m\left(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d\right)\right|\\ &= O_p\left(\left(h_1^4+\frac{1}{nh_1^d}\right)^{1/2}\right).\end{aligned}$$

First, we can show that $\Delta_n^{(1.1.1.b)}(z|x_{\underline{2}})$ is negligible similarly as for $\Delta_n^{(2)}(z|x_{\underline{2}})$.

$$\begin{aligned}|\Delta_n^{(1.1.1.b)}(z|x_{\underline{2}})| &\leq \frac{(1+o_p(1))}{h_m^3}\int_0^1\int_0^1 \frac{\partial}{\partial x_1}m(x_{\underline{2}},x_2)K_m\left(\frac{m(x_{\underline{2}},x_2)-z}{h_m}\right) \\ &\quad \left|K_m'\left(\frac{m(u_1,x_{\underline{1}})-m(x_{\underline{2}},x_2)}{h_m}\right)\right|\left(\hat{m}(u_1,x_{\underline{1}})-m(u_1,x_{\underline{1}})\right)^2 dx_2du_1 \\ &= O_p\left(\frac{1}{h_m}\left(h_1^4+\frac{1}{nh_1^d}\right)^{1/2}\right) = o_p\left(\frac{1}{\sqrt{nh_1^d}}\right). \qquad (2.25)\end{aligned}$$

The other term $\Delta_n^{(1.1.1.a)}(z|x_{\underline{2}})$ can be split up into bias and variance term in a similar manner as in the proof of Theorem 2.12 [see (2.10)]

$$\Delta_n^{(1.1.1.a)}(z|x_{\underline{2}}) = (1+o_p(1))(\Delta_n^{(1.1.1.a.a)}(z|x_{\underline{2}})+\Delta_n^{(1.1.1.a.b)}(z|x_{\underline{2}})), \qquad (2.26)$$

where

$$\begin{aligned}\Delta_n^{(1.1.1.a.a)}(z|x_{\underline{2}}) &= -\frac{1}{n^3h_m^2}\sum_{i,j,k=1}^n \frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)\times \\ &\quad K_m\left(\frac{m(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)-m(x_{\underline{2}},\frac{i}{n})}{h_m}\right)\times \\ &\quad K_H\left((\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)^T-\mathbf{X}_k\right)\frac{m(\mathbf{X}_k)-m(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)}{p(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)} \\ \Delta_n^{(1.1.1.a.b)}(z|x_{\underline{2}}) &= -\frac{1}{n^3h_m^2}\sum_{i,j,k=1}^n \frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})\times \\ &\quad K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)K_m\left(\frac{m(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)-m(x_{\underline{2}},\frac{i}{n})}{h_m}\right)\times \\ &\quad K_H\left((\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)^T-\mathbf{X}_k\right)\frac{\sigma(\mathbf{X}_k)\varepsilon_k}{p(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)}.\end{aligned}$$

Now we calculate the expectation and the variance of the first term. For $n\to\infty$, we approximate the sums by the corresponding integrals.

$$\begin{aligned}&E[\Delta_n^{(1.1.1.a.a)}(z|x_{\underline{2}})] \\ &= -\frac{(1+o(1))}{h_m^2}\int\frac{\partial}{\partial x_1}m(x_{\underline{2}},x_2)K_m\left(\frac{m(x_{\underline{2}},x_2)-z}{h_m}\right)K_m\left(\frac{m(u_1,x_{\underline{1}})-m(x_{\underline{2}},x_2)}{h_m}\right) \\ &\quad K_H((u_1,x_{\underline{1}})^T-\mathbf{v})\frac{m(\mathbf{v})-m(u_1,x_{\underline{1}})}{p(u_1,x_{\underline{1}})}p(\mathbf{v})dx_2du_1d\mathbf{v} \\ &= -\frac{(1+o(1))}{h_m^2}\int\frac{\partial}{\partial x_1}m(x_{\underline{2}},x_2)K_m\left(\frac{m(x_{\underline{2}},x_2)-z}{h_m}\right)\times\end{aligned}$$

$$
\begin{aligned}
& K_m\left(\frac{m(u_1,x_{\underline{1}})-m(x_{\underline{2}},x_2)}{h_m}\right)K(\mathbf{w})\times \\
& \frac{m(u_1-h_1w_1,x_{\underline{1}}-h_1w_{\underline{1}})-m(u_1,x_{\underline{1}})}{p(u_1,x_{\underline{1}})}p(u_1-h_1w_1,x_{\underline{1}}-h_1w_{\underline{1}})dx_2du_1d\mathbf{w} \\
= & -\frac{(1+o(1))\kappa_2(K)}{h_m^2}\int\frac{\partial}{\partial x_1}m(x_{\underline{2}},x_2)K_m\left(\frac{m(x_{\underline{2}},x_2)-z}{h_m}\right)\times \\
& K_m\left(\frac{m(u_1,x_{\underline{1}})-m(x_{\underline{2}},x_2)}{h_m}\right)\times \\
& h_1^2\sum_{l=1}^{d}\left[\frac{\partial^2}{\partial x_l^2}m(u_1,x_{\underline{1}})+2\frac{\frac{\partial}{\partial x_l}m(u_1,x_{\underline{1}})\frac{\partial}{\partial x_l}p(u_1,x_{\underline{1}})}{p(u_1,x_{\underline{1}})}\right]dx_2du_1 \\
= & -\frac{(1+o(1))\kappa_2(K)}{h_m}\int K_m\left(\frac{m(x_{\underline{2}},x_2)-z}{h_m}\right)K_m(s)ds \qquad (2.27)\\
& h_1^2\sum_{l=1}^{d}\left[\frac{\partial^2}{\partial x_l^2}m(x_1,x_{\underline{1}})+2\frac{\frac{\partial}{\partial x_l}m(x_1,x_{\underline{1}})\frac{\partial}{\partial x_l}p(x_1,x_{\underline{1}})}{p(x_1,x_{\underline{1}})}\right]dx_2 \\
= & -(1+o(1))\frac{\kappa_2(K)h_1^2}{\frac{\partial}{\partial x_2}m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))}\times \qquad (2.28)\\
& \sum_{l=1}^{d}\left[\frac{\partial^2}{\partial x_l^2}m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))+2\frac{\frac{\partial}{\partial x_l}m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))\frac{\partial}{\partial x_l}p(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))}{p(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))}\right] \\
= & -(1+o(1))b_2(x_{\underline{2}},z).
\end{aligned}
$$

In line (2.27), the substitution $s=\frac{m(u_1,x_{\underline{1}})-m(x_{\underline{2}},x_2)}{h_m}$ is applied. By the substitution, the derivative $\frac{\partial}{\partial x_1}m(x_{\underline{2}},x_2)$ disappears. Finally, the substitution $w=\frac{m(x_{\underline{2}},x_2)-z}{h_m}$ is used in line (2.28) and provides the factor $\frac{1}{\frac{\partial}{\partial x_2}m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))}$ in the bias. This term only contributes to the bias. When we compute its variance, we can show that it is asymptotically negligible.

$$
\begin{aligned}
& \mathrm{Var}(\Delta_n^{(1.1.1.a.a)}(z|x_{\underline{2}})) \\
= & \frac{1}{n^6h_m^4}\mathrm{Var}\Bigg(\sum_{i,j,k=1}^{n}\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)K_m\left(\frac{m(\frac{j}{n},\frac{i}{n},x_3,\dots,x_d)-m(x_{\underline{2}},\frac{i}{n})}{h_m}\right) \\
& K_H\left((\frac{j}{n},\frac{i}{n},x_3,\dots,x_d)^T-\mathbf{X}_k\right)\frac{m(\mathbf{X}_k)-m(\frac{j}{n},\frac{i}{n},x_3,\dots,x_d)}{p(\frac{j}{n},\frac{i}{n},x_3,\dots x_d)}\Bigg) \\
\le & \frac{1}{n^5h_m^4}E\Bigg[\Bigg(\sum_{i,j=1}^{n}\frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)K_m\left(\frac{m(\frac{j}{n},\frac{i}{n},x_3,\dots,x_d)-m(x_{\underline{2}},\frac{i}{n})}{h_m}\right) \\
& K_H\left((\frac{j}{n},\frac{i}{n},x_3,\dots,x_d)^T-\mathbf{X}_k\right)\frac{m(\mathbf{X}_k)-m(\frac{j}{n},\frac{i}{n},x_3,\dots,x_d)}{p(\frac{j}{n},\frac{i}{n},x_3,\dots x_d)}\Bigg)^2\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
= & \frac{(1+o(1))}{nh_m^2}\int\Bigg(\int \frac{\partial}{\partial x_1}m(x_{\underline{2}},x_2)K_m\left(\frac{m(x_{\underline{2}},x_2)-z}{h_m}\right)K_m\left(\frac{m(v_1,x_{\underline{1}})-m(x_{\underline{2}},x_2)}{h_m}\right) \\
& K_H((v_1,x_{\underline{1}})^T-\mathbf{u})\frac{m(\mathbf{u})-m(v_1,x_{\underline{1}})}{p(v_1,x_{\underline{1}})}dv_1dx_2\Bigg)^2 p(u_{\underline{2}},u_2)d\mathbf{u} \\
= & \frac{(1+o(1))}{nh_m}\int\Bigg(\int \frac{\frac{\partial}{\partial x_1}m(x_{\underline{2}},x_2)}{\frac{\partial}{\partial x_1}m(m_1^{-1}(m(x_{\underline{2}},x_2)+h_mw_1|x_{\underline{1}}),x_{\underline{1}})}K_m\left(\frac{m(x_{\underline{2}},x_2)-z}{h_m}\right) \\
& K_m(w_1)K_H\left((m_1^{-1}(m(x_{\underline{2}},x_2)+h_mw_1|x_{\underline{1}}),x_{\underline{1}})^T-\mathbf{u}\right) \\
& \frac{m(\mathbf{u})-m(m_1^{-1}(m(x_{\underline{2}},x_2)+h_mw_1|x_{\underline{1}}),x_{\underline{1}})}{p(m_1^{-1}(m(x_{\underline{2}},x_2)+h_mw_1|x_{\underline{1}}),x_{\underline{1}})}dw_1dx_2\Bigg)^2 p(u_{\underline{2}},u_2)d\mathbf{u} \\
= & \frac{(1+o(1))}{n}\int\Bigg(\int \frac{K_m(w_2)}{\frac{\partial}{\partial x_2}m(x_{\underline{2}},m_2^{-1}(z+h_mw_2|x_{\underline{2}}))}K_H((x_{\underline{2}},m_2^{-1}(z+h_mw_2|x_{\underline{2}}))^T-\mathbf{u}) \\
& \frac{m(\mathbf{u})-m(x_{\underline{2}},m_2^{-1}(z+h_mw_2|x_{\underline{2}}))}{p(x_{\underline{2}},m_2^{-1}(z+h_mw_2|x_{\underline{2}}))}dw_2\Bigg)^2 p(u_{\underline{2}},u_2)d\mathbf{u} \\
= & \frac{(1+o(1))}{n}\int \frac{K_H^2((x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))^T-\mathbf{u})}{\left(\frac{\partial}{\partial x_2}m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))\right)^2}\frac{\left(m(\mathbf{u})-m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))\right)^2}{p^2(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))}p(u_{\underline{2}},u_2)d\mathbf{u} \\
= & \frac{(1+o(1))}{nh_1^d}\int \frac{K^2(\mathbf{v})\left(m(x_{\underline{2}}-h_1v_{\underline{2}},m_2^{-1}(z|x_{\underline{2}})-h_1v_2)-m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))\right)^2}{\left(\frac{\partial}{\partial x_2}m(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))\right)^2p^2(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))} \\
& p(x_{\underline{2}}-h_1v_{\underline{2}},m_2^{-1}(z|x_{\underline{2}})-h_1v_2)dv_{\underline{2}}dv_2 \\
= & O\left(\frac{1}{nh_1^{d-2}}\right)=o\left(\frac{1}{nh_1^d}\right).
\end{aligned}
$$

In the first equation, we used the i.i.d. assumption of the data. Then we applied successively the following substitutions: $w_1=\frac{m(v_1,x_{\underline{1}})-m(x_{\underline{2}},x_2)}{h_m}$ with respect to v_1, $w_2=\frac{m(x_{\underline{2}},x_2)-z}{h_m}$ with respect to x_2, and finally $\mathbf{v}=\frac{(x_{\underline{2}},m_2^{-1}(z|x_{\underline{2}}))^T-\mathbf{u}}{h_1}$ componentwise. Note that for the first two substitutions we need $h_m=o(h_1)$.

The remaining term $\Delta_n^{(1.1.1.a.b)}(z|x_{\underline{2}})$ has obviously zero expectation. We calculate its variance standardized with $\sqrt{nh_1^d}$ by using similar substitutions as for $\mathrm{Var}(\Delta_n^{(1.1.1.a.a)}(z|x_{\underline{2}}))$.

$$
\begin{aligned}
& \mathrm{Var}(\sqrt{nh_1^d}\Delta_n^{(1.1.1.a.b)}(z|x_{\underline{2}})) \\
= & \frac{h_1^d}{n^5h_m^4}\mathrm{Var}\Bigg(\sum_{i,j,k=1}^n \frac{\partial}{\partial x_1}m(x_{\underline{2}},\frac{i}{n})K_m\left(\frac{m(x_{\underline{2}},\frac{i}{n})-z}{h_m}\right)K_m\left(\frac{m(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)-m(x_{\underline{2}},\frac{i}{n})}{h_m}\right) \\
& K_H\left((\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)^T-\mathbf{X}_k\right)\frac{\sigma(\mathbf{X}_k)\varepsilon_k}{p(\frac{j}{n},\frac{i}{n},x_3,\ldots,x_d)}\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{h_1^d}{n^4 h_m^4} \int \left(\sum_{i,j=1}^{n} \frac{\partial}{\partial x_1} m(x_{\underline{2}}, \frac{i}{n}) K_m \left(\frac{m(x_{\underline{2}}, \frac{i}{n}) - z}{h_m} \right) K_m \left(\frac{m(\frac{j}{n}, \frac{i}{n}, x_3, \ldots, x_d) - m(x_{\underline{2}}, \frac{i}{n})}{h_m} \right) \right. \\
&\quad \left. K_H \left((\frac{j}{n}, \frac{i}{n}, x_3, \ldots, x_d)^T - \mathbf{u} \right) \frac{\sigma(u_{\underline{2}}, u_2)}{p(\frac{j}{n}, \frac{i}{n}, x_3, \ldots, x_d)} \right)^2 p(u_{\underline{2}}, u_2) du_{\underline{2}} du_2 \\
&= \frac{(1+o(1)) h_1^d}{h_m^4} \int \left(\int \frac{\partial}{\partial x_1} m(x_{\underline{2}}, x_2) K_m \left(\frac{m(x_{\underline{2}}, x_2) - z}{h_m} \right) K_m \left(\frac{m(v_1, x_{\underline{1}}) - m(x_{\underline{2}}, x_2)}{h_m} \right) \right. \\
&\quad \left. K_H((v_1, x_{\underline{1}})^T - \mathbf{u}) \frac{\sigma(u_{\underline{2}}, u_2)}{p(v_1, x_{\underline{1}})} dv_1 dx_2 \right)^2 p(u_{\underline{2}}, u_2) du_{\underline{2}} du_2 \\
&= \frac{(1+o(1)) h_1^d}{h_m^2} \int \left(\int K_m \left(\frac{m(x_{\underline{2}}, x_2) - z}{h_m} \right) K_m(w_1) K_H((x_1, x_{\underline{1}})^T - \mathbf{u}) \frac{\sigma(u_{\underline{2}}, u_2)}{p(x_1, x_{\underline{1}})} dw_1 dx_2 \right)^2 \\
&\quad p(u_{\underline{2}}, u_2) du_{\underline{2}} du_2 \\
&= (1+o(1)) h_1^d \int \left(\int \frac{K_m(w_2) K_H((x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))^T - \mathbf{u})}{\frac{\partial}{\partial x_2} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))} \frac{\sigma(u_{\underline{2}}, u_2)}{p(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})} dw_2 \right)^2 p(u_{\underline{2}}, u_2) du_{\underline{2}} du_2 \\
&= (1+o(1)) h_1^d \int \frac{K_H^2((x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))^T - \mathbf{u})}{\left(\frac{\partial}{\partial x_2} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))\right)^2} \frac{\sigma^2(u_{\underline{2}}, u_2)}{p^2(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))} p(u_{\underline{2}}, u_2) du_{\underline{2}} du_2 \\
&= (1+o(1)) \int \frac{K^2(\mathbf{v}) \sigma^2(x_{\underline{2}} - h_1 v_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}) - h_1 v_2)}{\left(\frac{\partial}{\partial x_2} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))\right)^2 p^2(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))} \\
&\quad p(x_{\underline{2}} - h_1 v_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}) - h_1 v_2) dv_{\underline{2}} dv_2 \\
&= (1+o(1)) \frac{\sigma^2(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))}{\left(\frac{\partial}{\partial x_2} m(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))\right)^2 p(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))} ||K||_2^2 \\
&= (1+o(1)) s_2^2(x_{\underline{2}}, z).
\end{aligned}
$$

To conclude the assertion of the Theorem, we have to prove the Lyapunov condition with $\delta = 2$ for the term $\Delta_n^{(1.1.1.a.b)}(z|x_{\underline{2}})$.

$$
\begin{aligned}
&\sum_{k=1}^{n} E \left[\left(\frac{\sqrt{n h_1^d}}{n^3 h_m^2} \sum_{i,j=1}^{n} \frac{\partial}{\partial x_1} m(x_{\underline{2}}, \frac{i}{n}) K_m \left(\frac{m(x_{\underline{2}}, \frac{i}{n}) - z}{h_m} \right) K_m \left(\frac{m(\frac{j}{n}, \frac{i}{n}, x_3, \ldots, x_d) - m(x_{\underline{2}}, \frac{i}{n})}{h_m} \right) \right. \right. \\
&\quad \left. \left. K_H \left((\frac{j}{n}, \frac{i}{n}, x_3, \ldots, x_d)^T - \mathbf{X}_k \right) \frac{\sigma(\mathbf{X}_k) \varepsilon_k}{p(\frac{j}{n}, \frac{i}{n}, x_3, \ldots, x_d)} \right)^4 \right] \\
&= \frac{E[\varepsilon_1^4](1+o(1)) h_1^{2d}}{n h_m^8} \int \left(\int \frac{\partial}{\partial x_1} m(x_{\underline{2}}, x_2) K_m \left(\frac{m(x_{\underline{2}}, x_2) - z}{h_m} \right) K_m \left(\frac{m(v_1, x_{\underline{1}}) - m(x_{\underline{2}}, x_2)}{h_m} \right) \right.
\end{aligned}
$$

$$K_H\left((v_1, x_{\underline{1}})^T - \mathbf{u}\right)\frac{\sigma(\mathbf{u})}{p(v_1, x_{\underline{1}})}dv_1 dx_2\Bigg)^4 p(u_{\underline{2}}, u_2)du_{\underline{2}}du_2$$

$$= \frac{E[\varepsilon_1^4](1+o(1))h_1^{2d}}{nh_m^4}\int\left(\int K_m\left(\frac{m(x_{\underline{2}}, x_2) - z}{h_m}\right)K_H((x_1, x_{\underline{1}})^T - \mathbf{u})\frac{\sigma(\mathbf{u})}{p(x_1, x_{\underline{1}})}dx_2\right)^4$$
$$p(u_{\underline{2}}, u_2)du_{\underline{2}}du_2$$
$$= \frac{E[\varepsilon_1^4](1+o(1))h_1^{2d}}{n}\int K_H^4((x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})^T - \mathbf{u})\frac{\sigma^4(\mathbf{u})}{p^4(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})}p(u_{\underline{2}}, u_2)du_{\underline{2}}du_2$$
$$= \frac{E[\varepsilon_1^4](1+o(1))}{nh_1^d}\left(\int K^4(\mathbf{v})d\mathbf{v}\right)\frac{\sigma^4(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}}))}{p^3(x_{\underline{2}}, m_2^{-1}(z|x_{\underline{2}})} = O\left(\frac{1}{nh_1^d}\right).$$

The asymptotic normality of the random variable $\sqrt{nh_1^d}\Delta_n^{(1.1.1.a.b)}(z|x_{\underline{d}})$ follows by the Lyapunov's theorem, i.e.,

$$\sqrt{nh_1^d}\Delta_n^{(1.1.1.a.b)}(z|x_{\underline{d}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_2^2(x_{\underline{2}}, z)),$$

where the quantity $s_2^2(x_{\underline{2}}, z)$ is defined in Theorem 2.14. The assertion of the Theorem follows now by combining (2.20), (2.21), (2.22), (2.23), (2.24), (2.25), (2.26), and the detailed analysis of the terms $\Delta_n^{(1.1.1.a.a)}(z|x_{\underline{d}})$ and $\Delta_n^{(1.1.1.a.b)}(z|x_{\underline{d}})$. □

2.3.6 Sketch of the Proof 2.16

The proof of Theorem 2.16 follows in a similar manner as the proof of Theorem 2.14. As in the other proofs of this section, we assume $h_1 = h_2 = \ldots = h_d$ and $n = N$. Using Lemma 1.8 with $h_m = o(h_1)$ and $x_{\underline{d}} = (x_1, \ldots, x_{d-1})^T$, the vector without x_d, shows that it is sufficient to analyze

$$\tilde{\Delta}_n(z|x_{\underline{d}}) = (\hat{m}_{I_{1,\ldots,d}}^{-1}(z|x_{\underline{d}}) - m_{I_d}^{-1}(z|x_{\underline{d}}))$$

with

$$m_{I_d}^{-1}(z|x_{\underline{d}}) = \frac{1}{nh_m}\sum_{i=1}^n\int_{-\infty}^z K_m\left(\frac{m(x_{\underline{d}}, \frac{i}{n}) - u}{h_m}\right)du$$

as the approximation of $m_d^{-1}(z|x_{\underline{d}})$. The difference $\tilde{\Delta}_n(z|x_{\underline{d}})$ can be further decomposed by a Taylor expansion and we obtain

$$\tilde{\Delta}_n(z|x_{\underline{d}}) = \tilde{\Delta}_n^{(1)}(z|x_{\underline{d}}) + \frac{1}{2}\tilde{\Delta}_n^{(2)}(z|x_{\underline{d}}),$$

where

$$\tilde{\Delta}_n^{(1)}(z|x_{\underline{d}}) = -\frac{1}{nh_m}\sum_{i=1}^n K_m\left(\frac{m(x_{\underline{d}}, \frac{i}{n}) - z}{h_m}\right)\left(\hat{m}_{I_{1,\ldots,d-1}}\left(x_{\underline{d}}, \frac{i}{n}\right) - m\left(x_{\underline{d}}, \frac{i}{n}\right)\right),$$
$$\tilde{\Delta}_n^{(2)}(z|x_{\underline{d}}) = -\frac{1}{nh_m^2}\sum_{i=1}^n K_m'\left(\frac{\xi_i(x_{\underline{d}}, z) - z}{h_m}\right)\left(\hat{m}_{I_{1,\ldots,d-1}}\left(x_{\underline{d}}, \frac{i}{n}\right) - m\left(x_{\underline{d}}, \frac{i}{n}\right)\right)^2$$

for a number $\xi_i(x_{\underline{d}}, z)$ between $\hat{m}_{I_{1,\ldots,d-1}}\left(x_{\underline{d}}, \frac{i}{n}\right)$ and $m\left(x_{\underline{d}}, \frac{i}{n}\right)$. As in the proof of Theorem 2.14, the term $\tilde{\Delta}_n^{(2)}(z|x_{\underline{d}})$ is asymptotically negligible. The term $\tilde{\Delta}_n^{(1)}(z|x_{\underline{d}})$ can be broken down into an expression with $\hat{m}$ as the term $\Delta_n^{(1)}(z|x_{\underline{2}})$ in the proof of Theorem 2.14. Therefore one has to go back all the monotonizing steps by applying a Taylor expansion in each step which results in something similar to $\Delta_n^{(1.1.1.a)}(z|x_{\underline{2}})$ in the proof of Theorem 2.14 where the kernel K_m appears d times for d monotonizing steps. In this sketch of the proof, we just give a heuristic intuition for that. Roughly, we can approximate $\tilde{\Delta}_n(z|x_{\underline{d}})$ in the following way

$$\begin{aligned}
\tilde{\Delta}_n^{(1)}(z|x_{\underline{d}}) &= -\frac{1}{nh_m}\sum_{i=1}^{n} K_m\left(\frac{m(x_{\underline{d}}, \frac{i}{n}) - z}{h_m}\right)\left(\hat{m}_{I_{1,\ldots,d-1}}\left(x_{\underline{d}}, \frac{i}{n}\right) - m\left(x_{\underline{d}}, \frac{i}{n}\right)\right) \\
&\approx -\frac{1}{h_m}\int_0^1 K_m\left(\frac{m(x_{\underline{d}}, x_d) - z}{h_m}\right)\left(\hat{m}_{I_{1,\ldots,d-1}}(x_{\underline{d}}, x_d) - m(x_{\underline{d}}, x_d)\right) dx_d \\
&\approx -\int_{-1}^{1} K_m(v)\frac{\left(\hat{m}_{I_{1,\ldots,d-1}}(x_{\underline{d}}, m_d^{-1}(z + h_m v|x_{\underline{d}})) - m(x_{\underline{d}}, m_d^{-1}(z + h_m v|x_{\underline{d}}))\right)}{\frac{\partial}{\partial x_d} m(x_{\underline{d}}, m_d^{-1}(z + h_m v|x_{\underline{d}}))} dv \\
&\approx -\frac{\left(\hat{m}_{I_{1,\ldots,d-1}}(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}})) - m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}}))\right)}{\frac{\partial}{\partial x_d} m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}}))}.
\end{aligned}$$

One has to be careful with the approximation in the last line. But in this manner, we can describe the asymptotic behavior of $\hat{m}^{-1}_{I_{1,\ldots,d}}$ in terms of $\hat{m}_{I_{1,\ldots,d-1}}$ evaluated in the point $\left(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}})\right)^T$ and standardized by $\frac{1}{\frac{\partial}{\partial x_d} m(x_{\underline{d}}, m_d^{-1}(z|x_{\underline{d}}))}$. If we assume that $\hat{m}_{I_{1,\ldots,d-1}}$ has the same behavior as $\hat{m}_{I_{1,2}}$ the assertion of the Theorem follows easily [see Theorem 2.15]. □

2.4 Finite sample properties and data analysis

In this section, we investigate the performance of the new multivariate monotone estimate in the two-dimensional case by means of a simulation study and data examples. We emphasize two aspects, to analyze the difference of the monotone estimates due to the order of monotonization and the improvement made by incorporating the monotonicity constraints versus the unconstrained estimator.

2.4.1 A small simulation study

In the following, we simulate i.i.d. data belonging to the nonparametric regression model

$$Y_j = m(X_{j1}, X_{j2}) + \sigma(X_{j1}, X_{j2})\varepsilon_j, \tag{2.29}$$

where it is known that $m(x_1, x_2)$ is strictly monotone increasing in x_1 and x_2. In order to avoid boundary effects, the local linear estimator $\hat{m}_{LL}(x_1, x_2)$ is used as unconstrained estimator [see Example 2.1 for an explicit expression] implemented with a product kernel based on two Epanechnikov kernels, i.e.

$$K(\mathbf{x}) = K(x_1, x_2) = \frac{3}{4}(1 - x_1^2)\frac{3}{4}(1 - x_2^2)I_{[-1,1]}(x_1)I_{[-1,1]}(x_2).$$

The bandwidth is set to

$$h_1 = h_2 = \left(\frac{\tilde{\sigma}^2}{n}\right)^{1/6},$$

where $\tilde{\sigma}^2 = \int \sigma^2(u_1, u_2)du_1du_2$ denotes the integrated variance and n is the sample size. For the kernel of the monotonization procedure K_m the Epanechnikov kernel is used as well. The number of nodes N for evaluating the integral is chosen as $N = 51$ and the bandwidth is given by $h_m = h_1^3$.

Example 2.21 In the first study, we consider the regression model with

$$m(x_1, x_2) = \frac{1}{2}\left(x_2 + \frac{1}{6\pi}\sin(6\pi x_2)\right)\left(1 + (2x_1 - 1)^3\right), \tag{2.30}$$

which is a strictly increasing function with respect to both arguments. For the error distribution in model (2.29), a standard normal distribution is used, 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 2.3, a typical result of one simulation run is displayed. In order to avoid the domination by boundary effects, all graphics in this example will be presented on the square $[0.05, 0.95]^2$.

We observe that the local linear estimate is obviously not isotone with respect to both arguments (see the upper right panel in Figure 2.3). The estimates $\hat{m}_{I_1}$ and $\hat{m}_{I_2}$ look substantially different because isotonization is only performed with respect to one coordinate (the middle panel in Figure 2.3). On the other hand, substantial differences between the isotone estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ (obtaining by interchanging the order of isotonization) are not visible as predicted by our asymptotic theory (see lower panel in Figure 2.3).

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 two-dimensional plots of the estimates of the regression function $m(x_1, x_2)$, where one coordinate has been fixed. In Figure 2.4, we show three typical simulations of the estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$, and the "true" curve, where the coordinate x_2 is fixed at 0.3, 0.5, and 0.7. The left part of the figure corresponds 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 $x_1 = 0.3, 0.5,$ and 0.7 are depicted in Figure 2.5. We observe a reasonable performance of the estimates in all cases and again no substantial difference between the estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ obtained by interchanging the order of isotonization.

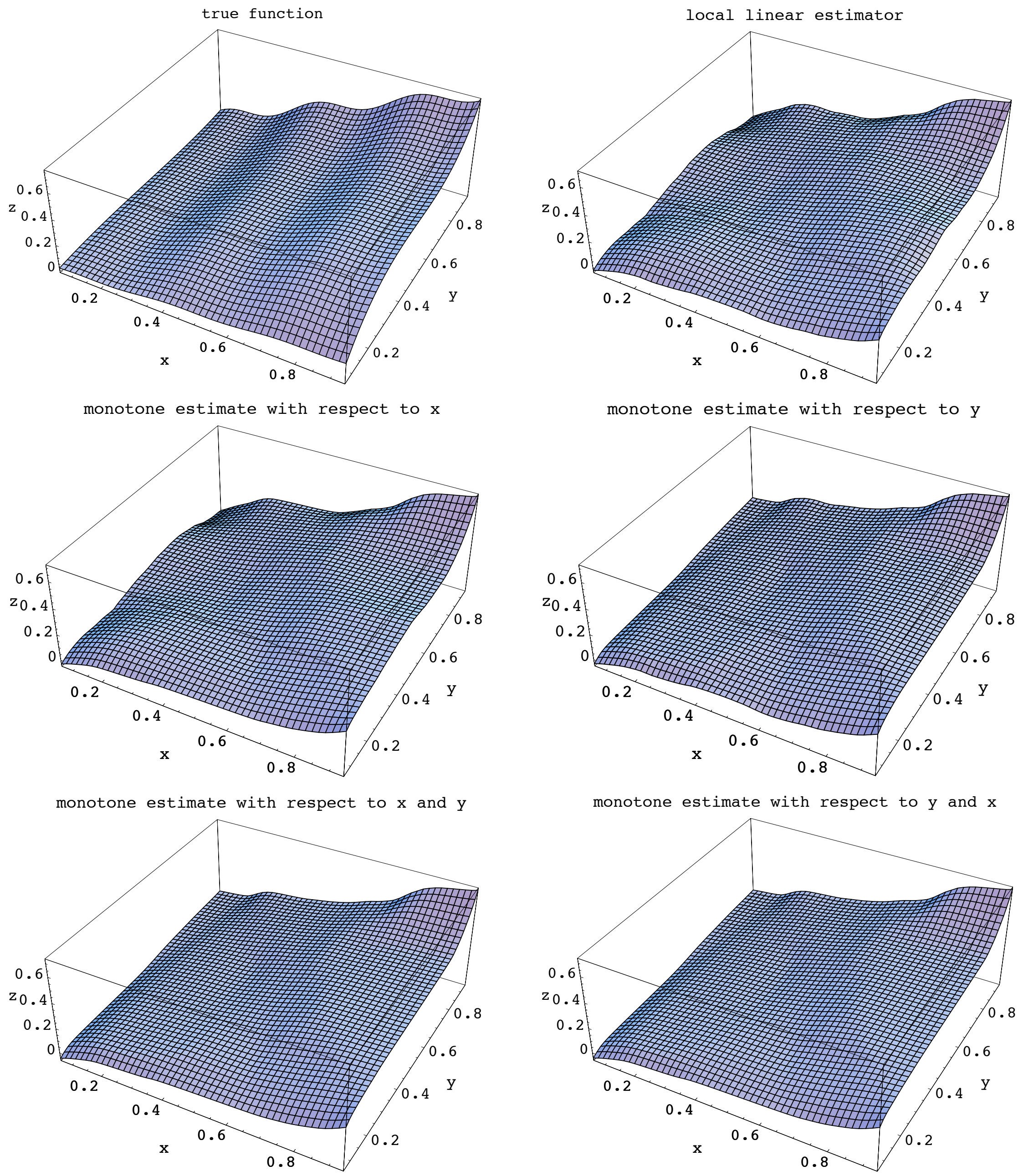

Figure 2.3: The regression function (2.30) and its different estimates. Upper panel: the regression function m (left) and the local linear estimator $\hat{m}$ (right) based on $n = 400$ observations with variance $\sigma^2 = 0.1$. Middle panel: monotone estimate $\hat{m}_{I_1}$ with respect to the first coordinate (left) and monotone estimate $\hat{m}_{I_2}$ with respect to the 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).

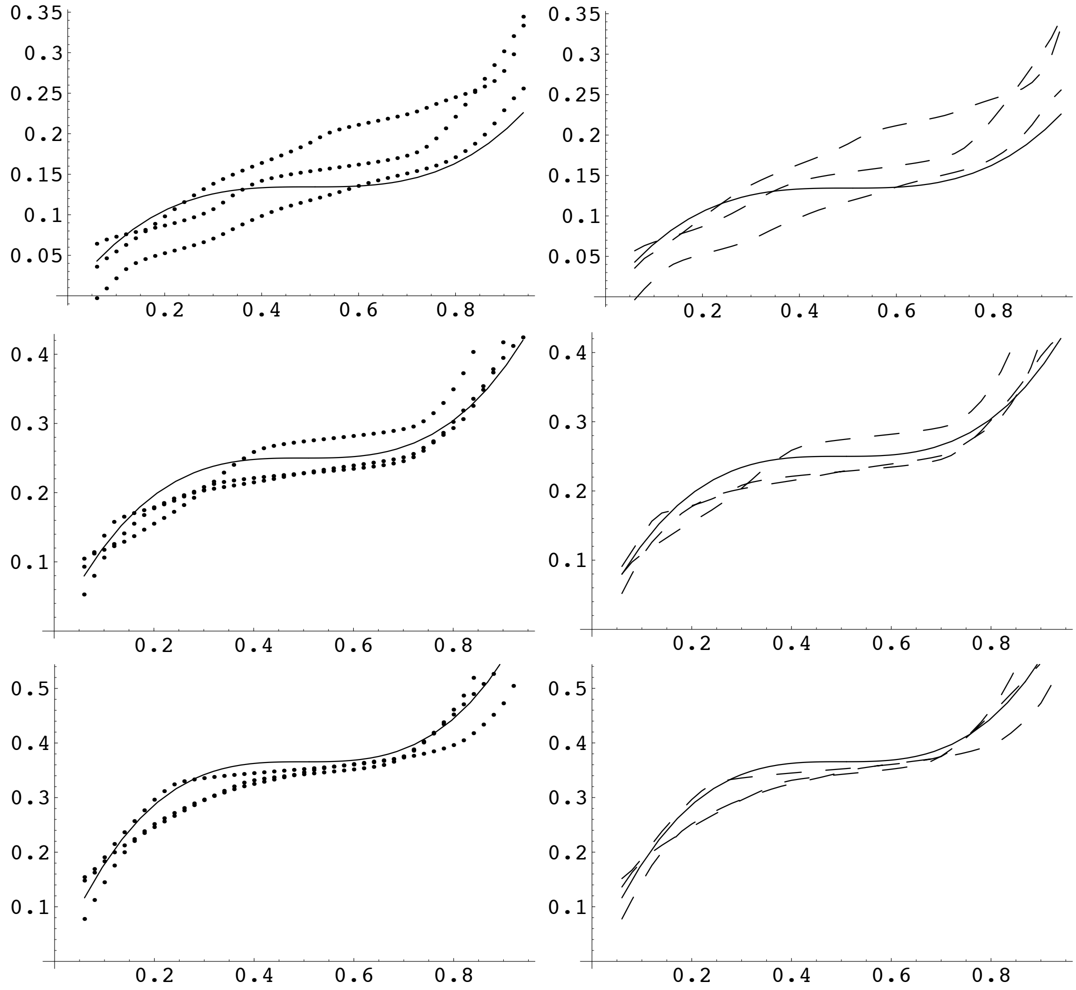

Figure 2.4: The regression function m and its monotone estimates $\hat{m}_{I_{1,2}}$ (dotted lines) and $\hat{m}_{I_{2,1}}$ (dashed lines) in comparison to each other. The three estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$, respectively, are obtained from three different simulation runs. The figure shows the two dimensional functions fixed in x_2: $x_2 = 0.3$ (upper panel), $x_2 = 0.5$ (middle panel), and $x_2 = 0.7$ (lower panel). The solid curve corresponds to the "true" regression function.

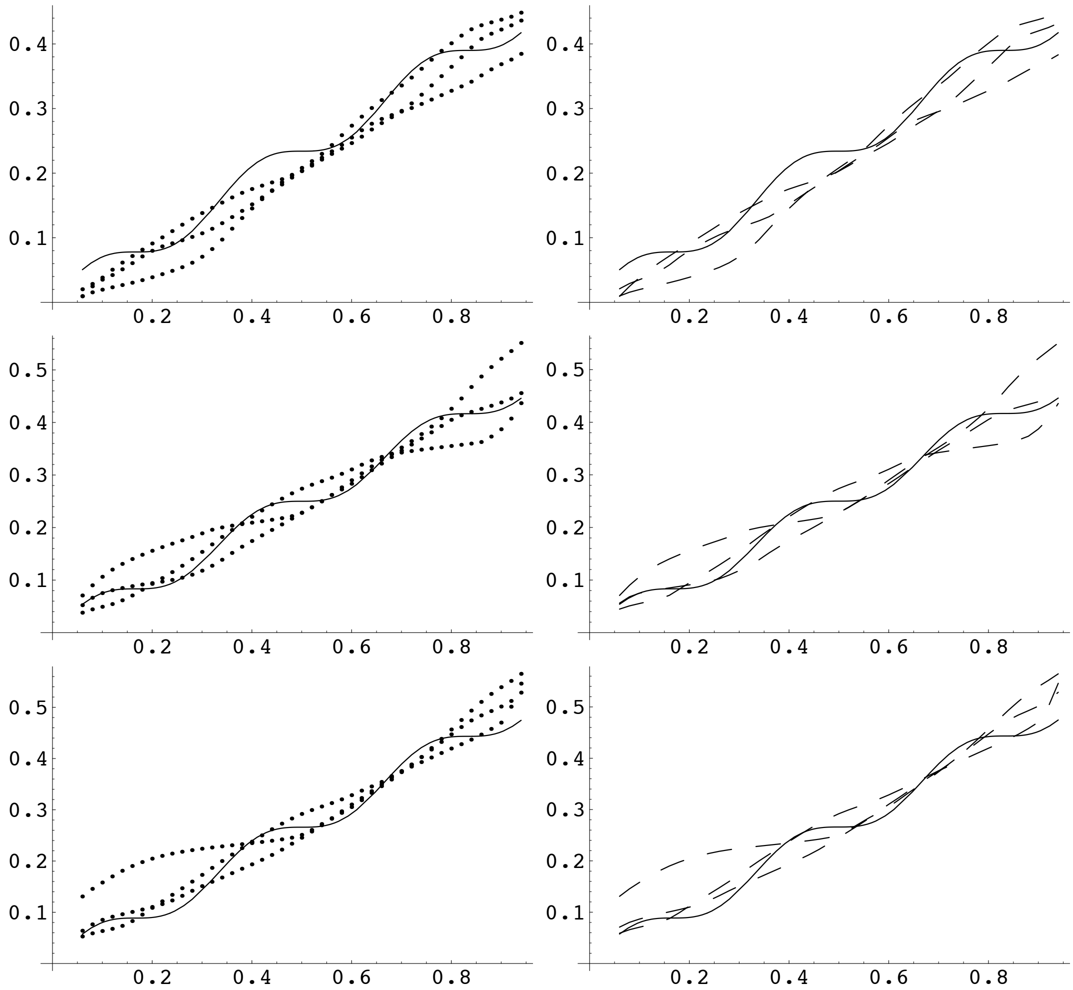

Figure 2.5: The regression function m and its monotone estimates $\hat{m}_{I_{1,2}}$ (dotted lines) and $\hat{m}_{I_{2,1}}$ (dashed lines) in comparison to each other. The three estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$, respectively, are obtained from three different simulation runs. The figure shows the two dimensional functions fixed in x_1: $x_1 = 0.3$ (upper panel), $x_1 = 0.5$ (middle panel), and $x_1 = 0.7$ (lower panel). The solid curve corresponds to the "true" regression function.

Example 2.22 In this example, we compare the isotonized estimates with the unconstrained local linear estimates. Again to avoid the domination by boundary effects, we calculate the mean squared error of both estimates over the square $[0.1, 0.9]^2$. In Figure 2.6, we display the ratios

$$r(x_1, x_2) = \frac{\mathrm{MSE}[\hat{m}(x_1, x_2)]}{\mathrm{MSE}[\hat{m}_{I_{1,2}}(x_1, x_2)]}$$

in this region for the regression function (2.30) and the functions

$$\begin{aligned}
m(x_1, x_2) &= \frac{1}{2}\left[x_1 + x_2 + \frac{1}{6\pi}\sin(6\pi(x_1 + x_2))\right], && (2.31)\\
m(x_1, x_2) &= \sin\left(\frac{\pi}{4}(x_1 + x_2)\right), && (2.32)\\
m(x_1, x_2) &= \left(2x_1x_2 + \frac{1}{2}x_1^2 - 2x_2^2 + 2x_2 - \frac{1}{2}\right) I\{1 \leq 2x_2 \leq x_1 + 1\} && (2.33)\\
&\quad + \left(2x_2x_1 - \frac{1}{2}x_1^2\right) I\{x_1 \leq 2x_2 < 1\} + 2x_2^2 I\{2x_2 < x_1\} + x_1 I\{x_1 + 1 \leq 2x_2\}.
\end{aligned}$$

Again the sample size is $n = 400$ and the errors are standard normal with constant variance $\sigma^2 = 0.1$. The ratio of the mean squared errors are 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 can be substantial. For example in model (2.30), there are several regions in $[0.1, 0.9]^2$, where the ratio is larger than 1.7. The improvement caused by the monotonization in model (2.31) and (2.32) is not so large but still clearly visible.
Note that model (2.33) corresponds to the distribution function of the random variable $Z = \frac{1}{2}(X_1 + X_2)$, where X_1, X_2 are independent with an uniform distribution on the interval $[0, 1]$. This model does not satisfy the assumptions of our theoretical results since for $2x_2 \geq x_1 + 1$, we have $m(x_1, x_2) = x_1$, and this function is not strictly increasing with respect to x_2. Similarly, we have $m(x_1, x_2) = 2x_2^2$ if $2x_2 < x_1$ in the model (2.33). Observe also that there are some (small) areas where the unconstrained local linear estimate yields a smaller mean squared error in model (2.33). Nevertheless, 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.

2.4.2 Data examples

To explore the performance of the strictly monotone estimates in terms of real data examples, we present to typical settings for monotonicity constraints in applications. These

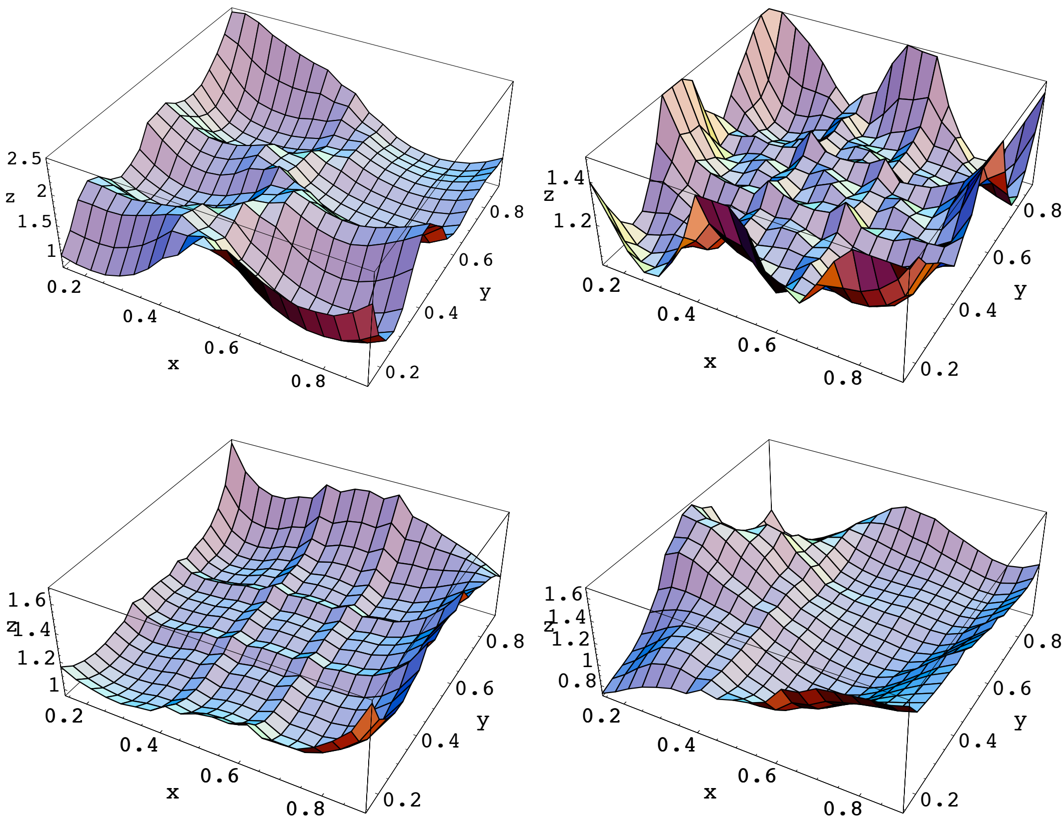

Figure 2.6: The simulated ratio of the mean squared errors of the local linear estimate and its monotonization with respect to the x_1 and x_2 coordinate. The four surfaces correspond to the regression functions (2.30): upper left panel, (2.31): upper right panel, (2.32): lower left panel, and (2.33): lower right panel.

examples are implemented with the R package monoProc [see Scheder (2007) in R Development Core Team (2007)] and can be reproduced by the reader. The difficulty in the implementation of nonparametric kernel techniques is to assign the parameters like the bandwidth or the kernel function.
As in the simulation study before, we apply the local linear estimator as unconstrained estimator for its superiority in the asymptotic behavior compared to the Nadaraya and Watson estimator. To get an estimate for the variance function σ in model (2.29), we use Spokoiny's variance estimate for high-dimensional regression models [see Spokoiny (2002)] to get an estimate $\hat{\sigma}^2$ of the integrated variance. The Epanechnikov kernel is used as kernel function. After that a two-step cross validation is implemented to determine the bandwidths $h_1 = h_2$ and h_m. In the first step, the bandwidths h_1, h_2 are chosen. The cross validation is performed as a leave-one-out-cross validation procedure over an appropriate interval centered at the point $\left(\frac{\hat{\sigma}^2}{n}\right)^{1/6}$ determining $h_1 = h_2$, where n is the sample size of the data. For more details on this method to assign the bandwidth for regression smoothers see Härdle (1990), p.152. In the second step of the cross validation, we determine the bandwidth h_m using the bandwidths $h_1 = h_2$ from the first step. This time the cross validation function is defined by

$$\mathrm{CV}(h_m) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{m}_{I,j}(X_{j1}, X_{j2}))^2$$

with $\hat{m}_{I,j}(x_1, x_2) =) = (\hat{m}_{I_{1,2},j}(x_1, x_2) + \hat{m}_{I_{2,1},j}(x_1, x_2))/2$ as the average of the two monotone estimates $\hat{m}_{I_{1,2},j}(x_1, x_2)$ and $\hat{m}_{I_{1,2},j}(x_1, x_2)$ (obtained by interchanging the order of isotoniziation) in which the jth observation is left out (denoted by the subscript j).

Example 2.23 A reasonable criterion for obesity is to calculate the body fat percentage of a person. The most accurate techniques like underwater weighing to obtain the body fat percentage are time consuming and expensive. Other methods to estimate this percentage suggest certain predictive equations depending on, e.g., weight, height, age, and several simple body measurements. The fat data set in the UsingR package contains these information of 252 men from 22 to 81 years [see Penrose et al. (1985)]. To simplify this problem, we consider weight and height as independent variables motivated by the body mass index (BMI)

$$BMI = \frac{\text{weight(kg)}}{\text{height(m)}^2}.$$

This is an restrictive simplification, but on the other hand there are also problems with the accuracy of body circumference measurements. Weight and height are in contrast to that, easily obtained quantities. Moreover, it is reasonable that in an average group of men, body fat should depend monotonically increasing on weight and monotonically decreasing on height. It has to be mentioned that two outliers in this data set have been disregarded (observation 39 and 42). For the fat data set, we obtain $\hat{\sigma}^2 = 39.48902$ using

Spokoiny's variance estimation. The cross validation function suggest the bandwidth $h_1 = h_2 = 0.5$ for the unconstrained local linear estimator. In the second cross validation, the bandwidth $h_m = 0.75$ for the monotonizing procedure is chosen out of the set $[0.1, 3.0]$ with $\mathrm{CV}(0.75) = 30.9634$. 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 $h_m = 0.75$ corresponds also the minimum of $\mathrm{CV}(h_m)$ if only $\hat{m}_{I_{1,2}}$ is under consideration. By comparison, the local linear estimator leads in this setup to a cross validation value of 30.952571. The monotone estimate of the regression function (body fat percentage as a function of weight and height) is depicted in Figure 2.7.

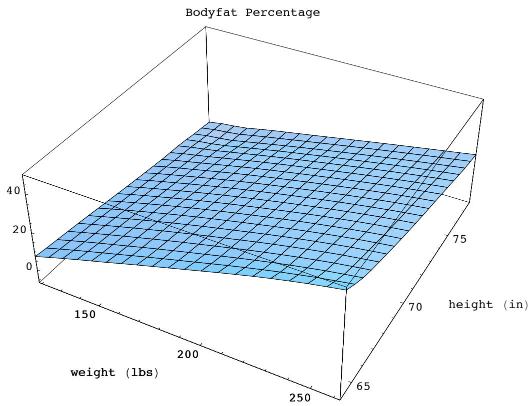

Figure 2.7: Monotone estimate of the body fat percentage as a function of weight and height.

Example 2.24 Often a monotone dependency between education and work experience and the corresponding salary is assumed. To examine the monotonic relationship between the frequency to of earning more than \$50 000 per year as response and education and age as explanatory variables, we analyze a reasonably clean extract of the 1994 Census data base found in Blake and Merz (1998) [see also Kohavi (1996)]. This data set 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 derive a data set with 639 observations, where age ranges from 19 to 90, and education level from 1 to 16. This extract includes only age groups in which at least one person earns more than \$50 000 per year. Furthermore, we restrict age to 27-70 and analyze therefore a sample of 516 observations. Spokoiny's variance estimation leads to $\hat{\sigma}^2 = 0.02563633$, and we obtain the bandwidths $h_1 = h_2 = 0.169$ by cross validation. Again we used a second cross validation to find the bandwidth h_m.
The cross validation function $CV(h_m)$ 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_m on the cross validation function is very small. We decided to use $h_m = 0.017$ for the further analysis, which is a minimizer of $CV(h_m)$ for the average of the two monotonized procedures $\hat{m}_I = (\hat{m}_{I_{1,2}} + \hat{m}_{I_{2,1}})/2$. But the difference to the minima of the estimates $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$, which are attained at $h_m = 0.013$ and in $h_m = 0.019$, respectively, is negligible.

$\hat{m}_I$	$\hat{m}_{I_{1,2}}$	$\hat{m}_{I_{2,1}}$	h_m	$\hat{m}_I$	$\hat{m}_{I_{1,2}}$	$\hat{m}_{I_{2,1}}$	h_m
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 2.1: Scores of the cross validation function for the different estimates with $h_1 = h_2 = 0.169$.

The estimates for the regression function are displayed in Figure 2.8. Since the difference between $\hat{m}_{I_{1,2}}$ and $\hat{m}_{I_{2,1}}$ 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.

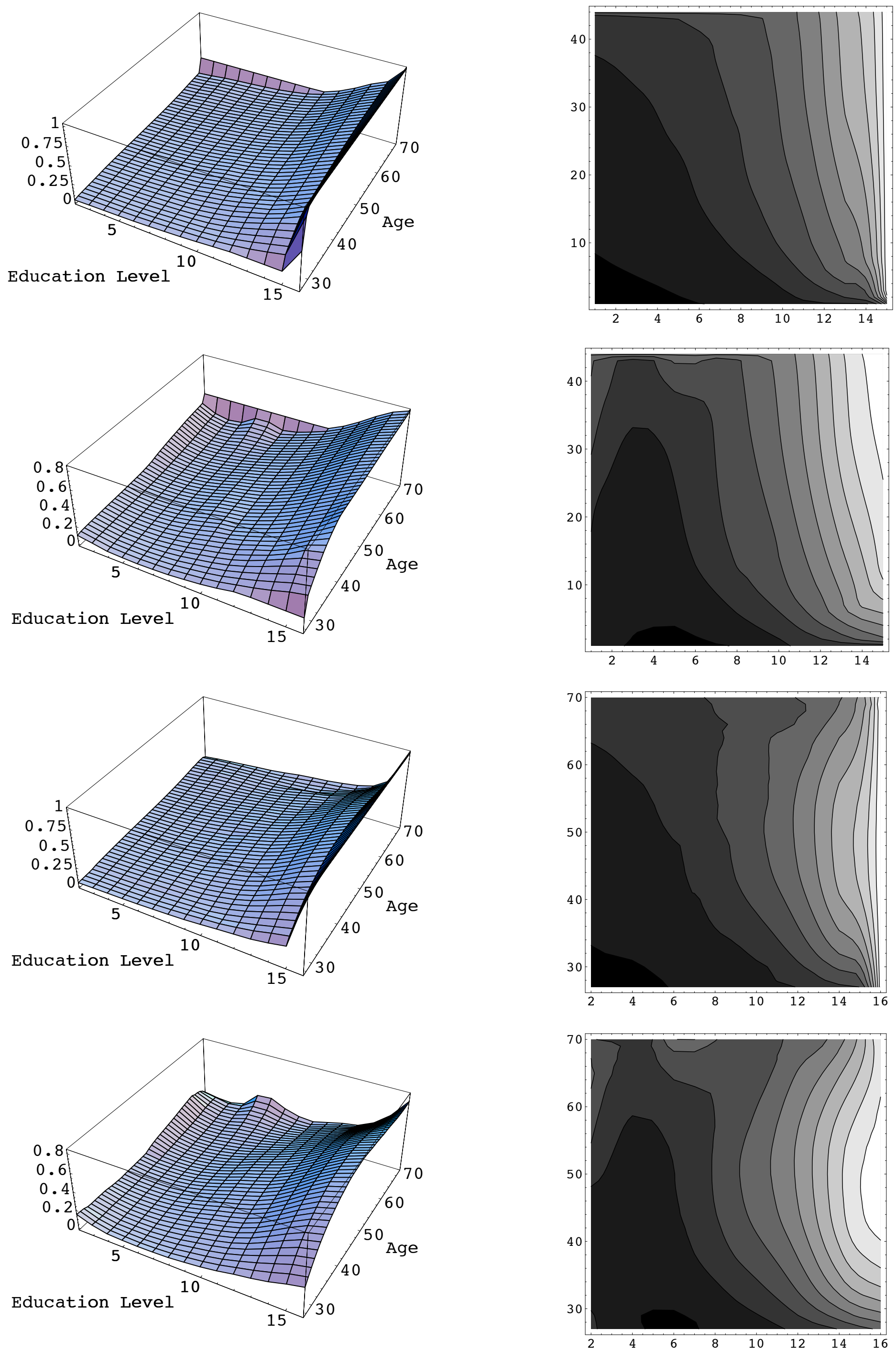

Figure 2.8: The results of the Census-Income Example. The estimates for the regression function and their corresponding Contour Plots from top to bottom with $h_1 = h_2 = 0.169$ and $h_m = 0.017$: $\hat{m}_I$, $\hat{m}_{I_1}$, $\hat{m}_{I_2}$, and $\hat{m}$.

Chapter 3

Additive quantile regression

3.1 Overview

This chapter is concerned with quantile regression models, which give a more sophisticated picture of a relationship between a predictor and a response variable. In particular, we consider additive conditional quantile models with high-dimensional predictors. In the following three sections, some fundamental aspects of the theory of conditional quantile models with high-dimensional covariates are summarized. First of all, the model itself is introduced. Then an estimator for the conditional quantile model is derived. Here we benefit again from the monotonizing procedure, since we use the first step to obtain an isotonized inverse of a possibly non-increasing estimate of the conditional distribution function [see Section 1.3 in Chapter 1]. To avoid the curse of dimensionality, additive models are utilized in Section 3.4. In Section 3.5, estimates for different contrasts in the additive conditional quantile model are developed and analyzed. Asymptotic normality of the estimates is established with the optimal one-dimensional rate of convergence. In the last section finally, some finite sample studies and data examples are given to illustrate the theory. Furthermore a comparison with the procedure introduced by De Gooijer and Zerom (2003) is conducted, which is most similar in spirit with the presented estimates.

3.2 Quantile regression models

In contrast to traditional regression models, quantile regression models offer a deeper insight into the relationship of the response variable Y and the one- or higher-dimensional explanatory variable $\mathbf{X}$. The regression model

$$Y = Q(\alpha|\mathbf{X}) + \varepsilon$$

is formulated as quantile model for fixed $\alpha \in (0,1)$. The error ε is defined as a random variable, whose α-quantile given $\mathbf{X} = \mathbf{x}$ is zero for almost every $\mathbf{x}$, so that $Q(\alpha|\mathbf{X})$ is the α-quantile of Y given $\mathbf{X} = \mathbf{x}$. The quantile model is specified for each α separately.
The median is the most common empirical quantile with $\alpha = 0.5$. For a simple numerical sample, it is defined as the number dividing the higher half of a sample, a population, or a probability distribution from the lower half. The median of a finite list of numbers can be found by ordering all observations from lowest value to highest value and picking the middle one. If there are an even number of observations, one often takes the mean of the two middle values. The median exemplifies easily why quantiles are more robust compared to the mean value of a sample. One outlier is enough to change the mean value substantial, whereas the median is not shifted noticeable. There are many data sets which suffer from this phenomena; e.g. salaries and wages are often skewed with few salaries at the extreme high end of the range, who have a strong influence on the average salary. This "average salary" conveys a false impression of how much an average person earns. The conditional quantile function of Y given $\mathbf{X} = \mathbf{x}$ allows to consider several quantiles and provides a deep insight into the relationship. In economics and finance, the value at risk (VaR) is a measure of market risk to a given confidence level $1-\alpha$, which is basically the quantile at α of the sample [see Pflug (2000) or Umantsev and Chernozhukov (2001)]. In general, there are two ways to define conditional quantiles.

Definition 3.1 *For fixed $\alpha \in (0,1)$, the α-conditional quantile of Y given $\mathbf{x} = (x_1, x_2, \ldots, x_d)^T$ is defined as the value $Q(\alpha|\mathbf{x})$ such that*

(i)

$$Q(\alpha|\mathbf{x}) = \inf\{t \in I\!R \mid F(t|\mathbf{x}) \geq \alpha\},$$

where $F(\cdot|\mathbf{x})$ denotes the conditional distribution function of Y given $\mathbf{X} = \mathbf{x}$ or equivalently,

(ii)

$$Q(\alpha|\mathbf{x}) = \arg\min_{a \in I\!R} E[\rho_\alpha(Y-a)|\mathbf{X} = \mathbf{x}],$$

where $\rho_\alpha(u) = |u| + (2\alpha - 1)u$ is the so-called "check-function".

Koenker and Bassett (1978) introduced quantile regression using the check-function as a supplement to least squares methods, which are used for estimating conditional mean functions. Their approach was a cornerstone in quantile regression. In the meantime, several authors worked on methodological and practical aspects of this method. Koenker (2005) summarized the results about quantile regression in a recent monograph.
Lately, nonparametric methods have found considerable interest in the context of conditional quantiles [see Yu and Jones (1998) or Yu and Jones (1997)]. Motivated by the observation that many one-dimensional nonparametric estimates of conditional quantiles are not monotone with respect to $\alpha \in (0,1)$ [see e.g. He (1997), Yu et al. (2003), or

Koenker (2005), Chapter 7], which is an embarrassing phenomenon in applications, we pursue a new approach. In the following, we construct the conditional quantile function via inversion of the distribution function. To estimate the conditional distribution function, we focus on nonparametric approaches. Many estimators of the distribution functions obtained by nonparametric regression techniques are not inherently monotone or positive, so the inverse cannot be obtained in a straightforward manner. We use the first step of the monotonizing procedure described in Chapter 1 to avoid these problems in a very practical way. The concept of non-increasing rearrangements has been successfully applied by Dette and Volgushev (2007) and Chernozhukov et al. (2007) in quantile regression. The last named authors use this concept to isotonize parametric (possibly crossing) quantile estimates. Dette and Volgushev (2007), however, isotonize and invert a nonparametric estimate of the conditional distribution function simultaneously in the context of a one-dimensional covariate. In the next section, we will extend this method to high-dimensional covariates. To deal with the curse of dimensionality, we consider additive models.

3.3 The conditional Quantile function

In this section, we discuss a method for estimating conditional quantile functions through reversing the conditional distribution function introduced by Dette and Volgushev (2007) in the case of a one-dimensional covariate. First of all, estimators for distribution functions are analyzed. Given the random sample $\{(\mathbf{X}_j, Y_j)\}_{j=1}^n$, where the explanatory variable $\mathbf{X}_i$ is a d-dimensional vector, the estimation problem can be viewed as a regression of Z_j on $\mathbf{X}_j$, where $Z_j = I\{Y_j \leq y\}$ and $E[Z_j|\mathbf{X}_j = \mathbf{x}] = P(Y_i \leq y|\mathbf{X}_i = \mathbf{x}) = F(y|\mathbf{x})$. So we can consider

$$Z_j = F(y|\mathbf{X}_j) + \sigma(y|\mathbf{X}_j)\varepsilon_j \tag{3.1}$$

as a nonparametric regression model, where the variance function depends on y [see Hall et al. (1999)]. The variance function $\sigma(y|\mathbf{x})$ can be specified in terms of $F(y|\mathbf{x})$. Precisely, we have

$$\sigma^2(y|\mathbf{x}) = E[(Z_j - F(y|\mathbf{x}))^2|\mathbf{X}_j = \mathbf{x}] = F(y|\mathbf{x})(1 - F(y|\mathbf{x})),$$

since $Z_j^2 = Z_j$. In this framework, many estimators for the conditional distribution function arise from the theory of nonparametric estimation techniques. Using kernel regression estimators, we derive a Nadaraya-Watson type estimator

$$\hat{F}_{NW}(y|\mathbf{x}) = \frac{\sum_{j=1}^n K_H(\mathbf{X}_j - \mathbf{x})I\{Y_j \leq y\}}{\sum_{j=1}^n K_H(\mathbf{X}_j - \mathbf{x})} \tag{3.2}$$

as in Chapter 2. In a similar manner the local linear estimator of the conditional distribution function is defined. In this chapter, we define K as a d-dimensional kernel of order $q \geq 2$ supported on $[-1, 1]^d$. Let $\nu = (\nu_1, \ldots, \nu_d)$ be a multi-index of integers with $\nu_l \geq 0$, so that $\mathbf{x}^\nu = x_1^{\nu_1} \ldots x_d^{\nu_d}$. Moreover, define $|\nu| = \sum_{l=1}^d \nu_l$. Then the kernel K satisfies the following conditions

(i) $\int_{[-1,1]^d} K(\mathbf{x})d\mathbf{x} = 1$,

(ii) $\int_{[-1,1]^d} |\mathbf{x}^\nu||K(\mathbf{x})|d\mathbf{x} < \infty$ for $|\nu| \leq q$,

(iii) $\int_{[-1,1]^d} \mathbf{x}^\nu K(\mathbf{x})d\mathbf{x} = 0$ for $1 \leq |\nu| \leq q-1$,

(iv) $\int_{[-1,1]^d} \mathbf{x}^\nu K(\mathbf{x})d\mathbf{x} \neq 0$ for some $|\nu| = q$.

Sometimes one assumes that the kernel function K is symmetric, i.e.,

$$K(-\mathbf{x}) = K(\mathbf{x}), \quad \mathbf{x} \in [-1,1]^d,$$

which implies that q is even. Note that the kernel K is not a probability density function anymore for $q > 2$. In particular, the kernel K is not positive anymore. Therefore we run into the problem that the estimate $\hat{F}_{NW}(y|\mathbf{x})$ violates the isotonicity constraint of a distribution function with respect to y. The same thing happens for the local linear estimate $\hat{F}_{LL}(y|\mathbf{x})$, since the local linear estimate can be expressed as a weighted average, where the weights are not necessarily positive even for a positive kernel K of order $q = 2$ [see Remark 1.5 in the univariate case]. Hall et al. (1999) suggest a reweighted Nadaraya and Watson estimator for the conditional distribution function with a kernel of order 2 achieving the more attractive bias of the local linear estimate. Basically, their estimate has the following form in the case of a one-dimensional covariate

$$\hat{F}_{RW}(y|x) = \frac{\sum_{j=1}^n p_j(x)K_h(x-X_j)I\{Y_j \leq y\}}{\sum_{j=1}^n p_j(x)K_h(x-X_j)},$$

where $p_j(x)$ denote weights depending on the data $X_1, \ldots, X_n$ satisfying $p_j \geq 0$ for $j = 1, \ldots, n$, $\sum_{j=1}^n p_j = 1$, and

$$\sum_{j=1}^n p_j(x)(x-X_j)K_h(x-X_j) = 0.$$

The weights p_j are chosen by maximizing $\prod_{j=1}^n p_j$.
Internal methods yield other estimates for the conditional distribution function. In the context of estimating an additive quantile function with high-dimensional covariates, the internalized estimator of the conditional distribution function is interesting from a computational point of view, that is

$$\hat{F}_{INW}(y|\mathbf{x}) = \sum_{j=1}^n \frac{K_H(\mathbf{x}-\mathbf{X}_j)I\{Y_j \leq y\}}{\sum_{i=1}^n K_H(\mathbf{X}_j - \mathbf{X}_i)}$$

[see Jones et al. (1994) or Kim et al. (1999)]. The main advantage of the internalized estimator is observable by writing it as a weighted sum

$$\hat{F}_{INW}(y|\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} K_H(\mathbf{x}-\mathbf{X}_j)\tilde{Y}_j$$

over the adjusted data $\tilde{Y}_j = I\{Y_j \le y\}/\hat{p}(\mathbf{X}_j)$, where

$$\hat{p}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_H(\mathbf{x}-\mathbf{X}_i).$$

On the other hand, as we have seen in Theorem 1.4 of Chapter 1, this estimate yields an additional bias and variance term.

There are other estimators for the conditional distribution function, but we will focus on the above mentioned nonparametric kernel estimators. The aim of this chapter is to derive an estimator for the conditional quantile function with many covariates from an estimator of the conditional distribution function as motivated by Definition 3.1. With this approach we have to deal with the problem that some estimates of the conditional distribution function are not monotone increasing with respect to y and the generalized inverse cannot be computed. To overcome this defect of some estimators, we propose a method which deals simultaneously with this lack and the problem of inversion. We apply the first step of the monotonizing procedure, which produces a monotonized inverse of the estimator of the conditional distribution function [see Section 1.3.2]. In order to fix ideas, let $G : \mathbb{R} \to [0,1]$ be a strictly increasing distribution function, which will be used as a transformation to the compact interval $[0,1]$, since $F(\cdot|\mathbf{x})$ might have unbounded support. The kernel K_m denotes a positive kernel with compact support on $[-1,1]$, and h_m denotes a bandwidth, then we define for a nonparametric estimator of the conditional distribution function $\hat{F}(y|\mathbf{x})$

$$\hat{G}_I(\alpha|\mathbf{x}) = \frac{1}{Nh_m}\sum_{i=1}^{N}\int_{-\infty}^{\alpha} K_m\left(\frac{\hat{F}(G^{-1}(\frac{i}{N})|\mathbf{x}) - u}{h_m}\right) du. \tag{3.3}$$

If $\hat{F}(y|\mathbf{x})$ is uniformly consistent and $N \to \infty$, $h_m \to 0$, it is intuitively clear that

$$\begin{aligned}\hat{G}_I(\alpha|\mathbf{x}) \;&\approx\; G_N(\alpha|\mathbf{x}) := \frac{1}{Nh_m}\sum_{i=1}^{N}\int_{-\infty}^{\alpha} K_m\left(\frac{F(G^{-1}(\frac{i}{N})|\mathbf{x}) - u}{h_m}\right) du \\ &\approx\; \int I\{F(G^{-1}(s)|\mathbf{x}) \le \alpha\}ds = G(Q(\alpha|\mathbf{x})),\end{aligned}$$

where $Q(\alpha|\mathbf{x}) := F^{-1}(\alpha|\mathbf{x})$. The last approximation follows for $h_m \to 0$ and $N \to \infty$ sufficiently fast as motivated before in Section 1.3.1. Consequently, we define

$$\hat{Q}_I(\alpha|\mathbf{x}) = G^{-1}(\hat{G}_I(\alpha|\mathbf{x})) \tag{3.4}$$

as the estimate of the conditional quantile $Q(\alpha|\mathbf{x})$, and

$$Q_N(\alpha|\mathbf{x}) = G^{-1}(G_N(\alpha|\mathbf{x})) \tag{3.5}$$

as a deterministic approximation of the conditional quantile $Q(\alpha|\mathbf{x})$. It will be demonstrated in the following sections that the choice of the function G has no impact on the asymptotic properties of the estimate. See Dette and Volgushev (2007) for a detailed proof. Moreover, even for realistic sample sizes the impact of the choice of G is negligible. Some practical recommendations regarding this choice will be given in Section 3.6. Note that the estimate $\hat{G}_I(\alpha|\mathbf{x})$ and $\hat{Q}_I(\alpha|\mathbf{x})$, respectively, are monotone increasing with respect to α provided that the kernel K_m is positive on its support.

3.4 Additive Models

Additive Models are a natural generalization of linear regression models. Let Y be the dependent variable and $\mathbf{X}$ the d-dimensional vector of explanatory variables. The additive structure of the regression function can be modeled by

$$m(\mathbf{X}) = g_1(X_1) + g_2(X_2) + \ldots + g_d(X_d) + c,$$

provided

$$E[g_k(X_k)] = \int g_k(x_k)p_k(x_k)dx_k = 0 \ \ (k = 1, \ldots, d)$$

for identifiability of the individual components, where p_k is the marginal density of X_k. These identification assumptions yield $E[Y] = c$. For high-dimensional estimation problems, additive models circumvent the curse of dimensionality. Stone (1985) showed that the additive regression function can be estimated with the optimal convergence rate of the one-dimensional case. This rate is $n^{-s/(2s+1)}$, where s is an index of smoothness of m, explicitly it indicates how often m is differentiable [usually one assumes $s = 2$].
The conditional quantile model focused in this chapter can be modeled through an additive model for $\alpha \in (0, 1)$ by

$$Q(\alpha|\mathbf{x}) = Q_1(\alpha|x_1) + Q_2(\alpha|x_2) + \ldots + Q_d(\alpha|x_d) + c(\alpha), \tag{3.6}$$

where for identifiability

$$E[Q_k(\alpha|X_k)] = 0 \ \ (k = 1, 2, \ldots, d) \ \text{ and } \ E[Q(\alpha|X)] = c(\alpha). \tag{3.7}$$

In contrast to the additive regression model, note that the constant term depends on α. Several authors have recommended this model for high-dimensional quantile regression [see e.g. Doksum and Koo (2000), De Gooijer and Zerom (2003), and Horowitz and Lee (2005) among others].

Essentially, there are three approaches to estimate the additive components of such models. The backfitting algorithm is proposed by Breiman and Friedman (1985) and Buja et al. (1989). Tjøstheim and Auestad (1994) and Linton and Nielsen (1995) introduced a marginal integration method. Finally, Andrews and Whang (1990) and Li (2000) considered splines to estimate additive models. An overview of the additive models and marginal effects can be found in Chapter 8 of Härdle et al. (2004).
In the following, we summarize the backfitting approach and the marginal integration method in the context of regression models. Both methods can be combined with kernel regression techniques, but are substantially different from each other.

3.4.1 Backfitting Algorithm

Originally, the backfitting algorithm was developed by Breiman and Friedman (1985) and Buja et al. (1989). It is a widely-used method to approximate the additive components. Hastie and Tibshirani (1990) motivate the backfitting algorithm as follows. Consider the optimization problem

$$\min_m E\left[(Y - m(\mathbf{X}))^2\right] \text{ such that } m(\mathbf{X}) = c + \sum_{l=1}^{d} g_l(X_l).$$

The solution of this problem is the best (in terms of the expected squared distance) additive predictor for $E[Y|\mathbf{X}]$. By the projection onto X_k, the optimization problem changes to

$$E[(Y - m(\mathbf{X}))|X_k] = 0,$$

which yields

$$g_k(X_k) = E\left[\left(Y - c - \sum_{l \neq k} g_l(X_l)\right)\Bigg|X_k\right] \tag{3.8}$$

for $k = 1, \ldots, d$. This system of equations can be formulated equivalently as

$$\begin{pmatrix} I & P_1 & P_1 & \cdots & P_1 \\ P_2 & I & P_2 & \cdots & P_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_d & P_d & P_d & \cdots & I \end{pmatrix} \begin{pmatrix} g_1(X_1) \\ g_2(X_2) \\ \vdots \\ g_d(X_d) \end{pmatrix} = \begin{pmatrix} P_1 Y - c \\ P_2 Y - c \\ \vdots \\ P_d Y - c \end{pmatrix}, \tag{3.9}$$

where $P_k = E[\cdot|X_k]$. For the empirical version of (3.9), the conditional expectation P_k is replaced by smoothers S_k on X_k and c by $\hat{c} = \bar{Y}$. The smoother matrices S_k depend only on the component X_k and change the problem to a one-dimensional smoothing problem. More precisely, the basic backfitting algorithm works as follows.

Backfitting Algorithm	
initialization	$\hat{c} = \bar{Y}, \hat{g}_k^{(0)} \equiv 0$ for $k = 1, \ldots, d$,
repeat	for $k = 1, \ldots, d$ the cycles
	$r_k = Y - \hat{c} - \sum_{l=1}^{k-1} \hat{g}_l^{(s+1)} - \sum_{l=k+1}^{d} \hat{g}_l^{(s)}$,
	$\hat{g}_k^{(s+1)}(\cdot) = S_k(r_k)$
until	convergence is reached

The backfitting estimate is an iterative solution and therefore not easy to analyze theoretically. Mammen et al. (1999) proposed a smoothed backfitting version, where they can develop asymptotic theory for their backfitting estimators. They write (3.8) as

$$g_k(X_k) = E[Y|X_k] - c - \sum_{l \neq k} E[g_l(X_l)|X_k]$$

and estimate the conditional expectation "correctly" in some sense. The population quantity $E[Y|X_k]$ is replaced by one-dimensional nonparametric estimates. For the Nadaraya-Watson estimate $\hat{m}$, it can be shown in a straightforward manner that

$$\int \hat{m}(\mathbf{x}) \frac{\hat{p}(\mathbf{x})}{\hat{p}_k(x_k)} dx_{\underline{k}} = \frac{\frac{1}{n}\sum_{j=1}^{n} K_h(x_k - X_{jk})Y_j}{\hat{p}_k(x_k)} = \hat{m}_k(x_k),$$

where $\hat{p}(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{l=1}^{d} K_h(x_l - X_{jl})$ is the empirical pdf of $\mathbf{X}$ and $\hat{p}_k(x_k)$ the corresponding empirical pdf of X_k for $k = 1, \ldots, d$. So the above equation system becomes to

$$g_k(x_k) = \hat{m}_k(x_k) - \bar{Y} - \sum_{l \neq k} \int g_l(x_l) \frac{\hat{p}(x_k, x_{\underline{k}})}{\hat{p}_k(x_k)} dx_{\underline{k}},$$

which means that the smoothing is done over the whole vector $\mathbf{X}$ instead of using only the subvector X_k as in the classical backfitting. This estimator achieves full oracle efficiency which means that it has the same bias and variance as the oracle estimator based on knowing the other components. A similar result can be attained for a local linear version [see Mammen et al. (1999) or Nielsen and Sperlich (2005)]. Since the backfitting approach relies on one-dimensional smooths, it is free from the curse of dimensionality. In particular, the smoothed backfitting estimator features the univariate rate of convergence.

3.4.2 Marginal Integration Approach

The basic idea of marginal integration is motivated again by looking at the identifiability assumptions. For $k = 1, \ldots, d$, it is assumed that

$$E[g_k(X_k)] = \int g_k(x) p_k(x) dx = 0,$$

where p_k is the marginal density of X_k. Let

$$X_{\underline{k}} = (X_1, \ldots, X_{(k-1)}, X_{(k+1)}, \ldots, X_d)$$

be the vector of all explanatory variables but X_k and $p_{\underline{k}}$ their joint pdf. For the additive regression function, we have

$$\begin{aligned}
\int m(\mathbf{x}) p_{\underline{k}}(x_{\underline{k}}) \prod_{l \neq k} dx_l &= E_{X_{\underline{k}}}[m(X_k, X_{\underline{k}})] \\
&= E_{X_{\underline{k}}}[\sum_{l \neq k} g_l(X_l) + g_k(X_k) + c] \\
&= g_k(X_k) + c.
\end{aligned}$$

The fundamental idea of the marginal integration approach is substantially different from the backfitting algorithm. If the underlying true model is not exactly additive, the two methods estimate different things. The backfitting algorithm fits the best additive model to the data, whereas the marginal impact is estimated by the marginal integration method. This marginal effect of X_k shows how Y changes on average while varying X_k. In the model

$$Y = m(\mathbf{X}) + \varepsilon = g_1(X_1) + g_2(X_2) + \ldots + g_d(X_d) + c + \varepsilon,$$

the marginal effect is the conditional expectation $E[Y|X_k]$, where the expectation is taken on the error distribution as well as on all other regressors. For estimation, we use this relation. First estimate the multidimensional regression function, which is denoted by $\hat{m}$, then integrate out the variables different from X_k. To create a general framework, let W be a deterministic weighting function with $\int dW(x) = 1$. We allow W to be discrete or continuous, so that the density w of W is either Lebesgue or a counting measure. In this case, the following contrast can be estimated

$$\gamma_W(x_k) = \int m(x_k, x_{\underline{k}}) dW(x_{\underline{k}}) = g_k(x_k) + c_k,$$

where $c_k = \int \sum_{l \neq k} g_l(x_l) dW(x_{\underline{l}})$ and $x_{\underline{k}} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_d)^T$.
A special case of W is the marginal probability measure of $X_{\underline{k}}$. In a typical estimation problem, this measure is unknown and the empirical distribution of $X_{\underline{k}}$ is used instead. In these cases, the constants $c_k = c$ are independent of W, which follows from the identifiability assumptions.
For the empirical distribution of $X_{\underline{k}}$, we obtain

$$\widehat{g_k(x_k) + c} = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(x_k, X_{i\underline{k}})$$

as estimator of the marginal effects.

3.5 Estimates of additive conditional quantiles

In this section, we utilize the marginal integration method to estimate contrast in quantile models. As underlying true model, we assume an additive quantile model of the form

$$Q(\alpha|\mathbf{x}) = Q_1(\alpha|x_1) + Q_2(\alpha|x_2) + \ldots + Q_d(\alpha|x_d) + c(\alpha),$$

where $\alpha \in (0,1)$ and $\mathbf{x} = (x_1, \ldots, x_d)^T$. As a first approach, we start with a $d-1$-dimensional known weighting function W on $x_{\underline{k}}$, the vector of all variables but x_k, and define the following contrast

$$\begin{aligned} \gamma_W(\alpha|x_k) &= \int Q(\alpha|\mathbf{x})W(dx_{\underline{k}}) = \int Q(\alpha|x_k, x_{\underline{k}})W(dx_{\underline{k}}) \\ &= Q(\alpha|x_k) + c_k(\alpha), \end{aligned} \tag{3.10}$$

where

$$c_k(x_k) = \int \sum_{l \neq k} Q_l(\alpha|x_l) dW(x_{\underline{k}}).$$

In Section 3.3, estimators for the conditional quantile function are introduced as the inverse of the distribution estimates. Recall the definition for this estimate in (3.4). In the case of a conditional distribution function $F(y|\mathbf{x})$ supported on the compact interval $[0,1]$, the function G corresponds to the uniform distribution, and the estimate (3.4) simplifies to

$$\hat{Q}_I(\alpha|x_k, x_{\underline{k}}) = \frac{1}{Nh_m} \sum_{i=1}^{N} \int_{-\infty}^{\alpha} K_m\left(\frac{\hat{F}(\frac{i}{N}|x_k, x_{\underline{k}}) - u}{h_m}\right) du,$$

where $\hat{F}(\frac{i}{N}|x_k, x_{\underline{k}})$ is a nonparametric estimator of the distribution function for instance a local polynomial estimator. For the sake of transparency, we restrict ourselves to the classical Nadaraya-Watson estimate, which is a local constant estimator [see definition (3.2)]. To distinguish between the variable of interest and the other ones, we use two different kernels, i.e.

$$\hat{F}_{NW}(y|x_k, x_{\underline{k}}) = \frac{\sum_{i=1}^{n} K_{h_1}(x_k - X_{ik}) L_{H_2}(x_{\underline{k}} - X_{i\underline{k}}) I\{Y_i \leq y\}}{\sum_{j=1}^{n} K_{h_1}(x_k - X_{ik}) L_{H_2}(x_{\underline{k}} - X_{i\underline{k}})},$$

where $K(\cdot)$ and $L(\cdot)$ are the kernel functions. $L(\cdot)$ is a $d-1$-dimensional kernel. h_1 is the bandwidth of the variable x_k and $H_2 = \mathrm{diag}(h_2, \ldots, h_2)$ is a diagonal $(d-1)\times(d-1)$-matrix with h_2 as diagonal elements, which is the bandwidth of the other variables.
To estimate (3.10) using the marginal integration method, we consider

$$\hat{\gamma}_W(\alpha|x_k) = \int \hat{Q}_I(\alpha|x_k, x_{\underline{k}})W(dx_{\underline{k}})$$

as an estimate and

$$\gamma_{W,N}(\alpha|x_k) \;=\; \int Q_N(\alpha|x_k, x_{\underline{k}})W(dx_{\underline{k}})$$

as an approximation of (3.10) [see (3.5) for a definition of $Q_N(\alpha|\mathbf{x})$]. We show in the following subsection that this estimate converges at the optimal one-dimensional rate. Through the monotonizing inversion step, we have to deal with an additional bandwidth h_m, which makes things a bit more complicated than in a usual marginal integration application. It will turn out that even for the two-dimensional case one has to trade off between taking $h_1 = h_2 = O(n^{-1/5})$ and keeping the approximation error between $\gamma_W(\alpha|x_k)$ and $\gamma_{W,N}(\alpha|x_k)$ small. So the monotonizing inversion adds some problems. On the other hand, all applications of marginal integration in a nonparametric setting run into this problem. For higher dimensions it is not possible to use h_2 at the rate $n^{-1/5}$. We explain this phenomenon in the following in deeper details.

3.5.1 Asymptotic Theory

To derive the asymptotic behavior of $\hat{\gamma}_W(\alpha|x_k)$, some conditions on the model, the bandwidths h_m, h_1, and h_2, and the kernels $K_m(\cdot)$, $K(\cdot)$, and $L(\cdot)$ are required. We assume model (3.1) for the conditional distribution function $F(y|\mathbf{x})$. Many model assumptions are related to the ones in Chapter 2, and we emphasize the analogy.

(A1′) $\mathbf{X}_j$, $j = 1, \ldots, n$ is an i.i.d. sample with a q times continuously differentiable positive density, say p, supported on $[0,1]^d$, i.e. $p \in C^q([0,1]^d)$. The partial derivatives with respect to the covariates $\mathbf{x} = (x_1, \ldots, x_d)^T$ are denoted by $\frac{\partial^s}{\partial x_k^s}$ $(s = 1, \ldots, q;\; k = 1, \ldots, d)$.

(A2′) For any $\mathbf{x} \in [0,1]^d$ and $F(\cdot|\mathbf{x}) : D \to [0,1]$ with $D \subseteq I\!R$, we assume $F(\cdot|\mathbf{x}) \in C^1(D)$ and $Q'(\alpha|\mathbf{x}) > 0$, where the function Q' denotes the derivative of the quantile function $Q(\alpha|\mathbf{x})$ with respect to α. Its existence in a neighborhood of the quantile of interest is assumed.

(A3′) For any $y \in D$, we have $F(y|\cdot) \in C^q([0,1]^d)$ and again the partial derivatives with respect to x_k are denoted by $\frac{\partial^s}{\partial x_k^s}$ for $s = 1, \ldots, q$ and $k = 1, \ldots, d$.

(A4′) The random error ε_j and $\mathbf{X}_j$ are independent, and $E[\varepsilon_j] = 0$, $E[\varepsilon_j^2] = 1$, and $E[\varepsilon_j^4] < \infty$.

(A5′) The kernels $K(\cdot)$, $L(\cdot)$, and $K_m(\cdot)$ have compact support. $K(\cdot)$ is a one-dimensional kernel of order 2, and $L(\cdot)$ is a $(d-1)$-dimensional kernel of order q [see Section 3.3]. The order q is determined by the bandwidth conditions. Recall the definitions of the scalars $\kappa_2(K) = \frac{1}{2}\int v^2 K(v)dv$ and $||K||_2^2$, which is the squared L_2-norm of K. $K_m(\cdot)$ is supposed to be a positive kernel of order 2 and twice continuously differentiable. Further $K_m'(\cdot)$ is assumed to be Lipschitz-continuous.

The bandwidths h_1, h_2, and h_m fulfill the following conditions

(B1′) $nh_1 \to \infty,\ \ nh_2^{d-1} \to \infty,\ \ nh_1h_2^{d-1} \to \infty$, and $nh_m \to \infty$

(B2′) $nh_1^5 = O(1), N = O(n)$

(B3′) $\frac{h_m}{h_1} = o(1)$

(B4′) $nh_2^{2q+1} = O(1)$

(B5′) $\frac{1}{nh_1h_2^{2(d-1)}h_m^2} = o(1)$

The last bandwidth condition (B5′) determines the order q of the kernel $L(\cdot)$. Under the above assumptions, we can state the following theorem.

Theorem 3.2 *If the assumptions (A1′)-(A5′) and (B1′)-(B5′) are satisfied, then we have for all* $\alpha \in (0,1)$ *and* $k = 1, \dots, d$

$$\sqrt{nh_1}(\hat{\gamma}_W(\alpha|x_k) - \gamma_W(\alpha|x_k) + b_k(\alpha|x_k)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_k^2(\alpha|x_k)),$$

where

$$\begin{aligned} b_k(\alpha|x_k) &= \kappa_2(K)h_1^2 \int \Bigg[\frac{\partial^2}{\partial x_k^2} F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}}) \\ &\quad + 2\frac{\frac{\partial}{\partial x_k} F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}}) \frac{\partial}{\partial x_k} p(x_k, x_{\underline{k}})}{p(x_k, x_{\underline{k}})} \Bigg] \frac{1}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})} dW(x_{\underline{k}}), \\ s_k^2(\alpha|x_k) &= ||K||_2^2 \int \frac{\alpha(1-\alpha)w^2(x_{\underline{k}})}{(F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}}))^2 p(x_k, x_{\underline{k}})} dx_{\underline{k}}. \end{aligned}$$

Remark 3.3 At this point it might be appropriate to explain the conditions regarding the bandwidths to a greater extent. The condition $nh_1h_2^{d-1} \to \infty$ is crucial for the marginal integration approach. For $d > 4$ it is not possible to choose $h_2 = O(n^{-1/5})$. Therefore, in some respect this approach still suffers from the curse of dimensionality. The way out is to reduce the bias in the directions not of interest by taking L to be a higher order kernel and over-smooth the variables $x_{\underline{k}}$ by using h_2 at the rate $n^{-\frac{1}{2q+1}}$ for $q > 2$. Otherwise, the bias term in the directions not of interest dominates the asymptotic properties of the estimate.
The condition (B3′) is necessary to get the approximation error between $\gamma_W(\alpha|x_k)$ and $\gamma_{W,N}(\alpha|x_k)$ small. It is possible to reduce this condition to $\frac{h_m}{h_1} = O(1)$, but in this case one gets an additional bias-term. There is a remark after the proof of the above theorem about how to weaken this bandwidth condition. The last bandwidth condition determines the order q of the kernel L and depends on h_m as well. Through the additional smoothing

parameter h_m, even for $d = 2$ one has to use higher order kernels for the variable not of interest. One can argue that this estimate suffers more from the curse of dimensionality than usual marginal integration estimates. But on the other hand, all other approaches in the additive quantile regression suffer from similar or other problems so far. De Gooijer and Zerom (2003) suggest an estimate which uses a reweighted Nadaraya-Watson estimate with positive kernel functions as estimate for the conditional distribution function, and obtain the superior bias properties of the local linear estimates. This estimates is monotone increasing because of the positive kernels. For $d > 4$, it is not possible to use positive kernel functions, since one has to use higher order kernels due to the condition $nh_1h_2^{d-1} \to \infty$. This means that their estimate works only for $d \leq 4$. Other approaches which apply the check function have the difficulty of crossing quantile curves. In other words, despite the drawbacks of our estimate it is easy applicable, works for $d > 4$, and produces monotone increasing contrast estimates for additive quantile models.

Remark 3.4 The theorem is formulated with the Nadaraya-Watson estimator as initial estimate for the conditional distribution function. There are many alternative estimators which can be used. Hall et al. (1999) proposed a reweighted Nadaraya-Watson estimate for the conditional distribution function, which comes with the superior bias properties of the local linear estimates. Local polynomial techniques provide tools for many other nonparametric estimators. For all these estimators, similar results can be formulated. In particular, if the conditional distribution function is estimated by a local linear estimate [see Masry and Fan (1997)]. Then the asymptotic normality of the resulting estimate is still true, but the bias term in Theorem 3.2 has to be replaced by

$$b_k(\alpha|x_k) = \kappa_2(K)h_1^2 \int \frac{\frac{\partial^2}{\partial x_k^2} F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})} dW(x_{\underline{k}}).$$

In the usual marginal integration applications, one often refers to *internally normalized* multivariate regression smoother as initial estimators. These estimates are computationally easier to integrate than the external estimates where the Nadaraya-Watson estimator belongs to. We discuss this extensively in the next subsection.

Remark 3.5 Note that we can relax the assumption of independent data. In a more general setup, we assume that the process $\{(\mathbf{X}_j, Y_j)\}_{-\infty}^{\infty}$ is α-mixing or strongly mixing, that is

$$\sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_k^{\infty}} |P(AB) - P(A)P(B)| = \alpha(k) \to 0 \quad \text{as } k \to \infty\ ,$$

where $\mathcal{F}_a^b$ denotes the σ-algebra generated by the random variables $\{(\mathbf{X}_j, Y_j), a \leq j \leq b\}$ [see Rosenblatt (1956)]. To retain the assertion of Theorem 3.2 for dependent data, we assume that the mixing coefficients $\alpha(k)$ fulfill

$$\sum_{j=1}^{\infty} j^a (\alpha(j))^{1/2} < \infty$$

for $a > \frac{1}{2}$ and that there exists a sequence $\{v_n\}$ of positive integers satisfying $v_n \to \infty$ and $v_n = o\left(\sqrt{nh_1^d}\right)$ such that

$$\sqrt{\frac{n}{h_1^d}}\alpha(v_n) \to 0, \quad \text{as } n \to \infty.$$

With these two additional assumptions, the assertion of Theorem 3.2 remains valid. For more details and a discussion of strongly mixing data in multivariate nonparametric regression settings see Masry (1996).

3.5.2 Non-crossing estimates of the marginal effects

In this section, we focus on estimates for the marginal effects $q_k(\alpha|x_k) = Q_k(\alpha|x_k) + c(\alpha)$ for $k = 1, \dots, d$ in the additive conditional quantile model. In the last section, the marginal integration estimates for a deterministic weighting function W were analyzed. Now we require a weighting function to estimate the marginal effects explicitly. Since the distribution $p_{\underline{k}}$ of $X_{\underline{k}}$ is unknown, we use the empirical version of it which yields the following estimate

$$\hat{q}_k(\alpha|x_k) = Q_k\widehat{(\alpha|x_k) + c}(\alpha) = \frac{1}{n}\sum_{j=1}^{n} \hat{Q}_I(\alpha|x_k, X_{j\underline{k}}),$$

where $\hat{Q}_I(\alpha|x_k, X_{j\underline{k}})$ is defined in (3.4). The estimate $\hat{q}_k(\alpha|x_k)$ can be regarded as the expectation of $\hat{Q}_I(\alpha|\mathbf{X})$ with respect to the empirical distribution of

$$X_{\underline{k}} = (X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_d)^T.$$

It is obviously monotone in α for fixed x_k. Note that by the strong law of large numbers and from the normalizing condition (3.7), we have

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} Q_N(\alpha|x_k, X_{j\underline{k}}) \quad &\xrightarrow{\text{a.s.}} \quad \int Q(\alpha|\mathbf{x})p_{\underline{k}}(x_{\underline{k}})dx_{\underline{k}} = Q_k(\alpha|x_k) + c(\alpha) =: q_k(\alpha|x_1), \\
\frac{1}{n}\sum_{j=1}^{n} Q_k(\alpha|X_{jk}) \quad &\xrightarrow{\text{a.s.}} \quad \int Q_k(\alpha|x_k)p_k(x_k)dx_k = E[Q_k(\alpha|X_k)] = 0,
\end{aligned}$$

where $p_{\underline{k}}$ denotes the marginal density of $X_{\underline{k}} = (X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_d)^T$ and p_k the marginal density of X_k. Consequently, if $\hat{Q}_I(\alpha|\mathbf{x})$ is a (uniformly) consistent estimate of $Q(\alpha|\mathbf{x})$ it follows that $\hat{q}_k(\alpha|x_k)$ is a consistent estimate of $q_k(\alpha|x_k) := Q_k(\alpha|x_k) + c(\alpha)$. Eventually,

$$\hat{Q}_k(\alpha|x_k) = \hat{q}_k(\alpha|x_k) - \frac{1}{n}\sum_{j=1}^{n} \hat{q}_k(\alpha|X_{jk})$$

defines a consistent estimate of $Q_k(\alpha|x_k)$. An estimate of the additive quantile function in (3.6) is given by

$$\hat{Q}_{\text{add}}(\alpha|\mathbf{x}) := \sum_{k=1}^{d} \hat{q}_k(\alpha|x_k) - \left(1 - \frac{1}{d}\right) \sum_{k=1}^{d} \frac{1}{n} \sum_{j=1}^{n} \hat{q}_k(\alpha|X_{jk}).$$

To get the asymptotic behavior of the estimate $\hat{q}_k(\alpha|x_k)$, some more sophisticated calculations are necessary as in the case of a deterministic weighting function. In the following theorem, we state the asymptotic results for $\hat{q}_k(\alpha|x_k)$, where the Nadaraya-Watson estimator is used to estimate the conditional distribution function in (3.4) and (3.3), respectively.

Theorem 3.6 *If the assumptions (A1')-(A5') and (B1')-(B5') are satisfied, then we have for any* $\alpha \in (0, 1)$ *and for* $k = 1, \ldots, d$

$$\sqrt{nh_1}(\hat{q}_k(\alpha|x_k) - q_k(\alpha|x_k) + b_k(\alpha|x_k)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_k^2(\alpha|x_k)),$$

where

$$\begin{aligned} b_k(\alpha|x_k) &= \kappa_2(K)h_1^2 \int \left[\frac{\partial^2}{\partial x_k^2} F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}}) \right. \\ &\quad \left. + 2\frac{\frac{\partial}{\partial x_k} F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}}) \frac{\partial}{\partial x_k} p(x_k, x_{\underline{k}})}{p(x_k, x_{\underline{k}})} \right] \frac{1}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})} p_{\underline{k}}(x_{\underline{k}}) dx_{\underline{k}}, \\ s_k^2(\alpha|x_k) &= ||K||_2^2 \int \frac{\alpha(1-\alpha)p_{\underline{k}}^2(x_{\underline{k}})}{(F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}}))^2 p(x_k, x_{\underline{k}})} dx_{\underline{k}}, \end{aligned}$$

and $\kappa_2(K)$, $||K||_2^2$ *are defined as in Theorem 3.2.*

It is again worth to mention that the local linear estimate as the initial estimate for the distribution function yields an improved bias term, which is

$$b_k(\alpha|x_k) = \kappa_2(K)h_1^2 \int \frac{\frac{\partial^2}{\partial x_k^2} F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})} p_{\underline{k}}(x_{\underline{k}}) dx_{\underline{k}}.$$

Remark 3.7 The asymptotic properties of the additive quantile function

$$\hat{Q}_{\text{add}}(\alpha|\mathbf{x}) = \sum_{k=1}^{d} \hat{q}_k(\alpha|x_k) - \left(1 - \frac{1}{d}\right) \sum_{k=1}^{d} \frac{1}{n} \sum_{j=1}^{n} \hat{q}_k(\alpha|X_{jk})$$

can be derived as well. The asymptotic bias of $\hat{Q}_{\text{add}}(\alpha|\mathbf{x})$ is

$$\sum_{k=1}^{d} b_k(\alpha|x_k) - \left(1 - \frac{1}{d}\right) \int b_k(\alpha|x_k) p_k(x_k) dx_k,$$

where $b_k(\alpha|x_k)$ is the bias of $\hat{q}_k(\alpha|x_k)$. The asymptotic variance is the sum of the variances of $\hat{q}_k(\alpha|x_k)$, since the terms in $Q_{\text{add}}(\alpha|\mathbf{x})$ are asymptotically uncorrelated.

Now we state a similar result for the internal estimator. This estimator is particularly interesting in the context of marginal integration, since the usual marginal integration estimate is computationally not very appealing. The internalized Nadaraya-Watson estimate of the conditional distribution function is of the following form

$$\hat{F}_{INW}(y|x_k, x_{\underline{k}}) = \sum_{i=1}^{n} \frac{K_{h_1}(x_k - X_{ik})L_{H_2}(x_{\underline{k}} - X_{i\underline{k}})I\{Y_i \le y\}}{\sum_{j=1}^{n} K_{h_1}(X_{jk} - X_{ik})L_{H_2}(X_{j\underline{k}} - X_{i\underline{k}})},$$

which is as we have seen in Section 3.3 basically a weighted sum over the adjusted data

$$\tilde{Y}_i = \frac{I\{Y_i \le y\}}{\frac{1}{n}\sum_{j=1}^{n} K_{h_1}(X_{jk} - X_{ik})L_{H_2}(X_{j\underline{k}} - X_{i\underline{k}})} = \frac{I\{Y_i \le y\}}{\hat{p}(X_{ik}, X_{i\underline{k}})}.$$

The corresponding estimate for $q_k(\alpha|x_1)$ is defined by

$$\tilde{q}_k(\alpha|x_k) = \frac{1}{n}\sum_{j=1}^{n} \tilde{Q}_I(\alpha|x_k, X_{j\underline{k}}),$$

where $\hat{F}_{INW}(y|x_k, x_{\underline{k}})$ is used to calculate the monotonized inverse $\tilde{Q}_I(\alpha|x_k, X_{j\underline{k}})$.

Theorem 3.8 *If the assumptions of Theorem 3.6 are satisfied, then we have for $\alpha \in (0,1)$ and $k = 1, \ldots, d$*

$$\sqrt{nh_1}(\tilde{q}_k(\alpha|x_k) - q_k(\alpha|x_k) + \tilde{b}_{1k}(\alpha|x_k) - \tilde{b}_{2k}(\alpha|x_k)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{s}_k^2(\alpha|x_k)),$$

where

$$\begin{aligned}
\tilde{b}_{1k}(\alpha|x_k) &= \kappa_2(K)h_1^2 \int \frac{\frac{\partial^2}{\partial x_k^2}F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})} p_{\underline{k}}(x_{\underline{k}})dx_{\underline{k}} \\
\tilde{b}_{2k}(\alpha|x_k) &= \kappa_2(K)h_1^2\alpha \int \frac{\frac{\partial^2}{\partial x_k^2}p(x_k, x_{\underline{k}})}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})p(x_k, x_{\underline{k}})} p_{\underline{k}}(x_{\underline{k}})dx_{\underline{k}} \\
\tilde{s}_k^2(\alpha|x_k) &= ||K||_2^2 \int \frac{\alpha p_{\underline{k}}^2(x_{\underline{k}})}{p(x_k, x_{\underline{k}})(F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}}))^2} dx_{\underline{k}}
\end{aligned}$$

The internalized marginal integration estimate $\tilde{q}_k(\alpha|x_k)$ is less efficient than the estimate $\hat{q}_k(\alpha|x_k)$, since it has an additional term in its bias and variance. Compare with Theorem 1.4 and Theorem 3.6, we obtain

$$\sigma^2(Q(\alpha|\mathbf{x})|\mathbf{x}) + F^2(Q(\alpha|\mathbf{x})|\mathbf{x}) = \alpha(1-\alpha) + \alpha^2 = \alpha \ge \alpha(1-\alpha).$$

There is some theory about the difference between external and internal approaches [see Chapter 1 for a discussion]. When using externalized methods, one smoothes the data first at hand, and then modifies the result to obtain an estimator. In the other case, one first modifies the data, or at least the empirical distribution function to get something like unbiasedness, and then smoothes. So the difference of this two approaches is quite fundamental. Jones et al. (1994) give some more explanation why internal methods are preferable.

Remark 3.9 Note that similar remarks can be stated as in Section 3.5. This means, e.g., that the assumptions of independent data can be relaxed.
The comparison of the estimate $\hat{q}_k(\alpha|x_k)$ and the estimate suggested by De Gooijer and Zerom (2003) is interesting since the two estimate are for all intents and purposes similar in spirit. Recall Remark 3.3, which applies for $\hat{q}_k(\alpha|x_k)$ as well. From asymptotic theory, the estimate of De Gooijer and Zerom (2003) behaves better in low-dimensional problems ($d \leq 4$), since less assumptions on the smoothness of the functions p and $F(y|\mathbf{x})$ have to be made. On the other side, the estimate $\hat{q}_k(\alpha|x_k)$ works for high-dimensional settings in contrast to the estimate of De Gooijer and Zerom (2003), which requires positive kernel functions. In Section 3.6, we conduct a comparison of this two estimates in terms of finite sample properties.

3.5.3 Proof of Theorem 3.2

For the sake of simplicity, we assume $N = n$, and that the transformation function G corresponds to the uniform distribution. Recall the definitions of $\hat{Q}_I(\alpha|\mathbf{x})$ and $Q_N(\alpha|\mathbf{x})$. As in the proof of Theorem 2.12 in previous chapter, we analyze the function

$$g\left(\frac{\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}}) - \alpha}{h_m}\right) = \int_{\frac{\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}})-\alpha}{h_m}}^{\infty} K_m(v)dv = \frac{1}{h_m}\int_{-\infty}^{\alpha} K_m\left(\frac{\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}}) - u}{h_m}\right) du,$$

where this time we utilize the conditional distribution function $F(y|\mathbf{x})$ and its estimate, respectively. With a Taylor expansion of degree 1, we obtain for $\xi_i = \xi_i(\alpha, x_k, x_{\underline{k}})$ between $F(\frac{i}{n}|x_k, x_{\underline{k}})$ and $\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}})$:

$$\begin{aligned} g\left(\frac{\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}}) - \alpha}{h_m}\right) &= \int_{\frac{\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}})-\alpha}{h_m}}^{\infty} K_m(v)dv \\ &= \int_{\frac{F(\frac{i}{n}|x_k, x_{\underline{k}})-\alpha}{h_m}}^{\infty} K_m(v)dv \\ &\quad -\frac{1}{h_m}K_m\left(\frac{F(\frac{i}{n}|x_k, x_{\underline{k}}) - \alpha}{h_m}\right)\left(\hat{F}\left(\frac{i}{n}|x_k, x_{\underline{k}}\right) - F\left(\frac{i}{n}|x_k, x_{\underline{k}}\right)\right) \\ &\quad -\frac{1}{2}\frac{1}{h_m^2}K_m'\left(\frac{\xi_i - \alpha}{h_m}\right)\left(\hat{F}\left(\frac{i}{n}|x_k, x_{\underline{k}}\right) - F\left(\frac{i}{n}|x_k, x_{\underline{k}}\right)\right)^2, \end{aligned}$$

where we have the following inequality for all $\alpha \in (0,1)$ and $i = 1, \dots, n$

$$\left|\xi_i - F\left(\frac{i}{n}|x_k, x_{\underline{k}}\right)\right| \le \left|\hat{F}\left(\frac{i}{n}|x_k, x_{\underline{k}}\right) - F\left(\frac{i}{n}|x_k, x_{\underline{k}}\right)\right|. \tag{3.11}$$

$\gamma_W(\alpha|x_k)$ can be approximated by $\gamma_{W,n}(\alpha|x_k)$ since h_m fulfills (B3′):

$$\begin{aligned}
\gamma_W(\alpha|x_k) - \gamma_{W,n}(\alpha|x_k) &= \int \left(Q(\alpha|x_k, x_{\underline{k}}) - Q_n(\alpha|x_k, x_{\underline{k}})\right) W(dx_{\underline{k}}) \qquad (3.12)\\
&= \int \Bigg[\frac{1}{2}\left(\int v^2 K_m(v)dv\right) h_m^2 Q''(\alpha|x_k, x_{\underline{k}})\\
&\qquad + o(h_m^2) + O\left(\frac{1}{nh_m}\right)\Bigg] W(dx_{\underline{k}})\\
&= o\left(\frac{1}{\sqrt{nh_1}}\right).
\end{aligned}$$

This is the expansion derived in Lemma 1.8. Taking the above approximation into account, it is enough to consider $\hat{\gamma}_W(\alpha|x_k) - \gamma_{W,n}(\alpha|x_k)$:

$$\begin{aligned}
\hat{\gamma}_W(\alpha|x_k) - \gamma_W(\alpha|x_k) &= \hat{\gamma}_W(\alpha|x_k) - \gamma_{W,n}(\alpha|x_k) + o\left(\frac{1}{\sqrt{nh_1}}\right)\\
&= \int \left[\hat{Q}_I(\alpha|x_k, x_{\underline{k}}) - Q_n(\alpha|x_k, x_{\underline{k}})\right] W(dx_{\underline{k}}) + o\left(\frac{1}{\sqrt{nh_1}}\right)\\
&= \int \frac{1}{n}\sum_{i=1}^n \left[\int_{\frac{\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}}) - \alpha}{h_m}}^{\infty} K_m(v)dv - \int_{\frac{F(\frac{i}{n}|x_k, x_{\underline{k}}) - \alpha}{h_m}}^{\infty} K_m(v)dv\right] W(dx_{\underline{k}})\\
&\quad + o\left(\frac{1}{\sqrt{nh_1}}\right)\\
&= \Delta_n^{(1)}(\alpha|x_k) + \frac{1}{2}\Delta_n^{(2)}(\alpha|x_k) + o\left(\frac{1}{\sqrt{nh_1}}\right),
\end{aligned}$$

where the terms $\Delta_n^{(1)}(\alpha|x_k)$ and $\Delta_n^{(2)}(\alpha|x_k)$ are defined by a Taylor expansion of order 1, i.e.

$$\begin{aligned}
\Delta_n^{(1)}(\alpha|x_k) &= -\frac{1}{nh_m}\int \sum_{i=1}^n K_m\left(\frac{F(\frac{i}{n}|x_k, x_{\underline{k}}) - \alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}}) - F(\frac{i}{n}|x_k, x_{\underline{k}})\right) W(dx_{\underline{k}}),\\
\Delta_n^{(2)}(\alpha|x_k) &= -\frac{1}{nh_m^2}\int \sum_{i=1}^n K'_m\left(\frac{\xi_i - \alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k, x_{\underline{k}}) - F(\frac{i}{n}|x_k, x_{\underline{k}})\right)^2 W(dx_{\underline{k}}).
\end{aligned}$$

In the first step, we show that $\Delta_n^{(2)}(\alpha|x_k) = o_p\left(\frac{1}{\sqrt{nh_1}}\right)$ for a number ξ_i, which fulfills (3.11).

$$
\begin{aligned}
& |\Delta_n^{(2)}(\alpha|x_k)| \\
= \;& \frac{1}{nh_m^2}\left|\int \sum_{i=1}^n K_m'\left(\frac{\xi_i-\alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k,x_{\underline{k}}) - F(\frac{i}{n}|x_k,x_{\underline{k}})\right)^2 W(dx_{\underline{k}})\right| \\
= \;& \frac{1}{nh_m^2}\left|\int \sum_{i=1}^n K_m'\left(\frac{F(\frac{i}{n}|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\left[1+\left(K_m'\left(\frac{F(\frac{i}{n}|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\right)^{-1}\times\right.\right. \\
& \left.\left(K_m'\left(\frac{\xi_i-\alpha}{h_m}\right) - K_m'\left(\frac{F(\frac{i}{n}|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\right)\right]\left(\hat{F}(\frac{i}{n}|x_k,x_{\underline{k}}) - F(\frac{i}{n}|x_k,x_{\underline{k}})\right)^2 W(dx_{\underline{k}})\Bigg| \\
= \;& \frac{(1+o_p(1))}{nh_m^2}\left|\int \sum_{i=1}^n K_m'\left(\frac{F(\frac{i}{n}|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k,x_{\underline{k}}) - F(\frac{i}{n}|x_k,x_{\underline{k}})\right)^2 W(dx_{\underline{k}})\right| \qquad (3.13) \\
\le \;& \frac{(1+o_p(1))}{nh_m^2}\int \sum_{i=1}^n \left|K_m'\left(\frac{F(\frac{i}{n}|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\right|\left(\hat{F}(\frac{i}{n}|x_k,x_{\underline{k}}) - F(\frac{i}{n}|x_k,x_{\underline{k}})\right)^2 W(dx_{\underline{k}}),
\end{aligned}
$$

where line (3.13) follows from the Lipschitz continuity of K_m', the bandwidth condition (B5′), and (3.11), since

$$
\begin{aligned}
\left|K_m'\left(\frac{\xi_i-\alpha}{h_m}\right) - K_m'\left(\frac{F(\frac{i}{n}|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\right| &\le L\left|\frac{\xi_i - F(\frac{i}{n}|x_k,x_{\underline{k}})}{h_m}\right| \\
&\le L\left|\frac{\hat{F}(\frac{i}{n}|x_k,x_{\underline{k}}) - F(\frac{i}{n}|x_k,x_{\underline{k}})}{h_m}\right| \\
&= O_p\left(\frac{\log n}{nh_1h_2^{d-1}h_m^2}\right)^{1/2} = o_p(1).
\end{aligned}
$$

For the last identity, the uniform convergence rate of $\hat{F}(\frac{i}{n}|x_k,x_{\underline{k}})$ is used, which can be found in Collomb and Härdle (1986) for higher dimensions.
Using the bandwidth condition (B5′), we obtain for the term $\Delta_n^{(2)}$ by the theorem of Tonelli approximating the sum over i by the corresponding integral

$$
\begin{aligned}
E[|\Delta_n^{(2)}(\alpha|x_k)|] &\le \frac{(1+o(1))}{h_m^2}\int\int_0^1 \left|K_m'\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\right| \times \\
&\qquad E\left[\left(\hat{F}(t|x_k,x_{\underline{k}}) - F(t|x_k,x_{\underline{k}})\right)^2\right] dtW(dx_{\underline{k}}) \\
&= \frac{(1+o(1))}{h_m}\int\int_0^1 |K_m'(s)|\,ds\frac{1}{F'(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})} \times \qquad (3.14) \\
&\qquad E\left[\left(\hat{F}(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}}) - F(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})\right)^2\right] dtW(dx_{\underline{k}}) \\
&= O\left(\frac{1}{h_m}\left(\frac{1}{nh_1h_2^{d-1}}\right)\right) = o_p\left(\frac{1}{\sqrt{nh_1}}\right).
\end{aligned}
$$

In line (3.14), the substitution $s = \frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}$ is applied. The last identity follows from the condition (B5').
Now we can turn to the remaining term $\Delta_n^{(1)}(\alpha|x_k)$. This term can be separated into a bias and a variance term. Again, we interpret the sum over i as approximation of the corresponding integral. Furthermore, we decompose $(\hat{F}(t|x_k,x_{\underline{k}}) - F(t|x_k,x_{\underline{k}}))$ and exchange $\hat{p}$ by p since $|\hat{p}(x_k,x_{\underline{k}}) - p(x_k,x_{\underline{k}})| = o_p(1)$.

$$\begin{aligned}
\Delta_n^{(1)}(\alpha|x_k) &= -\frac{(1+o_p(1))}{nh_m}\sum_{j=1}^n \int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\times \\
&\qquad K_{h_1}(x_k - X_{jk})\, L_{H_2}\left(x_{\underline{k}} - X_{j\underline{k}}\right)\left(\frac{I\{Y_j\le t\} - F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right) dtW(dx_{\underline{k}}) \\
&= -\frac{(1+o_p(1))}{nh_m}\sum_{j=1}^n \int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right) K_{h_1}(x_k - X_{jk})\, L_{H_2}\left(x_{\underline{k}} - X_{j\underline{k}}\right)\times \\
&\qquad \left(\frac{F(t|X_{jk},X_{j\underline{k}}) - F(t|x_k,x_{\underline{k}}) + \sigma(t|X_{jk},X_{j\underline{k}})\varepsilon_j}{p(x_k,x_{\underline{k}})}\right) dtW(dx_{\underline{k}}) \\
&= (1+o_p(1))\left(\Delta_n^{(1.1)}(\alpha|x_k) + \Delta_n^{(1.2)}(\alpha|x_k)\right),
\end{aligned}$$

where

$$\begin{aligned}
\Delta_n^{(1.1)}(\alpha|x_k) &= -\frac{1}{nh_m}\int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\sum_{j=1}^n K_{h_1}(x_k - X_{jk})\, L_{H_2}\left(x_{\underline{k}} - X_{j\underline{k}}\right)\times \\
&\qquad \left(\frac{F(t|X_{jk},X_{j\underline{k}}) - F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right) W(dx_{\underline{k}})dt, \\
\Delta_n^{(1.2)}(\alpha|x_k) &= -\frac{1}{nh_m}\sum_{j=1}^n \int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right) K_{h_1}(x_k - X_{jk}) L_{H_2}(x_{\underline{k}} - X_{j\underline{k}})\times \\
&\qquad \frac{\sigma(t|X_{jk},X_{j\underline{k}})\varepsilon_j}{p(x_k,x_{\underline{k}})} dtW(dx_{\underline{k}}).
\end{aligned}$$

Recall the explicit definition of

$$\sigma^2(t|X_{jk},X_{j\underline{k}}) = F(t|X_{jk},X_{j\underline{k}})(1 - F(t|X_{jk},X_{j\underline{k}})), \tag{3.15}$$

which is motivated in Section 3.3. The first term represents basically the bias term. To calculate the expectation and the variance, we use the following multi-index notation for the differential operator D and the multi-index $\nu = (\nu_1,\ldots,\nu_d)^T$

$$D^{\nu_{\underline{k}}} := D_1^{\nu_1}\ldots D_{k-1}^{\nu_{k-1}} D_{k+1}^{\nu_{k+1}}\ldots D_d^{\nu_d} = \frac{\partial^{\nu_1}}{\partial x_1^{\nu_1}}\cdots\frac{\partial^{\nu_{k-1}}}{\partial x_{k-1}^{\nu_{k-1}}}\frac{\partial^{\nu_{k+1}}}{\partial x_{k+1}^{\nu_{k+1}}}\cdots\frac{\partial^{\nu_d}}{\partial x_d^{\nu_d}}.$$

First we calculate the expectation

$$
\begin{aligned}
& E[\Delta_n^{(1.1)}(\alpha|x_k)] \\
= \; & -\frac{(1+o(1))}{h_m}\int\int_0^1\int K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)K_{h_1}(x_k-u_k)\,L_{H_2}\left(x_{\underline{k}}-u_{\underline{k}}\right)\times \\
& \left(\frac{F(t|u_k,u_{\underline{k}})-F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right)p(u_k,u_{\underline{k}})W(dx_{\underline{k}})dtdu_kdu_{\underline{k}} \\
= \; & -\frac{(1+o(1))}{h_m}\int\int_0^1\int K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)K(v_k)L(v_{\underline{k}})p(x_k-h_1v_k,x_{\underline{k}}-h_2v_{\underline{k}})\times \\
& \left(\frac{F(t|x_k-h_1v_k,x_{\underline{k}}-h_2v_{\underline{k}})-F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right)W(dx_{\underline{k}})dtdv_kdv_{\underline{k}} \\
= \; & -\frac{(1+o(1))}{h_m}\int\int_0^1\int K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)K(v_k)L(v_{\underline{k}})\times \\
& \left[\frac{1}{2}h_1^2v_k^2\frac{\partial^2}{\partial x_k^2}F(t|x_k,x_{\underline{k}})+h_1^2v_k^2\frac{\frac{\partial}{\partial x_k}F(t|x_k,x_{\underline{k}})\frac{\partial}{\partial x_k}p(x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}+\frac{h_2^q}{q!}\sum_{|\nu_{\underline{k}}|=q}\frac{v_{\underline{k}}^{\nu_{\underline{k}}}}{p(x_k,x_{\underline{k}})}\times\right. \\
& \left.\left(D^{\nu_{\underline{k}}}\left(F(t|x_k,x_{\underline{k}})p(x_k,x_{\underline{k}})\right)-F(t|x_k,x_{\underline{k}})D^{\nu_{\underline{k}}}p(x_k,x_{\underline{k}})\right)\right]W(dx_{\underline{k}})dtdv_ldv_{\underline{l}} \\
= \; & -(1+o(1))\left(\int v^2K(v)dv\right)\int\int_{\frac{-\alpha}{h_m}}^{\frac{1-\alpha}{h_m}}K_m(t')dt'\frac{1}{F'(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})}\times \qquad (3.16) \\
& \left[\frac{1}{2}h_1^2\frac{\partial^2}{\partial x_k^2}F(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})+h_1^2\frac{\frac{\partial}{\partial x_k}F(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})\frac{\partial}{\partial x_k}p(x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right]W(dx_{\underline{k}}) \\
& +O_p(h_2^q) \\
= \; & -(1+o(1))h_1^2\mu_2(K)\int\frac{1}{F'(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})}\times \\
& \left[\frac{1}{2}\frac{\partial^2}{\partial x_k^2}F(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})+\frac{\frac{\partial}{\partial x_k}F(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})\frac{\partial}{\partial x_k}p(x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right]W(dx_{\underline{k}}) \\
& +O_p(h_2^q) \\
= \; & -(1+o(1))b_k(\alpha|x_k).
\end{aligned}
$$

In the second equation, we apply the substitution $v_k=\frac{x_k-u_k}{h_1}$ and $v_{\underline{k}}=\frac{x_{\underline{k}}-u_{\underline{k}}}{h_2}$ componentwise. In line (3.16), we use the substitution $t'=\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}$. The bounds of integration with respect to t' converge towards $\pm\infty$ that the integral is 1. We show in the following that the variance of this random variable is negligible.

$$\begin{aligned}
\mathrm{Var}(\Delta_n^{(1.1)}(\alpha|x_k)) &= \mathrm{Var}\Bigg(\frac{1}{nh_m} \int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right) \times \\
&\quad \sum_{j=1}^n K_{h_1}\left(x_k - X_{jk}\right) L_{H_2}\left(x_{\underline{k}} - X_{j\underline{k}}\right) \left(\frac{F(t|X_{jk},X_{j\underline{k}}) - F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right) W(dx_{\underline{k}})dt \Bigg) \\
&= \frac{1}{n^2h_m^2}\sum_{j=1}^n \mathrm{Var}\Bigg(\int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right) K_{h_1}\left(x_k - X_{jk}\right) L_{H_2}\left(x_{\underline{k}} - X_{j\underline{k}}\right) \times \\
&\quad \left(\frac{F(t|X_{jk},X_{j\underline{k}}) - F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right) W(dx_{\underline{k}})dt \Bigg) \\
&= \frac{1}{nh_m^2} Var\Bigg(\int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right) K_{h_1}\left(x_k - X_{1k}\right) K\left(x_{\underline{k}} - X_{1\underline{k}}\right) \times \\
&\quad \left(\frac{F(t|X_{1k},X_{1k}) - F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right) W(dx_{\underline{k}})dt \Bigg) \\
&\leq \frac{1}{nh_m^2}\int\Bigg(\int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right) K_{h_1}\left(x_k - u_k\right) L_{H_2}\left(x_{\underline{k}} - u_{\underline{k}}\right) \times \\
&\quad \left(\frac{F(t|u_k,u_{\underline{k}}) - F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right) W(dx_{\underline{k}})dt \Bigg)^2 p(u_k,u_{\underline{k}})du_k du_{\underline{k}} \\
&= \frac{1}{nh_m^2h_1}\int K^2(v_k)p(x_k - h_1v_k, u_{\underline{k}})\Bigg(\int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right) L_{H_2}\left(x_{\underline{k}} - u_{\underline{k}}\right) \times \\
&\quad \left(\frac{F(t|x_k - h_1v_k,u_{\underline{k}}) - F(t|x_k,x_{\underline{k}})}{p(x_k,x_{\underline{k}})}\right) W(dx_{\underline{k}})dt \Bigg)^2 dv_k du_{\underline{k}} \\
&= \frac{1}{nh_1}\int K^2(v_k)p(x_k - h_1v_k, u_{\underline{k}})\Bigg(\int\int_{\frac{-\alpha}{h_m}}^{\frac{1-\alpha}{h_m}} K_m(t')dt' L_{H_2}(x_{\underline{k}} - u_{\underline{k}}) \times \\
&\quad \left(\frac{F(Q(\alpha|x_k,x_{\underline{k}})|x_k - h_1v_k,u_{\underline{k}}) - F(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})}{F'(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})p(x_k,x_{\underline{k}})}\right) W(dx_{\underline{k}}) \Bigg)^2 dv_k du_{\underline{k}} \\
&= \frac{1}{nh_1}\int K^2(v_k)p(x_k,u_{\underline{k}})\frac{w^2(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2p^2(x_k,u_{\underline{k}})}\Bigg(\int L(v_{\underline{k}}) \times \\
&\quad \left(F(Q(\alpha|x_k,u_{\underline{k}})|x_k - h_1v_k,u_{\underline{k}}) - F(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}} + h_2v_{\underline{k}})\right) dv_{\underline{k}} \Bigg)^2 dv_k du_{\underline{k}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(1+o(1))}{nh_1} \int \frac{K^2(v_k)p(x_k,u_{\underline{k}})w^2(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2 p^2(x_k,u_{\underline{k}})} \Bigg(-h_1 v_k \frac{\partial}{\partial x_k} F(Q(\alpha|x_k,u_{\underline{k}})|\xi_k,u_{\underline{k}}) \\
&\quad - \frac{h_2^q}{q!} \sum_{|\nu_{\underline{k}}|=q} \left(\int v_{\underline{k}}^{\nu_{\underline{k}}} L(v_{\underline{k}}) dv_{\underline{k}} \right) D^{\nu_{\underline{k}}} F(Q(\alpha|x_k,u_{\underline{k}})|x_k,\xi_{\underline{k}}) \Bigg)^2 dv_k du_{\underline{k}} \\
&= O\left(\frac{1}{nh_1}(h_1+h_2^q)^2 \right) = o\left(\frac{1}{nh_1} \right).
\end{aligned}
$$

For this derivation similar substitutions are applied as for calculating the expectation. Recall that w is the density of the weighting function W.
Since ε_i and X_i are independent and $E[\varepsilon_i] = 0$, we obtain $E[\Delta_n^{(1.2)}(\alpha|x_k)] = 0$ for the second term. We show that the second term represents the variance term.

$$
\begin{aligned}
\operatorname{Var}(\sqrt{nh_1}\Delta_n^{(1.2)}(\alpha|x_k)) &= \frac{h_1}{nh_m^2} \sum_{j=1}^n \operatorname{Var}\Bigg(\int \int_0^1 K_m \left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m} \right) \times \\
&\quad K_{h_1}(x_k - X_{jk}) L_{H_2}(x_{\underline{k}} - X_{j\underline{k}}) \frac{\sigma(t|X_{jk},X_{j\underline{k}})\varepsilon_j}{p(x_k,x_{\underline{k}})} dt W(dx_{\underline{k}}) \Bigg) \\
&= \frac{h_1}{h_m^2} \int \Bigg(\int \int_0^1 K_m \left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m} \right) K_{h_1}(x_k - u_k) L_{H_2}(x_{\underline{k}} - u_{\underline{k}}) \times \\
&\quad \frac{\sigma(t|u_k,u_{\underline{k}})}{p(x_k,x_{\underline{k}})} dt W(dx_{\underline{k}}) \Bigg)^2 p(u_k,u_{\underline{k}}) du_k du_{\underline{k}} \\
&= h_1 \int p(u_k,u_{\underline{k}}) \Bigg(\int \int_{\frac{-\alpha}{h_m}}^{\frac{1-\alpha}{h_m}} K_m(t') \frac{\sigma(Q(\alpha+h_m t'|x_k,x_{\underline{k}})|u_k,u_{\underline{k}})}{F'(Q(\alpha+h_m t'|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})} dt' \times \\
&\quad K_{h_1}(x_k - u_k) L_{H_2}(x_{\underline{k}} - u_{\underline{k}}) \frac{1}{p(x_k,x_{\underline{k}})} W(dx_{\underline{k}}) \Bigg)^2 du_k du_k \\
&= h_1 \int p(u_k,u_{\underline{k}}) K_{h_1}^2(x_k - u_k) \Bigg(\int L_{H_2}(x_{\underline{k}} - u_{\underline{k}}) \times \\
&\quad \frac{\sigma(Q(\alpha|x_k,x_{\underline{k}})|u_k,u_{\underline{k}})}{F'(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})p(x_k,x_{\underline{k}})} W(dx_{\underline{k}}) \Bigg)^2 du_k u_{\underline{k}} \\
&= h_1 \int p(u_k,u_{\underline{k}}) K_{h_1}^2(x_k - u_k) \times \\
&\quad \left(\int L(v_{\underline{k}}) \frac{\sigma(Q(\alpha|x_k,u_{\underline{k}}+h_2 v_{\underline{k}})|u_k,u_{\underline{k}}) w(u_{\underline{k}}+h_2 v_{\underline{k}}) p(x_k,u_{\underline{k}}+h_2 v_{\underline{k}}) dv_{\underline{k}}}{F'(Q(\alpha|x_k,u_{\underline{k}}+h_2 v_{\underline{k}})|x_k,u_{\underline{k}}+h_2 v_{\underline{k}}) p^2(x_k,u_{\underline{k}})} \right)^2 du_k du_{\underline{k}}
\end{aligned}
\tag{3.17}
$$

$$
\begin{aligned}
&= (1+o(1)) \int \frac{\sigma^2(Q(\alpha|x_k,u_{\underline{k}})|x_k-h_1v_k,u_{\underline{k}})K^2(v_k)w^2(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2p(x_k-h_1v_k,u_{\underline{k}})}dv_kdu_{\underline{k}} \\
&= (1+o(1)) \int K^2(v)dv \int \frac{\sigma^2(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})w^2(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2p(x_k,u_{\underline{k}})}du_{\underline{k}} \\
&= (1+o(1)) \int K^2(v)dv \int \frac{\alpha(1-\alpha)w^2(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2p(x_k,u_{\underline{k}})}du_{\underline{k}},
\end{aligned}
$$

where w is the density of the probability measure W, and the definition of the variance function σ in the nonparametric model of the conditional distribution function is used [see (3.15)]. Line (3.17) shows why the condition $\lim_{n\to\infty}\frac{h_1}{h_m}=\infty$ does not affect the convergence rate, as the argument in the kernel functions is not changed by the substitution $t'=\frac{F(t|x_k,u_{\underline{k}})-\alpha}{h_m}$. In other words, the arguments of the kernel functions are independent of t.

Finally, the weak convergence of

$$\sqrt{nh_1}\Delta_n^{(1.2)}(\alpha|x_k) \xrightarrow{\mathcal{D}} \mathcal{N}(0,s_k^2(\alpha|x_k))$$

follows as a consequence of the Lyapunov's Theorem for $\delta=2$ since

$$
\begin{aligned}
&\sum_{j=1}^n E\Bigg[\frac{\sqrt{nh_1}}{nh_m}\int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)K_{h_1}(x_k-X_{jk})\,L_{H_2}\left(x_{\underline{k}}-X_{j\underline{k}}\right)\times \\
&\qquad \frac{\sigma(t|X_{jk},X_{j\underline{k}})\varepsilon_j}{p(x_k,x_{\underline{k}})}dtW(dx_{\underline{k}})\Bigg]^4 \\
&= \frac{h_1^2}{n^2h_m^4}\sum_{j=1}^n E\Bigg[\int\int_0^1 K_m\left(\frac{F(t|x_k,x_{\underline{k}})-\alpha}{h_m}\right)K_{h_1}(x_k-X_{jk})\,L_{H_2}\left(x_{\underline{k}}-X_{j\underline{k}}\right)\times \\
&\qquad \frac{\sigma(t|X_{jk},X_{j\underline{k}})\varepsilon_j}{p(x_k,x_{\underline{k}})}dtW(dx_{\underline{k}})\Bigg]^4 \\
&= \frac{h_1^2E[\varepsilon_1^4]}{n}\int K_{h_1}^4(x_k-u_k)\left(\int\frac{\sigma(Q(\alpha|x_k,x_{\underline{k}})|u_k,u_{\underline{k}})L_{H_2}\left(x_{\underline{k}}-u_{\underline{k}}\right)}{F'(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})p(x_k,x_{\underline{k}})}W(dx_{\underline{k}})\right)^4\times \\
&\qquad p(u_k,u_{\underline{k}})du_kdu_{\underline{k}} \\
&= \frac{E[\varepsilon_1^4]}{nh_1}\int K^4(v_k)p(x_k-h_1v_k,u_{\underline{k}})\times \\
&\qquad \left(\int\frac{\sigma(Q(\alpha|x_k,x_{\underline{k}})|x_k-h_1v_k,u_{\underline{k}})L_{H_2}(x_{\underline{k}}-u_{\underline{k}})}{F'(Q(\alpha|x_k,x_{\underline{k}})|x_k,x_{\underline{k}})p(x_k,x_{\underline{k}})}W(dx_{\underline{k}})\right)^4dv_kdu_{\underline{k}} \\
&= (1+o(1))\frac{E[\varepsilon_1^4]}{nh_1}\left(\int K^4(v)dv\right)\int p(x_k,u_{\underline{k}})\times
\end{aligned}
$$

$$\begin{aligned}
&\left(\int \frac{L(v_{\underline{k}})\sigma(Q(\alpha|x_k, u_{\underline{k}} + h_2 v_{\underline{k}})|x_k, u_{\underline{k}})w(u_{\underline{k}} + h_2 x_{\underline{k}})}{F'(Q(\alpha|x_k, u_{\underline{k}} + h_2 v_{\underline{k}})|x_k, u_{\underline{k}} + h_2 v_{\underline{k}})p(x_k, u_{\underline{k}} + h_2 v_{\underline{k}})} dv_{\underline{k}} \right)^4 du_{\underline{k}} \\
= \;& (1 + o(1)) \frac{E[\varepsilon_1^4]}{nh_1} \left(\int K^4(v)dv \right) \int \frac{\sigma^4(Q(\alpha|x_k, u_{\underline{k}})|x_k, u_{\underline{k}})w^4(u_{\underline{k}})}{(F'(Q(\alpha|x_k, u_{\underline{k}})|x_k, u_{\underline{k}}))^4 p^3(x_k, u_{\underline{k}})} du_{\underline{k}} \\
= \;& (1 + o(1)) \frac{E[\varepsilon_1^4]}{nh_1} \left(\int K^4(v)dv \right) \int \frac{\alpha^2(1-\alpha)^2 w^4(u_{\underline{k}})}{(F'(Q(\alpha|x_k, u_{\underline{k}})|x_k, u_{\underline{k}}))^4 p^3(x_k, u_{\underline{k}})} du_{\underline{k}} = O\left(\frac{1}{nh_1} \right).
\end{aligned}$$

This completes the proof of the theorem. □

Remark 3.10 In equation (3.12), the bandwidth condition (B3') comes into play. This quite strong condition can be weakened by assuming $\frac{h_m^3}{h_1^2} = o(1)$ and $\frac{h_m}{h_1} = O(1)$. This means that an additional bias term appears in the expansion of $\hat{\gamma}_W(\alpha|x_k) - \gamma_W(\alpha|x_k)$, which is

$$c(\alpha|x_k) = \kappa_2(K_m)h_m^2 \int Q''(\alpha|x_k, x_{\underline{k}})dW(x_{\underline{k}}).$$

The first term of the second line of (3.12) adds to the bias term, and the remainder is $o\left(\frac{1}{\sqrt{nh_1}}\right)$.

3.5.4 Proof of Theorem 3.6

Again we assume $N = n$ and G be a uniform distribution to simplify the proof. By the law of the iterated logarithm, we observe

$$q_k(\alpha|x_k) = \frac{1}{n}\sum_{j=1}^n Q(\alpha|x_k, X_{j\underline{k}}) + O\left(\sqrt{\frac{\log\log n}{n}}\right) \;\; P-\text{a.s.}$$

Furthermore, $Q(\alpha|x_k, X_{j\underline{k}})$ can be approximated by $Q_n(\alpha|x_k, X_{j\underline{k}})$, which is defined in (3.5). Applying Lemma 1.8 and using the bandwidth condition (B3'), we obtain

$$\begin{aligned}
q_k(\alpha|x_k) &= \frac{1}{n}\sum_{j=1}^n Q_n(\alpha|x_k, X_{j\underline{k}}) + \frac{1}{n}\sum_{j=1}^n \left(Q(\alpha|x_k, X_{j\underline{k}}) - Q_n(\alpha|x_k, X_{j\underline{k}})\right) + o_p\left(\frac{1}{\sqrt{nh_1}}\right) \\
&= \frac{1}{n}\sum_{j=1}^n Q_n(\alpha|x_k, X_{j\underline{k}}) + o_p\left(\frac{1}{\sqrt{nh_1}}\right).
\end{aligned}$$

Hence to derive the asymptotic behavior of $\hat{q}_k(\alpha|x_k) - q_k(\alpha|x_k)$, we apply a Taylor expansion to the simplified difference

$$\begin{aligned}
\hat{q}_k(\alpha|x_k) - q_k(\alpha|x_k) &= \frac{1}{n}\sum_{j=1}^n \left[\hat{Q}_I(\alpha|x_k, X_{j\underline{k}}) - Q_n(\alpha|x_k, X_{j\underline{k}})\right] + o_p\left(\frac{1}{\sqrt{nh_1}}\right) \\
&= \Delta_n^{(1)}(\alpha|x_k) + \frac{1}{2}\Delta_n^{(2)}(\alpha|x_k) + o_p\left(\frac{1}{\sqrt{nh_1}}\right),
\end{aligned}$$

where

$$\Delta_n^{(1)}(\alpha|x_k) = -\frac{1}{n^2 h_m}\sum_{j=1}^{n}\sum_{i=1}^{n} K_m\left(\frac{F(\frac{i}{n}|x_k,X_{j\underline{k}})-\alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k,X_{j\underline{k}})-F(\frac{i}{n}|x_k,X_{j\underline{k}})\right),$$

$$\Delta_n^{(2)}(\alpha|x_k) = -\frac{1}{n^2 h_m^2}\sum_{j=1}^{n}\sum_{i=1}^{n} K_m'\left(\frac{\xi_i-\alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k,X_{j\underline{k}})-F(\frac{i}{n}|x_k,X_{j\underline{k}})\right)^2,$$

and $\xi_i = \xi_i(\alpha,x_k,X_{j\underline{k}})$ satisfies $|\xi_i - F(\frac{i}{n}|x_k,X_{j\underline{k}})| \le |\hat{F}(\frac{i}{n}|x_k,X_{j\underline{k}}) - F(\frac{i}{n}|x_k,X_{j\underline{k}})|$ for $i=1,\dots,n$. In the first step, we show that $\Delta_n^{(2)}(\alpha|x_k) = o_p\left(\frac{1}{\sqrt{nh_1}}\right)$. We observe as in the proof of Theorem 3.2 using the Lipschitz continuity of $K_m'(\cdot)$

$$\begin{aligned}
|\Delta_n^{(2)}(\alpha|x_k)| &= \frac{1}{n^2h_m^2}\left|\sum_{j=1}^{n}\sum_{i=1}^{n} K_m'\left(\frac{\xi_i-\alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k,X_{j\underline{k}})-F(\frac{i}{n}|x_k,X_{j\underline{k}})\right)^2\right| \\
&= \frac{1}{n^2h_m^2}\left|\sum_{j=1}^{n}\sum_{i=1}^{n} K_m'\left(\frac{F(\frac{i}{n}|x_k,X_{j\underline{k}})-\alpha}{h_m}\right)\left[1+\left(K_m'\left(\frac{F(\frac{i}{n}|x_k,X_{j\underline{k}})-\alpha}{h_m}\right)\right)^{-1}\times\right.\right. \\
&\left.\left.\left(K_m'\left(\frac{\xi_i-\alpha}{h_m}\right)-K_m'\left(\frac{F(\frac{i}{n}|x_k,X_{j\underline{k}})-\alpha}{h_m}\right)\right)\right]\left(\hat{F}(\frac{i}{n}|x_k,X_{j\underline{k}})-F(\frac{i}{n}|x_k,X_{j\underline{k}})\right)^2\right| \\
&= \frac{(1+o_p(1))}{n^2h_m^2}\left|\sum_{j=1}^{n}\sum_{i=1}^{n} K_m'\left(\frac{F(\frac{i}{n}|x_k,X_{j\underline{k}})-\alpha}{h_m}\right)\left(\hat{F}(\frac{i}{n}|x_k,X_{j\underline{k}})-F(\frac{i}{n}|x_k,X_{j\underline{k}})\right)^2\right|.
\end{aligned}$$

Finally, we get from the bandwidth condition (B5′)

$$\begin{aligned}
E\left[|\Delta_n^{(2)}(\alpha|x_k)||X_{j\underline{k}}=x_{\underline{k}}\right] &\le \frac{(1+o_p(1))}{n^2h_m^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\left|K_m'\left(\frac{F(\frac{i}{n}|x_k,x_{\underline{k}})-\alpha}{h_m}\right)\right|\times \\
&\quad E\left[\left(\hat{F}(\frac{i}{n}|x_k,x_{\underline{k}})-F(\frac{i}{n}|x_k,x_{\underline{k}})\right)^2\middle|X_{j\underline{k}}=x_{\underline{k}}\right] \\
&= O_p\left(\frac{1}{h_m}\left(\frac{1}{nh_1h_2^{d-1}}\right)\right) = o_p\left(\frac{1}{\sqrt{nh_1}}\right).
\end{aligned}$$

Thus, we can concentrate on the remaining term $\Delta_n^{(1)}(\alpha|x_k)$ which can be decomposed into bias and variance part. The sum over i is interpreted as approximation of the corresponding integral. We obtain observing the representation (3.1) and exchanging $\hat{p}$ by p since $|\hat{p}(x_k,X_{j\underline{k}})-p(x_k,X_{j\underline{k}})| = o_p(1)$

$$\begin{aligned}
\Delta_n^{(1)}(\alpha|x_k) &= -\frac{(1+o(1))}{nh_m}\sum_{j=1}^{n}\int_0^1 K_m\left(\frac{F(t|x_k,X_{j\underline{k}})-\alpha}{h_m}\right)\left(\hat{F}(t|x_k,X_{j\underline{k}})-F(t|x_k,X_{j\underline{k}})\right)dt \\
&= -\frac{(1+o_p(1))}{n^2h_m}\sum_{j=1}^{n}\sum_{m=1}^{n}\int_0^1 K_m\left(\frac{F(t|x_k,X_{j\underline{k}})-\alpha}{h_m}\right)K_{h_1}(x_k-X_{mk})\times
\end{aligned}$$

$$
\begin{aligned}
& L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \frac{I\{Y_m \le t\} - F(t|x_k, X_{j\underline{k}})}{p(x_k, X_{j\underline{k}})} dt \\
= \; & -\frac{(1+o_p(1))}{n^2} \sum_{j=1}^n \sum_{m=1}^n \int_{-\frac{\alpha}{h_m}}^{\frac{1-\alpha}{h_m}} K_m(t') K_{h_1}(x_k - X_{mk}) L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \times \\
& \left(\frac{F(Q(\alpha + h_m t'|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}}) - F(Q(\alpha + h_m t'|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}{F'(Q(\alpha + h_m t'|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) p(x_k, X_{j\underline{k}})} \right. \\
& \left. + \frac{\sigma(Q(\alpha + h_m t'|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}})\varepsilon_m}{F'(Q(\alpha + h_m t'|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) p(x_k, X_{j\underline{k}})} \right) dt' \\
= \; & -\frac{(1+o_p(1))}{n^2} \sum_{j=1}^n \sum_{m=1}^n K_{h_1}(x_k - X_{mk}) L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \times \\
& \left(\frac{F(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) p(x_k, X_{j\underline{k}})} \right. \\
& \left. + \frac{\sigma(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}})\varepsilon_m}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) p(x_k, X_{j\underline{k}})} \right) \\
= \; & (1+o_p(1)) \left(\Delta_n^{(1.1)}(\alpha|x_k) + \Delta_n^{(1.2)}(\alpha|x_k) \right), \qquad (3.18)
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_n^{(1.1)}(\alpha|x_k) &= -\frac{1}{n^2 h_m} \sum_{j=1}^n \sum_{m=1}^n K_{h_1}(x_k - X_{mk}) L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \times \\
& \left(\frac{F(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) p(x_k, X_{j\underline{k}})} \right), \\
\Delta_n^{(1.2)}(\alpha|x_k) &= -\frac{1}{n^2 h_m} \sum_{j=1}^n \sum_{m=1}^n K_{h_1}(x_k - X_{mk}) L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \times \qquad (3.19) \\
& \left(\frac{\sigma(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}})\varepsilon_m}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) p(x_k, X_{j\underline{k}})} \right).
\end{aligned}
$$

The terms $\Delta_n^{(1.1)}$ and $\Delta_n^{(1.2)}$ are now investigated separately. First of all, $\Delta_n^{(1.1)}(\alpha|x_k)$ can be written as

$$
\Delta_n^{(1.1)}(\alpha|x_k) = -\frac{1}{n} \sum_{j=1}^n \eta_j(\alpha|x_k),
$$

where

$$
\begin{aligned}
\eta_j(\alpha|x_k) &= \frac{1}{n} \sum_{m=1}^n K_{h_1}(x_k - X_{mk}) L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \\
& \frac{(F(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}))}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) p(x_k, X_{j\underline{k}})}.
\end{aligned}
$$

Now we break η_j into two uncorrelated terms: $E[\eta_j(\alpha|x_k)|\mathbf{X}_j]$ and $(\eta_j(\alpha|x_k) - E[\eta_j(\alpha|x_k)|\mathbf{X}_j])$. For the conditional expectation of $\eta_j(\alpha|x_k)$, we have

$$\begin{aligned}
E[\eta_j(\alpha|x_k)|\mathbf{X}_j] &= \int K_{h_1}(x_k - u_k)\, L_{H_2}\left(X_{j\underline{k}} - u_{\underline{k}}\right) \times \\
&\quad \frac{(F(Q(\alpha|x_k, X_{j\underline{k}})|u_k, u_{\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}))}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})p(x_k, X_{j\underline{k}})} p(u_k, u_{\underline{k}}) du_k du_{\underline{k}} \\
&= \int K(v_k) L(v_{\underline{k}}) p(x_k - h_1 v_k, X_{j\underline{k}} - h_2 v_{\underline{k}}) \\
&\quad \frac{(F(Q(\alpha|x_k, X_{j\underline{k}})|x_k - h_1 v_k, X_{j\underline{k}} - h_2 v_{\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}))}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})p(x_k, X_{j\underline{k}})} dv_k dv_{\underline{k}} \\
&= (1 + o_p(1)) h_1^2 \kappa_2(K) \Bigg(\frac{\frac{\partial^2}{\partial x_k^2} F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})} \\
&\quad + 2 \frac{\frac{\partial}{\partial x_k} F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})p(x_k, X_{j\underline{k}})} \Bigg) + O_p(h_2^q).
\end{aligned}$$

It is easy to see that

$$\begin{aligned}
E[\Delta_n^{(1.1)}(\alpha|x_k)] &= \int\int (F(Q(\alpha|x_k, w_{\underline{k}})|x_k - h_1 v_k, w_{\underline{k}} - h_2 v_{\underline{k}}) - F(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})) \\
&\quad \frac{K(v_k) L(v_{\underline{k}}) p(x_k - h_1 v_k, w_{\underline{k}} - h_2 v_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})p(x_k, w_{\underline{k}})} p_{\underline{k}}(w_{\underline{k}}) dv_k dv_{\underline{k}} dw_{\underline{k}} \\
&= (1 + o(1)) h_1^2 \kappa_2(K) \int \Bigg(\frac{\frac{\partial^2}{\partial x_k^2} F(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})} \\
&\quad + 2 \frac{\frac{\partial}{\partial x_k} F(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})p(x_k, w_{\underline{k}})} \Bigg) p_{\underline{k}}(w_{\underline{k}}) dw_{\underline{k}} + O(h_2^q) \\
&= -(1 + o(1)) b_k(\alpha|x_k) + o\left(\frac{1}{\sqrt{nh_1}}\right).
\end{aligned}$$

For the variance of $\Delta_n^{(1.1)}(\alpha|x_k)$, we observe

$$\begin{aligned}
& E[(\eta_j(\alpha|x_k) - E[\eta_j(\alpha|x_k)|X_j])^2|\mathbf{X}_j] \\
\leq\ & \frac{1}{n} \int \left(\frac{K_{h_1}(x_k - u_k)\, L_{H_2}\left(X_{j\underline{k}} - u_{\underline{k}}\right)}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})p(x_1, x_k, X_{j\underline{k}})} \right)^2 \\
& (F(Q(\alpha|x_k, X_{j\underline{k}})|u_k, u_{\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}))^2 du_k du_{\underline{k}} \\
=\ & \frac{1}{nh_1 h_2^{d-1}} \int \frac{K^2(v_k) L^2(v_{\underline{k}})}{(F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}))^2 p^2(x_k, X_{j\underline{k}})} \\
& (F(Q(\alpha|x_k, X_{j\underline{k}})|x_k - h_1 v_k, X_{j\underline{k}} - h_2 v_{\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}))^2 dv_k dv_{\underline{k}}
\end{aligned}$$

$$= \quad O_p\left(\frac{1}{nh_1h_2^{d-1}}(h_1^2+h_2^2)\right).$$

Since $E[E[\eta_j(\alpha|x_k)|\mathbf{X}_j](\eta_j(\alpha|x_k)-E[\eta_j(\alpha|\mathbf{X}_j])]=0$, we can estimate the variance of $\Delta_n^{(1.1)}(\alpha|x_k)$ by

$$\begin{aligned}\mathrm{Var}(\Delta_n^{(1.1)}(\alpha|x_k)) &= \mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^n E[\eta_j(\alpha|x_k)|\mathbf{X}_j]+\frac{1}{n}\sum_{j=1}^n(\eta_j(\alpha|x_k)-E[\eta_j(\alpha|x_k)|\mathbf{X}_j])\right)\\ &= \mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^n E[\eta_j(\alpha|x_k)|\mathbf{X}_j]\right)+\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^n(\eta_j(\alpha|x_k)-E[\eta_j(\alpha|x_k)|\mathbf{X}_j])\right).\end{aligned}$$

For the first term, we have

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^n E[\eta_j(\alpha|x_k)|\mathbf{X}_j]\right)\\ =\;& \frac{1}{n}\mathrm{Var}\left(E[\eta_j(\alpha|x_k)|\mathbf{X}_j]\right)\le\frac{1}{n}E[(E[\eta_j(\alpha|x_k)|\mathbf{X}_j])^2]\\ =\;& \frac{1}{n}\int\Bigg(\int K(v_k)L(v_{\underline{k}})p(x_k-h_1v_k,u_{\underline{k}}-h_2v_{\underline{k}})\\ &\qquad\frac{(F(Q(\alpha|x_k,u_{\underline{k}}|x_k-h_1v_k,u_{\underline{k}}-h_2v_{\underline{k}})-F(Q(\alpha|x_k,u_{\underline{k}}),x_k,u_{\underline{k}}))}{F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})p(x_k,u_{\underline{k}})}dv_kdv_{\underline{k}}\Bigg)^2p_{\underline{k}}(u_{\underline{k}})du_{\underline{k}}\\ =\;& \frac{(1+o(1))}{n}\int\Bigg(h_1^2\kappa_2(K)\Bigg(\frac{\frac{1}{2}\frac{\partial^2}{\partial x_k^2}F(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}{F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}+\frac{\frac{\partial}{\partial x_k}F(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}{F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})p(x_k,u_{\underline{k}})}\Bigg)\\ &+\frac{h_2^q}{q!}\sum_{|\nu_{\underline{k}}|=q}\int v_{\underline{k}}^{\nu_{\underline{k}}}L(v_{\underline{k}})dv_{\underline{k}}\Bigg(\frac{D^{\nu_{\underline{k}}}(F(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})p(x_k,u_{\underline{k}}))}{F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})p(x_k,u_{\underline{k}})}\\ &\qquad+\frac{F(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})D^{\nu_{\underline{k}}}p(x_k,u_{\underline{k}})}{F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})p(x_k,u_{\underline{k}})}\Bigg)\Bigg)^2p_{\underline{k}}(u_{\underline{k}})du_{\underline{k}}\\ =\;& O\left(\frac{(h_1^2+h_2^q)^2}{n}\right)=o\left(\frac{1}{nh_1}\right).\end{aligned}$$

On the other hand, we obtain for the second term

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^n(\eta_j(\alpha|x_k)-E[\eta_j(\alpha|x_k)|\mathbf{X}_j])\right)\\ =\;& E\left[\frac{1}{n}E\left[(\eta_j(\alpha|x_k)-E[\eta_j(\alpha|x_k)|\mathbf{X}_j])^2|\mathbf{X}_j\right]\right]\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{n^2h_1h_2^{d-1}}\int \frac{K^2(v_k)L^2(v_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}}))^2p^2(x_k,u_{\underline{k}})} \times \\
&\quad \left(F(Q(\alpha|x_k,u_{\underline{k}})|x_k-h_1v_k,u_{\underline{k}}-h_2v_{\underline{k}})-F(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})\right)^2 dv_k dv_{\underline{k}} du_{\underline{k}} \\
&= O\left(\frac{(h_1+h_2)^2}{n^2h_1h_2^{d-1}}\right) = o\left(\frac{1}{nh_1}\right).
\end{aligned}$$

This shows

$$\Delta_n^{(1.1)}(\alpha|x_k) + b_k(\alpha|x_k) = o_p\Big(\frac{1}{\sqrt{nh_1}}\Big). \tag{3.20}$$

We consider the term $\Delta_n^{(1.2)}$ in (3.19), which has expectation $E[\Delta^{(1.2)}(\alpha|x_k)] = 0$. To calculate the variance, we decompose $\Delta_n^{(1.2)}$ into

$$\Delta_n^{(1.2)}(\alpha|x_k) = -\frac{1}{n}\sum_{m=1}^{n} K_{h_1}(x_k - X_{mk})\,\varepsilon_m\beta_m(\alpha|x_k),$$

where

$$\beta_m(\alpha|x_k) = \frac{1}{n}\sum_{j=1}^{n}\frac{L_{H_2}\left(X_{j\underline{k}}-X_{m\underline{k}}\right)\sigma(Q(\alpha|x_k,X_{j\underline{k}})|X_{mk},X_{m\underline{k}})}{p(x_k,X_{j\underline{k}})F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})}.$$

Now we treat $\beta_m(\alpha|x_k)$ similar as $\eta_j(\alpha|x_k)$ and break it into $E[\beta_m(\alpha|x_k)|\mathbf{X}_m]$ and $\beta_m(\alpha|x_k) - E[\beta_m(\alpha|x_k)|\mathbf{X}_m]$. The conditional expectation yields

$$\begin{aligned}
E[\beta_m(\alpha|x_k)|\mathbf{X}_m] &= \int \frac{L_{H_2}\left(u_{\underline{k}}-X_{m\underline{k}}\right)\sigma(Q(\alpha|x_k,u_{\underline{k}})|X_{mk},X_{m\underline{k}})p_{\underline{k}}(u_{\underline{k}})}{p(x_k,u_{\underline{k}})F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}du_{\underline{k}} \\
&= \frac{\sigma(Q(\alpha|x_k,X_{m\underline{k}})|X_{mk},X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}})}{p(x_k,X_{m\underline{k}})F'(Q(\alpha|x_k,X_{m\underline{k}})|x_k,X_{m\underline{k}})} + O_p(h_2^q).
\end{aligned}$$

Furthermore we have

$$\begin{aligned}
&E\left[(\beta_m(\alpha|x_k)-E[\beta_m(\alpha|x_k)|\mathbf{X}_m])^2|\mathbf{X}_m\right] \\
&\leq \frac{1}{n}\int\left(\frac{L_{H_2}\left(u_{\underline{k}}-X_{m\underline{k}}\right)\sigma(Q(\alpha|x_k,u_{\underline{k}})|X_{mk},X_{m\underline{k}})}{p(x_k,u_{\underline{k}})F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}\right)^2 p_{\underline{k}}(u_{\underline{k}})du_{\underline{k}} \\
&= \frac{(1+o_p(1))}{nh_2^{d-1}}\frac{\left(\int L^2(v_{\underline{k}})dv_{\underline{k}}\right)\sigma^2(Q(\alpha|x_k,X_{m\underline{k}})|X_{mk},X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}})}{p^2(x_k,X_{m\underline{k}})\left(F'(Q(\alpha|x_k,X_{m\underline{k}})|x_k,X_{m\underline{k}})\right)^2} \\
&= o_p(1).
\end{aligned}$$

To calculate the variance of $\Delta_n^{(1.2)}(\alpha|x_k))$, we use a similar decomposition as for $\Delta_n^{(1.1)}(\alpha|x_k))$

$$\begin{aligned}
&\mathrm{Var}(\sqrt{nh_1}\Delta_n^{(1.2)}(\alpha|x_k)) \\
&= \mathrm{Var}\left(\frac{\sqrt{nh_1}}{n}\sum_{m=1}^{n}K_{h_1}(x_k-X_{mk})\,\varepsilon_m E[\beta_m(\alpha|x_k)|\mathbf{X}_m]\right) \\
&\quad +\mathrm{Var}\left(\frac{\sqrt{nh_1}}{n}\sum_{m=1}^{n}K_{h_1}(x_k-X_{mk})\,\varepsilon_m(\beta_m(\alpha|x_k)-E[\beta_m(\alpha|x_k)|\mathbf{X}_m])\right).
\end{aligned}$$

The first term yields

$$
\begin{aligned}
& \operatorname{Var}\left(\frac{\sqrt{nh_1}}{n}\sum_{m=1}^{n} K_{h_1}\left(x_k - X_{mk}\right)\varepsilon_m E[\beta_m(\alpha|x_k)|\mathbf{X}_m]\right) \\
= \;& h_1 \int K_{h_1}^2\left(x_k - v_k\right)\left(\int \frac{L_{H_2}\left(u_{\underline{k}} - v_{\underline{k}}\right)\sigma(Q(\alpha|x_k,u_{\underline{k}})|v_k,v_{\underline{k}})p_{\underline{k}}(u_{\underline{k}})}{p(x_k,u_{\underline{k}})F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}du_{\underline{k}}\right)^2 p(v_k,v_{\underline{k}})dv_k dv_{\underline{k}} \\
= \;& \int K^2(w_k)p(x_1 - h_1 w_k, v_{\underline{k}}) \\
& \left(\int \frac{L(u'_{\underline{k}})\sigma(Q(\alpha|x_k,v_{\underline{k}} - h_2 u'_{\underline{k}})|x_k - h_1 w_k, v_{\underline{k}})p_{\underline{k}}(v_{\underline{k}} - h_2 u'_{\underline{k}})}{p(x_k,v_{\underline{k}} - h_2 u'_{\underline{k}})F'(Q(\alpha|x_k,v_{\underline{k}} - h_2 u'_{\underline{k}})|x_k,v_{\underline{k}} - h_2 u'_{\underline{k}})}du'_{\underline{k}}\right)^2 dw_k dv_{\underline{k}} \\
= \;& (1+o(1))\int K^2(w_k)\sigma^2(Q(\alpha|x_k,v_{\underline{k}})|x_k - h_1 w_k, v_{\underline{k}})p(x_k - h_1 w_k, v_{\underline{k}}) \\
& \left(\frac{p_{\underline{k}}(v_{\underline{k}})}{p(x_k,v_{\underline{k}})F'(Q(\alpha|x_k,v_{\underline{k}})|x_k,v_{\underline{k}})} + h_2^q \frac{\sum_{|\nu_{\underline{k}}|=q}\left(\int {u'}_{\underline{k}}^{\nu_{\underline{k}}} L(u'_{\underline{k}})du'_{\underline{k}}\right) D^{\nu_{\underline{k}}} p_{\underline{k}}(v_{\underline{k}})}{p(x_k,v_{\underline{k}})F'(Q(\alpha|x_k,v_{\underline{k}})|x_k,v_{\underline{k}})}\right)^2 dw_k dv_{\underline{k}} \\
= \;& (1+o(1))\left(\int K^2(w_k)dw_k\right)\int \frac{\sigma^2(Q(\alpha|x_k,v_{\underline{k}})|x_k,v_{\underline{k}})p_{\underline{k}}^2(v_{\underline{k}})}{p(x_k,v_{\underline{k}})(F'(Q(\alpha|x_k,v_{\underline{k}})|x_k,v_{\underline{k}}))^2}dv_{\underline{k}} + o(1) \\
= \;& (1+o(1))\left(\int K^2(w_k)dw_k\right)\int \frac{\alpha(1-\alpha)p_{\underline{k}}^2(v_{\underline{k}})}{p(x_k,v_{\underline{k}})(F'(Q(\alpha|x_k,v_{\underline{k}})|x_k,v_{\underline{k}}))^2}dv_{\underline{k}} + o(1) \\
= \;& (1+o(1))s_k^2(\alpha|x_k) + o(1)
\end{aligned}
$$

and the second term

$$
\begin{aligned}
& \operatorname{Var}\left(\frac{\sqrt{nh_1}}{n}\sum_{m=1}^{n} K_{h_1}\left(x_k - X_{mk}\right)\varepsilon_m(\beta_m(\alpha|x_k) - E[\beta_m(\alpha|x_k)|\mathbf{X}_m])\right) \\
= \;& \frac{h_1}{n}E\left[E\left[\left(\sum_{m=1}^{n} K_{h_1}\left(x_k - X_{mk}\right)\varepsilon_m(\beta_m(\alpha|x_k) - E[\beta_m(\alpha|x_k)|\mathbf{X}_m])\right)^2\middle|\mathbf{X}_m\right]\right] \\
= \;& h_1 E\left[E\left[K_{h_1}^2\left(x_k - X_{mk}\right)\left(\beta_m(\alpha|x_k) - E[\beta_m(\alpha|x_k)|\mathbf{X}_m]\right)^2\middle|\mathbf{X}_m\right]\right] \\
\le \;& \frac{h_1}{nh_2^{d-1}}\int \frac{K_{h_1}^2\left(x_k - u_k\right)\left(\int L^2(v_{\underline{k}})dv_{\underline{k}}\right)\sigma^2(Q(\alpha|x_k,u_{\underline{k}})|u_k,u_{\underline{k}})p_{\underline{k}}(u_{\underline{k}})}{p^2(x_k,u_k)(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2}p(u_k,u_{\underline{k}})du_k du_{\underline{k}} \\
= \;& \frac{1}{nh_2^{d-1}}\left(\int K^2(v_k)dv_k\right)\left(\int L^2(v_{\underline{k}})dv_{\underline{k}}\right)\frac{\alpha(1-\alpha)p_{\underline{k}}(u_{\underline{k}})}{p^2(x_k,u_{\underline{k}})(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2} \\
= \;& O\left(\frac{1}{nh_2^{d-1}}\right).
\end{aligned}
$$

Combining these results, we obtain

$$\operatorname{Var}(\sqrt{nh_1}\Delta_n^{(1.2)}(\alpha|x_k)) = s_k^2(\alpha|x_k) + o(1).$$

A similar calculation shows that Lyapunov's condition for $\delta = 2$ is satisfied for the leading term of $\Delta_n^{(1.2)}$, that is

$$\begin{aligned}
&\sum_{m=1}^{n} E\left[\frac{\sqrt{nh_1}}{n} K_{h_1}\left(x_k - X_{mk}\right)\varepsilon_m E[\beta_m(\alpha|x_k)|\mathbf{X}_m]\right]^4 \\
= \;& \frac{E[\varepsilon_1^4]h_1^2}{n} \int K_{h_1}^4\left(x_k - u_k\right)\sigma^4(Q(\alpha|x_k, v_{\underline{k}})|u_k, u_{\underline{k}}) p(u_k, u_{\underline{k}}) \\
&\left(\frac{p_{\underline{k}}(u_{\underline{k}})}{p(x_k, u_{\underline{k}})F'(Q(\alpha|x_k, u_{\underline{k}})|x_k, u_{\underline{k}})} + \frac{h_2^q}{q!}\frac{\sum_{|\nu_{\underline{k}}|=q}\left(\int v_{\underline{k}}^{\nu_{\underline{k}}} L(v_{\underline{k}})dv_{\underline{k}}\right) D^{\nu_{\underline{k}}} p_{\underline{k}}(v_{\underline{k}})}{p(x_k, u_{\underline{k}})F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}})}\right)^4 du_k du_{\underline{k}} \\
= \;& \frac{E[\varepsilon_1^4]}{nh_1}\left(\int K^4(v)dv\right)\int \frac{\alpha^2(1-\alpha)^2 p_{\underline{k}}^4(u_{\underline{k}})}{(F'(Q(\alpha|x_k, u_{\underline{k}})|x_k, u_{\underline{k}}))^4 p^3(x_k, u_{\underline{k}})} du_{\underline{k}} + O\left(\frac{h_2^q}{nh_1}\right) = O\left(\frac{1}{nh_1}\right),
\end{aligned}$$

which establishes the weak convergence

$$\sqrt{nh_1}\Delta_n^{(1.2)}(\alpha|x_k) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s_k^2(\alpha|x_k)).$$

A combination with (3.18) and (3.20) yields the assertion of Theorem 3.6.

□

3.5.5 Proof of Theorem 3.8

As in the previous proofs, we assume $N = n$, and that the transformation function G is in fact a uniform distribution . We use again the law of the iterated logarithm and Lemma 1.8 to approximate $q_k(\alpha|x_k)$. The following decomposition is derived and is analyzed more precisely in the rest of the proof

$$\begin{aligned}
\tilde{q}_k(\alpha|x_k) - q_k(\alpha|x_k) &= \frac{1}{n}\sum_{j=1}^{n}\left(\tilde{Q}_I(\alpha|x_k, X_{j\underline{k}}) - Q_n(\alpha|x_k, X_{j\underline{k}})\right) + o_p\left(\frac{1}{\sqrt{nh_1}}\right) \\
&= \tilde{\Delta}_n^{(1)}(\alpha|x_k) + \frac{1}{2}\tilde{\Delta}_n^{(2)}(\alpha|x_k) + o_p\left(\frac{1}{\sqrt{nh_1}}\right),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\Delta}_n^{(1)}(\alpha|x_k) &= -\frac{1}{n^2 h_m}\sum_{j=1}^{n}\sum_{i=1}^{n} K_m\left(\frac{F(\frac{i}{n}|x_k, X_{j\underline{k}}) - \alpha}{h_m}\right)\left(\tilde{F}(\frac{i}{n}|x_k, X_{j\underline{k}}) - F(\frac{i}{n}|x_k, X_{j\underline{k}})\right), \\
\tilde{\Delta}_n^{(2)}(\alpha|x_k) &= -\frac{1}{n^2 h_m^2}\sum_{j=1}^{n}\sum_{i=1}^{n} K_m'\left(\frac{\tilde{\xi}_i - \alpha}{h_m}\right)\left(\tilde{F}(\frac{i}{n}|x_k, X_{j\underline{k}}) - F(\frac{i}{n}|x_k, X_{j\underline{k}})\right)^2
\end{aligned}$$

are obtained by a Taylor expansion with the number $\tilde{\xi}_i = \tilde{\xi}_i(\alpha, x_k, X_{j\underline{k}})$ satisfying

$$|\tilde{\xi}_i - F(\frac{i}{n}|x_k, X_{j\underline{k}})| \leq |\tilde{F}(\frac{i}{n}|x_k, X_{j\underline{k}}) - F(\frac{i}{n}|x_k, X_{j\underline{k}})|$$

for $i = 1, \ldots, n$. The second term $\tilde{\Delta}_n^{(2)}(\alpha|x_k)$ can be treated as in the proof of Theorem 3.6 and it follows that $\tilde{\Delta}_n^{(2)}(\alpha|x_k) = o_p\left(\frac{1}{\sqrt{nh_1}}\right)$. Now we turn to the remaining term $\tilde{\Delta}_n^{(1)}(\alpha|x_k)$, which requires a more sophisticated treatment and decomposition using the special structure of the internalized Nadaraya-Watson estimate

$$\begin{aligned}
\tilde{\Delta}_n^{(1)}(\alpha|x_k) &= -\frac{(1+o(1))}{nh_m}\sum_{j=1}^{n} K_m\left(\frac{F(t|x_k, X_{j\underline{k}}) - \alpha}{h_m}\right)\left(\tilde{F}(t|x_k, X_{j\underline{k}}) - F(t|x_k, X_{j\underline{k}})\right) dt \\
&= (1+o(1))(\tilde{\Delta}_n^{(1.1)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.2)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.3)}(\alpha|x_k)),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\Delta}_n^{(1.1)}(\alpha|x_k) &= -\frac{1}{nh_m}\sum_{j=1}^{n}\int_0^1 K_m\left(\frac{F(t|x_k, X_{j\underline{k}}) - \alpha}{h_m}\right) \times \\
&\quad \left[\frac{1}{n}\sum_{m=1}^{n}\frac{K_{h_1}(x_k - X_{mk})\, L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right)\sigma(t|X_{mk}, X_{m\underline{k}})\varepsilon_m}{\hat{p}(X_{mk}, X_{m\underline{k}})}\right] dt, \\
\tilde{\Delta}_n^{(1.2)}(\alpha|x_k) &= -\frac{1}{nh_m}\sum_{j=1}^{n}\int_0^1 K_m\left(\frac{F(t|x_k, X_{j\underline{k}}) - \alpha}{h_m}\right) \times \\
&\quad \left[\frac{1}{n}\sum_{m=1}^{n}\frac{K_{h_1}(x_k - X_{mk})\, L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right)(F(t|X_{mk}, X_{m\underline{k}}) - F(t|x_1, X_{j\underline{k}}))}{\hat{p}(X_{mk}, X_{m\underline{k}})}\right] dt, \\
\tilde{\Delta}_n^{(1.3)}(\alpha|x_k) &= -\frac{1}{nh_m}\sum_{j=1}^{n}\int_0^1 K_m\left(\frac{F(t|x_k, X_{j\underline{k}}) - \alpha}{h_m}\right) F(t|x_k, X_{j\underline{k}}) \times \\
&\quad \left[\frac{1}{n}\sum_{m=1}^{n}\frac{K_{h_1}(x_k - X_{mk})\, L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right)}{\hat{p}(X_{mk}, X_{m\underline{k}})} - 1\right] dt.
\end{aligned}$$

This is a similar decomposition as in Kim et al. (1999). Now we study these terms step by step. The first term $\tilde{\Delta}_n^{(1.1)}(\alpha|x_k)$ has apparently zero expectation, since $E[\varepsilon_m] = 0$ for $m = 1, \ldots, n$. To calculate the variance, we use a similar analysis as in Chen et al. (1996).

$$\begin{aligned}
\tilde{\Delta}_n^{(1.1)}(\alpha|x_k) &= -\frac{1}{n^2 h_m}\sum_{j=1}^{n}\sum_{m=1}^{n}\int_0^1 K_m\left(\frac{F(t|x_k, X_{j\underline{k}}) - \alpha}{h_m}\right) \\
&\quad \frac{K_{h_1}(x_k - X_{mk})\, L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right)\sigma(t|X_{mk}, X_{m\underline{k}})\varepsilon_m}{\hat{p}(X_{mk}, X_{m\underline{k}})} dt \\
&= -\frac{1}{n^2}\sum_{j=1}^{n}\sum_{m=1}^{n}\int_{-\frac{\alpha}{h_m}}^{\frac{1-\alpha}{h_m}} K_m(t') \times \\
&\quad \frac{K_{h_1}(x_k - X_{mk})\, L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right)\sigma(Q(\alpha + h_m t'|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}})\varepsilon_m}{F'(Q(\alpha + h_m t'|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})\hat{p}(X_{mk}, X_{m\underline{k}})} dt'
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{(1+o_p(1))}{n^2} \times \\
&\quad \sum_{j=1}^{n}\sum_{m=1}^{n} \frac{K_{h_1}\left(x_k - X_{mk}\right) L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \sigma(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}})\varepsilon_m}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})\hat{p}(X_{mk}, X_{m\underline{k}})} \\
&= -\frac{(1+o_p(1))}{n} \sum_{m=1}^{n} \frac{K_{h_1}\left(x_k - X_{mk}\right)\varepsilon_m \eta_m(\alpha|x_k)}{p(X_{mk}, X_{m\underline{k}})},
\end{aligned}
$$

where

$$
\eta_m(\alpha|x_k) = \frac{1}{n}\sum_{j=1}^{n} \frac{L_{H_2}\left(X_{j\underline{k}} - X_{m\underline{k}}\right) \sigma(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}})}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}.
$$

To understand why we can substitute $\hat{p}$ in the denominator by p, we multiply by

$$
\frac{\hat{p}(X_{mk}, X_{m\underline{k}})}{p(X_{mk}, X_{m\underline{k}})} + \left(1 - \frac{\hat{p}(X_{mk}, X_{m\underline{k}})}{p(X_{mk}, X_{m\underline{k}})}\right)
$$

and use the fact that

$$
(p(X_{mk}, X_{m\underline{k}}) - \hat{p}(X_{mk}, X_{m\underline{k}}))^2 = O_p\left(h_1^4 + h_2^{2q} + \frac{1}{nh_1h_2^{d-1}}\right).
$$

Then the term $\eta_m(\alpha|x_k)$ is split up into two terms

$$
\eta_m(\alpha|x_k) = E[\eta_m(\alpha|x_k)|\mathbf{X}_m] + (\eta_m(\alpha|x_k) - E[\eta_m(\alpha|x_k)|\mathbf{X}_m]),
$$

which yields two uncorrelated terms decomposing $\tilde{\Delta}_n^{(1.1)}(\alpha|x_k)$ by

$$
\tilde{\Delta}_n^{(1.1)}(\alpha|x_k) = (1+o_p(1))(\tilde{\Delta}_n^{(1.1.a)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.1.b)}(\alpha|x_k)),
$$

where

$$
\begin{aligned}
\tilde{\Delta}_n^{(1.1.a)}(\alpha|x_k) &= -\frac{1}{n}\sum_{m=1}^{n} \frac{K_{h_1}\left(x_k - X_{mk}\right) E[\eta_m(\alpha|x_k)|\mathbf{X}_m]\varepsilon_m}{p(X_{mk}, X_{m\underline{k}})}, \\
\tilde{\Delta}_n^{(1.1.b)}(\alpha|x_k) &= -\frac{1}{n}\sum_{m=1}^{n} \frac{K_{h_1}\left(x_k - X_{mk}\right)\left(\eta_m(\alpha|x_k) - E[\eta_m(\alpha|x_k)|\mathbf{X}_m]\right)\varepsilon_m}{p(X_{mk}, X_{m\underline{k}})}.
\end{aligned}
$$

For the first term, we calculate $E[\eta_m(\alpha|x_k)|\mathbf{X}_m]$

$$\begin{aligned}
E[\eta_m(\alpha|x_k)|\mathbf{X}_m] &= \int \frac{L_{H_2}(w_{\underline{k}} - X_{m\underline{k}})\sigma(Q(\alpha|x_k, w_{\underline{k}})|X_{mk}, X_{m\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})} p_{\underline{k}}(w_{\underline{k}})dw_{\underline{k}} \\
&= \int \frac{L(v_{\underline{k}})\sigma(Q(\alpha|x_k, X_{m\underline{k}} + h_2 v_{\underline{k}})|X_{mk}, X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}} + h_2 v_{\underline{k}})}{F'(Q(\alpha|x_k, X_{m\underline{k}} + h_2 v_{\underline{k}})|x_k, X_{m\underline{k}} + h_2 v_{\underline{k}})} dv_{\underline{k}} \\
&= \int \frac{L(v_{\underline{k}})\sigma(Q(\alpha|x_k, X_{m\underline{k}})|X_{mk}, X_{m\underline{k}})}{F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})} \times \\
&\quad \left(p_{\underline{k}}(X_{m\underline{k}}) + \sum_{|\nu_{\underline{k}}|=q} \frac{h_2^q v_{\underline{k}}^{\nu_{\underline{k}}}}{q!} D^{\nu_{\underline{k}}} p_{\underline{k}}(X_{m\underline{k}}) \right) dv_{\underline{k}} \\
&= \frac{\sigma(Q(\alpha|x_k, X_{m\underline{k}})|X_{mk}, X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}})}{F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})} + \\
&\quad \frac{h_2^q}{q!}\sigma(Q(\alpha|x_k, X_{m\underline{k}})|X_{mk}, X_{m\underline{k}}) \frac{\sum_{|\nu_{\underline{k}}|=q}\left(\int v_{\underline{k}}^{\nu_{\underline{k}}} L(v_{\underline{k}})dv_{\underline{k}}\right) D^{\nu_{\underline{k}}} p_{\underline{k}}(X_{m\underline{k}})}{F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})} \\
&= \frac{\sigma(Q(\alpha|x_k, X_{m\underline{k}})|X_{mk}, X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}})}{F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})} + O_p(h_2^q).
\end{aligned}$$

We obtain for $v = (v_k, v_{\underline{k}})$

$$\begin{aligned}
&\mathrm{Var}(\sqrt{nh_1}\tilde{\Delta}_n^{(1.1.a)}(\alpha|x_k)) \\
&= \frac{nh_1}{n} \int \left(\frac{K_{h_1}(x_k - v_k)E[\eta_m(\alpha|x_k)|\mathbf{X}_m = \mathbf{v}]}{p(v_k, v_{\underline{k}})} \right)^2 p(v_k, v_{\underline{k}})dv_k dv_{\underline{k}} \\
&= h_1 \int \frac{K_{h_1}^2(x_k - v_k)\, E[\eta_m(\alpha|x_k)|\mathbf{X}_m = \mathbf{v}]^2}{p(v_k, v_{\underline{k}})} dv_k dv_{\underline{k}} \\
&= \int \frac{K^2(w_k)\sigma^2(Q(\alpha|x_k, v_{\underline{k}})|x_k - h_1 w_k, v_{\underline{k}})p_{\underline{k}}^2(v_{\underline{k}})}{p(x_k - h_1 w_k, v_{\underline{k}})(F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} dw_k dv_{\underline{k}} \\
&\quad + \frac{h_2^q}{q!} \int \frac{K^2(w_k)\sigma^2(Q(\alpha|x_k, v_{\underline{k}})|x_k - h_1 w_k, v_{\underline{k}})p_{\underline{k}}(v_{\underline{k}})}{p(x_k - h_1 w_k, v_{\underline{k}})(F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} \times \\
&\quad \sum_{|\nu_{\underline{k}}|=q} \left(\int w_{\underline{k}}^{\nu_{\underline{k}}} L(w_{\underline{k}})dw_{\underline{k}}\right) D^{\nu_{\underline{k}}} p_{\underline{k}}(v_{\underline{k}}) dw_k dv_{\underline{k}} \\
&\quad + \frac{h_2^{2q}}{(q!)^2} \int \frac{K^2(w_k)\sigma^2(Q(\alpha|x_k, v_{\underline{k}})|x_k - h_1 w_k, v_{\underline{k}})}{p(x_k - h_1 w_k, v_{\underline{k}})(F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} \times \\
&\quad \left(\sum_{|\nu_{\underline{k}}|=q} \left(\int w_{\underline{k}}^{\nu_{\underline{k}}} L(w_{\underline{k}})dw_{\underline{k}}\right) D^{\nu_{\underline{k}}} p_{\underline{k}}(v_{\underline{k}}) \right)^2 dw_k dv_{\underline{k}} \\
&= ||K||_2^2 \int \frac{\sigma^2(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}})p_{\underline{k}}(v_{\underline{k}})}{p(x_k, v_{\underline{k}})(F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} dv_{\underline{k}} + O(h_1 + h_2^q)
\end{aligned}$$

$$= \quad ||K||_2^2 \int \frac{\alpha(1-\alpha)p_{\underline{k}}(v_{\underline{k}})}{p(x_k, v_{\underline{k}})(F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} dv_{\underline{k}} + O(h_1 + h_2^q).$$

Furthermore we can show the Lyapunov condition for the leading term of $\tilde{\Delta}_n^{(1.1.a)}(\alpha|x_k)$ with $\delta = 2$

$$\begin{aligned}
&\sum_{m=1}^{n} E\left[\left(\frac{\sqrt{nh_1}}{n} \frac{K_{h_1}(x_k - X_{mk})\,\sigma(Q(\alpha|x_k, X_{m\underline{k}})|X_{mk}, X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}})\varepsilon_m}{p(X_{mk}, X_{m\underline{k}})F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})}\right)^4\right] \\
&= \frac{E[\varepsilon_1^4]h_1^2}{n} \int \left(\frac{K_{h_1}(x_k - w_k)\,\sigma(Q(\alpha|x_k, w_{\underline{k}})|w_k, w_{\underline{k}})p_{\underline{k}}(w_{\underline{k}})}{p(w_k, w_{\underline{k}})F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}\right)^4 p(w_k, w_{\underline{k}})dw_k dw_{\underline{k}} \\
&= \frac{E[\varepsilon_1^4](1+o(1))}{nh_1}\left(\int K^4(w_k)dw_k\right)\int \frac{p_{\underline{k}}^4(w_{\underline{k}})\sigma^4(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}{p^3(x_k, w_{\underline{k}})(F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}}))^4} dw_{\underline{k}} \\
&= \frac{E[\varepsilon_1^4](1+o(1))}{nh_1}\left(\int K^4(w_k)dw_k\right)\int \frac{p_{\underline{k}}^4(w_{\underline{k}})\alpha^2(1-\alpha)^2}{p^3(x_k, w_{\underline{k}})(F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}}))^4} dw_{\underline{k}}.
\end{aligned}$$

Therefore, we have

$$\sqrt{nh_1}\tilde{\Delta}_n^{(1.1.a)}(\alpha|x_k) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, ||K||_2^2 \int \frac{\alpha(1-\alpha)p_{\underline{k}}(v_{\underline{k}})}{p(x_k, v_{\underline{k}})(F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} dv_{\underline{k}}\right).$$

For the second term, we use

$$\begin{aligned}
&E\left[(\eta_m(\alpha|x_k) - E[\eta_m(\alpha|x_k)|\mathbf{X}_m])^2|\mathbf{X}_m\right] \\
&\leq \frac{1}{n}\int \left(\frac{L_{H_2}(w_{\underline{k}} - X_{m\underline{k}})\sigma(Q(\alpha|x_k, w_{\underline{k}})|X_{mk}, X_{m\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}\right)^2 p_{\underline{k}}(w_{\underline{k}})dw_{\underline{k}} \\
&= \frac{1}{nh_2^{d-1}} \int \frac{L^2(v_{\underline{k}})\sigma^2(Q(\alpha|x_k, X_{m\underline{k}} + h_2 v_{\underline{k}})|X_{mk}, X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}} + h_2 v_{\underline{k}})}{(F'(Q(\alpha|x_k, X_{m\underline{k}} + h_2 v_{\underline{k}}))^2} dv_{\underline{k}} \\
&= \frac{1}{nh_2^{d-1}}||L||_2^2 \frac{\sigma^2(Q(\alpha|x_k, X_{m\underline{k}})|X_{mk}, X_{m\underline{k}})p_{\underline{k}}(X_{m\underline{k}})}{(F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}}))^2} \\
&\quad + O_p\left(\frac{1}{nh_2^{d-2}}\right).
\end{aligned}$$

Using the leading part of $E\left[(\eta_m(\alpha|x_k) - E[\eta_m(\alpha|x_k)|\mathbf{X}_m])^2|\mathbf{X}_m\right]$, we get

$$\begin{aligned}
\mathrm{Var}(\tilde{\Delta}_n^{(1.1.b)}(\alpha|x_k)) &\leq \frac{(1+o(1))}{n^2 h_1 h_2^{d-1}}||L||_2^2 \int \frac{K^2(w_k)\sigma^2(Q(\alpha|x_k, v_{\underline{k}})|x_k - h_1 w_k, v_{\underline{k}}))p_{\underline{k}}(v_{\underline{k}})}{p(x_k - h_1 w_k, v_{\underline{k}})((F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} dw_k dv_{\underline{k}} \\
&= O\left(\frac{1}{n^2 h_1 h_2^{d-1}}\right) = o\left(\frac{1}{nh_1}\right).
\end{aligned}$$

This means that $\tilde{\Delta}_n^{(1.1.b)}(\alpha|x_k)$ has no further influence on the variance and we have

$$\sqrt{nh_1}\tilde{\Delta}_n^{(1.1)}(\alpha|x_k) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, ||K||_2^2 \int \frac{\alpha(1-\alpha)p_{\underline{k}}(v_{\underline{k}})}{p(x_k, v_{\underline{k}})(F'(Q(\alpha|x_k, v_{\underline{k}})|x_k, v_{\underline{k}}))^2} dv_{\underline{k}}\right).$$

Now we turn to the term $\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)$

$$\begin{aligned}
\tilde{\Delta}_n^{(1.2)}(\alpha|x_k) &= -\frac{(1+o_p(1))}{n^2}\sum_{j=1}^n\sum_{m=1}^n \frac{K_{h_1}(x_k - X_{mk})L_{H_2}(X_{j\underline{k}} - X_{m\underline{k}})}{p(X_{mk}, X_{m\underline{k}})F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})} \times \\
&\quad (F(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})) \\
&= -\frac{(1+o_p(1))}{n}\sum_{j=1}^n \tilde{\beta}_j(\alpha|x_k),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\beta}_j(\alpha|x_k) &= \frac{1}{n}\sum_{m=1}^n \frac{K_{h_1}(x_k - X_{mk})L_{H_2}(X_{j\underline{k}} - X_{m\underline{k}})}{p(X_{mk}, X_{m\underline{k}})F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})} \times \\
&\quad (F(Q(\alpha|x_k, X_{j\underline{k}})|X_{mk}, X_{m\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})).
\end{aligned}$$

Again we break $\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)$ into two uncorrelated parts

$$\tilde{\Delta}_n^{(1.2)}(\alpha|x_k) = (1+o_p(1))(\tilde{\Delta}_n^{(1.2.a)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.2.b)}(\alpha|x_k))$$

with

$$\begin{aligned}
\tilde{\Delta}_n^{(1.2.a)}(\alpha|x_k) &= -\frac{1}{n}\sum_{j=1}^n E[\tilde{\beta}_j(\alpha|x_k)|\mathbf{X}_j], \\
\tilde{\Delta}_n^{(1.2.b)}(\alpha|x_k) &= -\frac{1}{n}\sum_{j=1}^n \left(\tilde{\beta}_j(\alpha|x_k) - E[\tilde{\beta}_j(\alpha|x_k)|\mathbf{X}_j]\right).
\end{aligned}$$

First we calculate the expectation.

$$\begin{aligned}
E[\tilde{\beta}_j(\alpha|x_k)|\mathbf{X}_j] &= \int \frac{K_{h_1}(x_k - w_k)L_{H_2}(X_{j\underline{k}} - w_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})} \times \\
&\quad (F(Q(\alpha|x_k, X_{j\underline{k}})|w_k, w_{\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})dw_k dw_{\underline{k}} \\
&= \int \frac{K(v_k)L(v_{\underline{k}})}{F'(Q(\alpha|x_k, X_{j\underline{k}} - h_2 v_{\underline{k}})|x_k, X_{j\underline{k}} - h_2 v_{\underline{k}})} \times \\
&\quad (F(Q(\alpha|x_k, X_{j\underline{k}})|x_k - h_1 v_k, X_{j\underline{k}} - h_2 v_{\underline{k}}) - F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})dv_k dv_{\underline{k}} \\
&= \frac{\int v_k^2 K(v_k)dv_k}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}\left(\frac{h_1^2}{2}\frac{\partial^2}{\partial x_k^2}F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) + O_p(h_2^q)\right)
\end{aligned}$$

It is straightforward to see that

$$\begin{aligned}
E[\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)] &= (1+o(1))E[\tilde{\Delta}_n^{(1.2.a)}(\alpha|x_k)] \\
&= -(1+o(1))\kappa_2(K)h_1^2\int \frac{\frac{\partial^2}{\partial x_k^2}F(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})}p_{\underline{k}}(x_{\underline{k}})dx_{\underline{k}} \\
&= -(1+o(1))\tilde{b}_{1k}(\alpha|x_k).
\end{aligned}$$

To estimate the variance of $\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)$, we have for the first term

$$\begin{aligned}
&\mathrm{Var}(\tilde{\Delta}_n^{(1.2.a)}(\alpha|x_k)) \\
\leq\ & \frac{1}{n}\int\Bigg(\int \frac{K(v_k)L(v_{\underline{k}})}{F'(Q(\alpha|x_k,w_{\underline{k}}-h_2v_{\underline{k}})|x_k,w_{\underline{k}}-h_2v_{\underline{k}})}\times \\
&\big(F(Q(\alpha|x_k,w_{\underline{k}})|x_k-h_1v_k,w_{\underline{k}}-h_2v_{\underline{k}})-F(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}})\big)\,dv_kdv_{\underline{k}}\Bigg)^2 p_{\underline{k}}(w_{\underline{k}})dw_{\underline{k}} \\
=\ & \frac{1}{n}\int\Bigg(h_1^2\kappa_2(K)\frac{\frac{\partial^2}{\partial x_k^2}F(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}})}{F'(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}})}+ \\
&\frac{h_2^q}{q!}\sum_{|\nu_{\underline{k}}|=q}\Big(\int v_{\underline{k}}^{\nu_{\underline{k}}}L(v_{\underline{k}})dv_{\underline{k}}\Big)\frac{D^{\nu_{\underline{k}}}F(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}})}{F'(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}})}\Bigg)^2 p_{\underline{k}}(w_{\underline{k}})dw_{\underline{k}} \\
=\ & o\left(\frac{1}{nh_1}\right).
\end{aligned}$$

For the second term, we use

$$\begin{aligned}
&E[(\beta_j(\alpha|x_k)-E[\beta_j(\alpha|x_k)|\mathbf{X}_j])^2|\mathbf{X}_j] \\
\leq\ & \frac{1}{n}\int\left(\frac{K_{h_1}(x_k-w_k)L_{H_2}(X_{j\underline{k}}-w_{\underline{k}})}{p(w_k,w_{\underline{k}})F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})}\right)^2 \\
&(F(Q(\alpha|x_k,X_{j\underline{k}})|w_k,w_{\underline{k}})-F(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2p(w_k,w_{\underline{k}})dw_kdw_{\underline{k}} \\
=\ & \frac{1}{nh_1h_2^{d-1}}\int\frac{K^2(v_k)L^2(v_{\underline{k}})}{p(x_k-h_1v_k,X_{j\underline{k}}-h_2v_{\underline{k}})(F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2} \\
&(F(Q(\alpha|x_k,X_{j\underline{k}})|x_k-h_1v_k,X_{j\underline{k}}-h_2v_{\underline{k}})-F(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2dv_kdv_{\underline{k}} \\
=\ & \frac{(1+o_p(1))}{nh_1h_2^{d-1}}\int\frac{K^2(v_k)L^2(v_{\underline{k}})}{p(x_k,X_{j\underline{k}})(F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2} \\
&\left(-h_1v_k\frac{\partial}{\partial x_k}F(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})-h_2\sum_{l\neq k}v_lD^lF(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})\right)^2 dv_kdv_{\underline{k}} \\
=\ & O_p\left(\frac{(h_1+h_2)^2}{nh_1h_2^{d-1}}\right).
\end{aligned}$$

Then it is easily estimated that $\mathrm{Var}(\tilde{\Delta}_n^{(1.2.b)}(\alpha|x_k)) = o\left(\frac{1}{nh_1}\right)$. The term $\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)$ characterizes the first part $\tilde{b}_{1k}(\alpha|x_k)$ of the bias term.
Finally we turn to the last term $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$. We use a similar trick as in the sketch of the proof of Theorem 1.4, multiply the summand by

$$1=\left(\frac{\hat{p}(X_{mk},X_{m\underline{k}})}{p(X_{mk},X_{m\underline{k}})}+\left(1-\frac{\hat{p}(X_{mk},X_{m\underline{k}})}{p(X_{mk},X_{m\underline{k}})}\right)\right)$$

and obtain using the fact that

$$F(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}}) = \alpha$$

as the quantile function is the inverse of the distribution function

$$\begin{aligned}
\tilde{\Delta}_n^{(1.3)}(\alpha|x_k) &= -\frac{(1+o_p(1))}{n}\sum_{j=1}^n \frac{\alpha}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})} \times \\
&\quad \left[\frac{1}{n}\sum_{m=1}^n \frac{K_{h_1}(x_k - X_{mk})L_{H_2}(X_{j\underline{k}} - X_{m\underline{k}})}{\hat{p}(X_{mk}, X_{m\underline{k}})} - 1\right] \\
&= -\frac{(1+o_p(1))}{n}\sum_{j=1}^n \frac{\alpha}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})} \times \\
&\quad \frac{1}{n}\sum_{m=1}^n \frac{K_{h_1}(x_k - X_{mk})L_{H_2}(X_{j\underline{k}} - X_{m\underline{k}})}{\hat{p}(X_{mk}, X_{m\underline{k}})} \times \\
&\quad \left(\frac{\hat{p}(X_{mk}, X_{m\underline{k}})}{p(X_{mk}, X_{m\underline{k}})} + \left(1 - \frac{\hat{p}(X_{mk}, X_{m\underline{k}})}{p(X_{mk}, X_{m\underline{k}})}\right)\right) \\
&\quad + \frac{(1+o_p(1))}{n}\sum_{j=1}^n \frac{\alpha}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})} \\
&= (1+o_p(1))(\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k)),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k) &= -\frac{\alpha}{n}\sum_{j=1}^n \gamma_j^{(1)}(\alpha|x_k) \\
\tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k) &= -\frac{1}{n}\sum_{j=1}^n \frac{\alpha\gamma_j^{(2)}(\alpha|x_k)}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})} \\
\gamma_j^{(1)}(\alpha|x_k) &= \frac{1}{n}\sum_{m=1}^n \left(\frac{K_{h_1}(x_k - X_{jk})L_{H_2}(X_{m\underline{k}} - X_{j\underline{k}})}{p(X_{jk}, X_{j\underline{k}})F'(Q(\alpha|x_k, X_{m\underline{k}})|x_k, X_{m\underline{k}})}\right. \\
&\quad \left. - \frac{1}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}\right) \\
\gamma_j^{(2)}(\alpha|x_k) &= \frac{1}{n}\sum_{m=1}^n \frac{p(X_{mk}, X_{m\underline{k}}) - \hat{p}(X_{mk}, X_{m\underline{k}})}{p^2(X_{mk}, X_{m\underline{k}})} K_{h_1}(x_k - X_{mk})L_{H_2}(X_{j\underline{k}} - X_{m\underline{k}}).
\end{aligned}$$

Note that we changed the indexing for the second term in $\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k)$, which is helpful for the further analysis. We treat both terms as before and split them up into two uncorrelated terms induced by $E[\gamma_j^{(l)}(\alpha|x_k)|\mathbf{X}_j]$ and $\left(\gamma_j^{(l)}(\alpha|x_k) - E[\gamma_j^{(l)}(\alpha|x_k)|\mathbf{X}_j]\right)$ for $l = 1, 2$.

For $\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k)$, we calculate $E[\gamma_j(\alpha|x_k)|\mathbf{X}_j]$ in the first step

$$\begin{aligned}
E[\gamma_j^{(1)}(\alpha|x_k)|\mathbf{X}_j] &= \int\Bigg(\frac{K_{h_1}(x_k - X_{jk})L_{H_2}(u_{\underline{k}} - X_{j\underline{k}})}{p(X_{jk}, X_{j\underline{k}})F'(Q(\alpha|x_k, u_{\underline{k}})|x_k, u_{\underline{k}})} \\
&\quad - \frac{1}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}\Bigg)p_{\underline{k}}(u_{\underline{k}})du_{\underline{k}} \\
&= \int \frac{K_{h_1}(x_k - X_{jk})L(v_{\underline{k}})p_{\underline{k}}(X_{j\underline{k}} + h_2 v_{\underline{k}})}{p(X_{jk}, X_{j\underline{k}})F'(Q(\alpha|x_k, X_{j\underline{k}} + h_2 v_{\underline{k}})|x_k, X_{j\underline{k}} + h_2 v_{\underline{k}})}dv_{\underline{k}} \\
&\quad - \frac{1}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})} \\
&= \frac{1}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}\left(\frac{K_{h_1}(x_k - X_{jk})p_{\underline{k}}(X_{j\underline{k}})}{p(X_{jk}, X_{j\underline{k}})} - 1\right) \\
&\quad + \frac{h_2^q}{q!}\frac{K_{h_1}(x_k - X_{jk})}{p(X_{jk}, X_{j\underline{k}})} \times \\
&\quad \sum_{|\nu_{\underline{k}}|=q}\left(\int v_{\underline{k}}^{\nu_{\underline{k}}}L(v_{\underline{k}})dv_{\underline{k}}\right)D^{\nu_{\underline{k}}}\left(\frac{p_{\underline{k}}(X_{j\underline{k}})}{F'(Q(\alpha|x_k, X_{j\underline{k}})|x_k, X_{j\underline{k}})}\right).
\end{aligned}$$

By the rule of iterated expectation, we obtain

$$\begin{aligned}
&E[\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k)] \\
&= -E\left[\frac{\alpha}{n}\sum_{j=1}^{n} E\left[\gamma_j^{(1)}(\alpha|x_k)|\mathbf{X}_j\right]\right] \\
&= -\int \frac{\alpha p(w_k, w_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}\left(\frac{K_{h_1}(x_k - w_k)p_{\underline{k}}(w_{\underline{k}})}{p(w_k, w_{\underline{k}})} - 1\right)dw_k dw_{\underline{k}} \\
&\quad + \frac{h_2^q \alpha}{q!}\int K_{h_1}(x_k - w_k)\sum_{|\nu_{\underline{k}}|=q}\left(\int v_{\underline{k}}^{\nu_{\underline{k}}}L(v_{\underline{k}})dv_{\underline{k}}\right)D^{\nu_{\underline{k}}}\left(\frac{p_{\underline{k}}(w_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}\right)dw_k dw_{\underline{k}} \\
&= -\int \frac{\alpha K(w_k')p_{\underline{k}}(w_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}dw_k' dw_{\underline{k}} - \int \frac{\alpha p_{\underline{k}}(w_{\underline{k}})}{F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}})}dw_{\underline{k}} + O(h_2^q) \\
&= o(1).
\end{aligned}$$

To obtain the total expectation of $\Delta_n^{(1.3)}(\alpha|x_k)$, we have to analyze $\Delta_n^{(1.3.b)}(\alpha|x_k)$. First, we calculate $E[\gamma_j^{(2)}(\alpha|x_k)|\mathbf{X}_j]$. Hence, we apply the same trick as in the sketch of the proof of theorem 1.4 and use the asymptotic expectation of the kernel density estimator $\hat{p}$.

$$
\begin{aligned}
E[\gamma_j^{(2)}(\alpha|x_k)|\mathbf{X}_j] &= -E\Bigg[\frac{1}{n}\sum_{m=1}^{n} K_{h_1}(x_k - X_{mk})L_{H_2}(X_{j\underline{k}} - X_{m\underline{k}})\Bigg(\frac{h_1^2\kappa_2(K)\frac{\partial^2}{\partial x_k^2}p(X_{mk}, X_{m\underline{k}})}{p^2(X_{mk}, X_{m\underline{k}})} \\
&\qquad + \frac{\frac{h_2^q}{q!}\sum_{|\nu_{\underline{k}}|=q}\left(\int w_{\underline{k}}^{\nu_{\underline{k}}}L(w_{\underline{k}})dw_{\underline{k}}\right)D^{\nu_{\underline{k}}}p(X_{mk}, X_{m\underline{k}})}{p^2(X_{mk}, X_{m\underline{k}})}\Bigg)\Bigg|\mathbf{X}_j\Bigg] \\
&= -\int \frac{h_1^2\kappa_2(K)\frac{\partial^2}{\partial x_k^2}p(u_k, u_{\underline{k}}) + \frac{h_2^q}{q!}\sum_{|\nu_{\underline{k}}|=q}\left(\int w_{\underline{k}}^{\nu_{\underline{k}}}L(w_{\underline{k}})dw_{\underline{k}}\right)D^{\nu_{\underline{k}}}p(u_k, u_{\underline{k}})}{p(u_k, u_{\underline{k}})} \times \\
&\qquad K_{h_1}(x_k - u_k)L_{H_2}(X_{j\underline{k}} - u_{\underline{k}})du_k du_{\underline{k}} \\
&= -\int \frac{h_1^2\kappa_2(K)\frac{\partial^2}{\partial x_k^2}p(x_k - h_1 v_k, X_{j\underline{k}} - h_2 v_{\underline{k}})}{p(x_k - h_1 v_k, X_{j\underline{k}} - h_2 v_{\underline{k}})}K(v_k)L(v_{\underline{k}})dv_k dv_{\underline{k}} + O_p(h_2^q) \\
&= -h_1^2\kappa_2(K)\frac{\frac{\partial^2}{\partial x_k^2}p(x_k, X_{j\underline{k}})}{p(x_k, X_{j\underline{k}})} + o_p(h_1^2) + O_p(h_2^q).
\end{aligned}
$$

For the expectation of $\Delta_n^{(1.3.b)}(\alpha|x_k)$, we obtain

$$
\begin{aligned}
E[\Delta_n^{(1.3.b)}(\alpha|x_k)] &= (1 + o(1))h_1^2\kappa_2(K)\alpha\int \frac{\frac{\partial^2}{\partial x_k^2}p(x_k, x_{\underline{k}})p_{\underline{k}}(x_{\underline{k}})}{F'(Q(\alpha|x_k, x_{\underline{k}})|x_k, x_{\underline{k}})p(x_k, x_{\underline{k}})}dx_{\underline{k}} \\
&= (1 + o(1))\tilde{b}_{2k}(\alpha|x_k).
\end{aligned}
$$

Moreover, we have $E[\Delta_n^{(1.3)}(\alpha|x_k)] = (1 + o(1))\tilde{b}_{2k}(\alpha|x_k)$. This means that the bias is given by

$$
E[\tilde{\Delta}_n^{(1.2)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.3)}(\alpha|x_k)] = -(1 + o(1))(\tilde{b}_{1k}(\alpha|x_k) - \tilde{b}_{2k}(\alpha|x_k)).
$$

To get the variance of $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$, we consider at first the term $\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k)$. We break this term into two uncorrelated terms

$$
\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k) = \tilde{\Delta}_n^{(1.3.a.1)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.3.a.2)}(\alpha|x_k),
$$

where

$$
\begin{aligned}
\tilde{\Delta}_n^{(1.3.a.1)}(\alpha|x_k) &= -\frac{\alpha}{n}\sum_{j=1}^{n} E[\gamma_j^{(1)}(\alpha|x_k)|\mathbf{X}_j], \\
\tilde{\Delta}_n^{(1.3.a.2)}(\alpha|x_k) &= -\frac{\alpha}{n}\sum_{j=1}^{n}\left(\gamma_j^{(1)}(\alpha|x_k) - E[\gamma_j^{(1)}(\alpha|x_k)|\mathbf{X}_j]\right).
\end{aligned}
$$

To calculate the variance, we observe

$$
\begin{aligned}
& \operatorname{Var}(\sqrt{nh_1}\tilde{\Delta}_n^{(1.3.a.1)}(\alpha|x_k)) \\
= \; & \frac{h_1\alpha^2}{n}\sum_{j=1}^n \operatorname{Var}\left(E[\gamma_j^{(1)}(\alpha|x_k)|\mathbf{X}_j]\right) \\
\leq \; & h_1\alpha^2 \int\Bigg(\frac{1}{F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}\left(\frac{K_{h_1}(x_k-u_k)p_{\underline{k}}(u_{\underline{k}})}{p(u_k,u_{\underline{k}})}-1\right) \\
& +\frac{h_2^q}{q!}\frac{K_{h_1}(x_k-u_k)}{p(u_k,u_{\underline{k}})}\sum_{|\nu_{\underline{k}}|=q}\left(\int v_{\underline{k}}^{\nu_{\underline{k}}}L(v_{\underline{k}})dv_{\underline{k}}\right)D^{\nu_{\underline{k}}}\left(\frac{p_{\underline{k}}(u_{\underline{k}})}{F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}\right)\Bigg)^2 p(u_k,u_{\underline{k}})du_k du_{\underline{k}} \\
= \; & h_1\alpha^2\int\left(\frac{K_{h_1}(x_k-u_k)p_{\underline{k}}(u_{\underline{k}})}{p(u_k,u_{\underline{k}})}-1\right)^2\frac{p(u_k,u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2}du_k du_{\underline{k}}+O(h_2^q) \\
= \; & h_1\alpha^2\int\left(\frac{K_{h_1}^2(x_k-u_k)p_{\underline{k}}^2(u_{\underline{k}})}{p^2(u_k,u_{\underline{k}})}-2\frac{K_{h_1}(x_k-u_k)p_{\underline{k}}(u_{\underline{k}})}{p(u_k,u_{\underline{k}})}+1\right)\times \\
& \frac{p(u_k,u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2}du_k du_{\underline{k}}+O(h_2^q) \\
= \; & ||K||_2^2\alpha^2\int\frac{p_{\underline{k}}(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2 p(x_k,u_{\underline{k}})}du_{\underline{k}}+O(h_1+h_2^q).
\end{aligned}
$$

For the second term, we first compute

$$
\begin{aligned}
& E[(\gamma_j^{(1)}(\alpha|x_k)-E[\gamma_j^{(1)}(\alpha|x_k)|\mathbf{X}_j])^2|\mathbf{X}_j] \\
\leq \; & \frac{(1+o_p(1))}{n}\int\left(\frac{K_{h_1}(x_k-X_{jk})L_{H_2}(u_{\underline{k}}-X_{j\mathbf{k}})}{p(X_{jk},X_{j\underline{k}})F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}})}-\frac{1}{F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{\underline{k}})}\right)^2 p_{\underline{k}}(u_{\underline{k}})du_{\underline{k}} \\
= \; & \frac{(1+o_p(1))}{nh_2^{d-1}}||L||_2^2\frac{K_{h_1}^2(x_k-X_{jk})}{p^2(X_{jk},X_{j\underline{k}})(F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2}p_{\underline{k}}(X_{j\underline{k}}) \\
& -2\frac{(1+o_p(1))}{n}\frac{K_{h_1}(x_k-X_{jk})}{p(X_{jk},X_{j\underline{k}})(F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2}p_{\underline{k}}(X_{j\underline{k}}) \\
& +\frac{(1+o_p(1))}{n}\frac{1}{(F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2}.
\end{aligned}
$$

We obtain for the variance of $\tilde{\Delta}_n^{(1.3.a.2)}(\alpha|x_k)$

$$
\begin{aligned}
\operatorname{Var}(\tilde{\Delta}_n^{(1.3.a.2)}(\alpha|x_k)) \;&\leq\; \frac{\alpha^2}{n^2}\sum_{j=1}^n E\left[E[(\gamma_j^{(1)}(\alpha|x_k)-E[\gamma_j^{(1)}(\alpha|x_k)|\mathbf{X}_j])^2|\mathbf{X}_j]\right] \\
&=\; \frac{(1+o(1))\alpha^2}{n^2}\int\frac{p(u_k,u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2}\times
\end{aligned}
$$

$$\left(\frac{||L||_2^2}{h_2^{d-1}}\frac{K_{h_1}^2(x_k-u_k)p_{\underline{k}}(u_{\underline{k}})}{p^2(u_k,u_{\underline{k}})}+\frac{K_{h_1}(x_k-u_k)p_{\underline{k}}(u_{\underline{k}})}{p(u_k,u_{\underline{k}})}+1\right)du_k du_{\underline{k}}$$
$$= O\left(\frac{1}{nh_1}\left(\frac{1}{nh_2^{d-1}}+\frac{h_1}{n}\right)\right)=o\left(\frac{1}{nh_1}\right).$$

Now we analyze the term $\tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k)$. We can split this term up like we did before

$$\tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k)=\tilde{\Delta}_n^{(1.3.b.1)}(\alpha|x_k)+\tilde{\Delta}_n^{(1.3.b.2)}(\alpha|x_k),$$

where

$$\begin{aligned}
\tilde{\Delta}_n^{(1.3.b.1)}(\alpha|x_k) &= -\frac{\alpha}{n}\sum_{j=1}^n\frac{E[\gamma_j^{(2)}(\alpha|x_k)|\mathbf{X}_j]}{F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})}\\
&= \frac{\alpha}{n}\sum_{j=1}^n\frac{h_1^2\kappa_2(K)}{F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})}\left(\frac{\frac{\partial^2}{\partial x_k^2}p(x_k,X_{j\underline{k}})}{p(x_k,X_{j\underline{k}})}+o_p(1)\right)\\
\tilde{\Delta}_n^{(1.3.b.2)}(\alpha|x_k) &= -\frac{\alpha}{n}\sum_{j=1}^n\frac{(\gamma_j^{(2)}(\alpha|x_k)-E[\gamma_j^{(2)}(\alpha|x_k)|\mathbf{X}_j])}{F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})}.
\end{aligned}$$

For the first term, we use only the leading term indicated in the representation above

$$\begin{aligned}
\mathrm{Var}(\tilde{\Delta}_n^{(1.3.b.1)}(\alpha|x_k)) &\le \frac{h_1^4\kappa_2(K)^2\alpha^2}{n^2}E\left[\sum_{j=1}^n\frac{\frac{\partial^2}{\partial x_k^2}p(x_k,X_{j\underline{k}})}{F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}})p(x_k,X_{j\underline{k}})}\right]^2\\
&= \frac{h_1^4\kappa_2(K)^2\alpha^2}{n}\int\frac{\left(\frac{\partial^2}{\partial x_k^2}p(x_k,w_{\underline{k}})\right)^2}{(F'(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}}))^2p^2(x_k,w_{\underline{k}})}p_{\underline{k}}(w_{\underline{k}})dw_{\underline{k}}\\
&= O\left(\frac{h_1^4}{n}\right)=o\left(\frac{1}{nh_1}\right).
\end{aligned}$$

For the second term, we obtain using the asymptotic mean squared error of the kernel density estimator $\hat{p}$

$$\begin{aligned}
\mathrm{Var}(\tilde{\Delta}_n^{(1.3.b.2)}(\alpha|x_k)) &= \frac{\alpha^2}{n^2}E\Bigg[\sum_{j=1}^n\frac{E[(\gamma_j^{(2)}(\alpha|x_k)-E[\gamma_j^{(2)}(\alpha|x_k)|X_j])^2|\mathbf{X}_j]}{(F'(Q(\alpha|x_k,X_{j\underline{k}})|x_k,X_{j\underline{k}}))^2}\Bigg]\\
&\le \frac{\alpha^2}{n}\int\int\frac{h^4\kappa_2(K)\frac{\partial^2}{\partial x_k^2}p(w_k,w_{\underline{k}})+O(h_2^{2q})+\frac{1}{nh_1h_2^{d-1}}||K||_2^2||L||_2^2p(w_k,w_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2p^3(w_k,w_{\underline{k}})}\\
&\quad K_{h_1}^2(x_k-w_k)L_{H_2}^2(u_{\underline{k}}-w_{\underline{k}})dw_k dw_{\underline{k}}p_{\underline{k}}(u_{\underline{k}})du_{\underline{k}}\\
&= \frac{\alpha^2}{nh_1h_2^{d-1}}||K||_2^2||L||_2^2\int\frac{p_{\underline{k}}(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2}\times
\end{aligned}$$

$$\frac{h^4\kappa_2(K)\frac{\partial^2}{\partial x_k^2}p(x_k,u_{\underline{k}})+O(h_2^{2q})+\frac{1}{nh_1h_2^{d-1}}||K||_2^2||L||_2^2p(x_k,u_{\underline{k}})}{p^3(x_k,u_{\underline{k}})}du_{\underline{k}}$$
$$= O\left(\frac{h_1^3}{nh_2^{d-1}}+\frac{1}{nh_1nh_1h_2^{2(d-1)}}\right)=o\left(\frac{1}{nh_1}\right),$$

where the last identity follows by bandwidth condition (B5′). For the variance of $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$, we observe applying the Cauchy-Schwarz inequality for variances

$$\begin{aligned}
\mathrm{Var}(\sqrt{nh_1}\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)) &= \mathrm{Var}(\sqrt{nh_1}\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k))+nh_1\mathrm{Var}(\tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k))\\
&\quad +nh_1\mathrm{Cov}(\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k),\tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k))\\
&\leq \mathrm{Var}(\sqrt{nh_1}\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k))+nh_1\mathrm{Var}(\tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k))\\
&\quad +nh_1\sqrt{\mathrm{Var}(\tilde{\Delta}_n^{(1.3.a)}(\alpha|x_k))\mathrm{Var}(\tilde{\Delta}_n^{(1.3.b)}(\alpha|x_k))}\\
&= (1+o(1))||K||_2^2\int\frac{\alpha^2p_{\underline{k}}(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2p(x_k,u_{\underline{k}})}du_{\underline{k}}\\
&\quad +O\left(h_1+h_2^q\right)+O\left(\frac{1}{nh_2^{d-1}}+\frac{h_1}{n}\right)+O\left(\frac{h_1^4}{n}+\frac{h_1^4}{h_2^{d-1}}+\frac{1}{nh_1h_2^{d-1}}\right)\\
&\quad +O\left(\left(\frac{h_1^4}{n}+\frac{h_1^4}{h_2^{d-1}}+\frac{1}{nh_1h_2^{d-1}}\right)^{1/2}\right)\\
&= (1+o(1))||K||_2^2\int\frac{\alpha^2p_{\underline{k}}(u_{\underline{k}})}{(F'(Q(\alpha|x_k,u_{\underline{k}})|x_k,u_{\underline{k}}))^2p(x_k,u_{\underline{k}})}du_{\underline{k}}+o(1),
\end{aligned}$$

where the last identity follows by the bandwidth conditions (B1′)-(B5′). Moreover we show that $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$ fulfills the Lyapunov condition for $\delta=2$, where we focus on the dominating term in the variance

$$\begin{aligned}
&\sum_{m=1}^n E\left[\left(\frac{1}{n}\frac{\alpha}{F'(Q(\alpha|x_k,X_{m\underline{k}})|x_k,X_{m\underline{k}})}\left(\frac{K_{h_1}(x_k-X_{mk})p_{\underline{k}}(X_{m\underline{k}})}{p(X_{mk},X_{m\underline{k}})}-1\right)\right)^4\right]\\
&= \frac{1}{n^3}\int\frac{\alpha^4}{(F'(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}}))^4}\left[\frac{K_{h_1}(x_k-w_k)p_{\underline{k}}(w_{\underline{k}})}{p(w_k,w_{\underline{k}})}-1\right]^4p(w_k,w_{\underline{k}})dw_kdw_{\underline{k}}\\
&= \frac{1}{n^3}\int\frac{\alpha^4}{(F'(Q(\alpha|x_k,w_{\underline{k}})|x_k,w_{\underline{k}}))^4}\left(\frac{1}{h_3}\frac{K^4(w_k)p_{\underline{k}}^4(w_{\underline{k}})}{p^3(x_k-h_1w_k,w_{\underline{k}})}-\frac{4}{h_1^2}\frac{K^3(w_k)p_{\underline{k}}^3(w_{\underline{k}})}{p^2(x_k-h_1w_k,w_{\underline{k}})}\right.\\
&\qquad\left.+\frac{6}{h_1}\frac{K^2(w_k)p_{\underline{k}}^2(w_{\underline{k}})}{p(x_k-h_1w_k,w_{\underline{k}})}-4K(w_k)p_{\underline{k}}(w_{\underline{k}})+p(w_k,w_{\underline{k}})\right)dw_kdw_{\underline{k}}\\
&= O\left(\frac{1}{n^3h_1^2}+\frac{1}{n^3h_1^2}+\frac{1}{n^3h_1}+\frac{1}{n}\right)=o(1).
\end{aligned}$$

Hence, applying the Central Limit Theorem for $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$, it follows that

$$\sqrt{nh_1}(\tilde{\Delta}_n^{(1.3)}(\alpha|x_k) - \tilde{b}_{2k}(\alpha|x_k)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, ||K||_2^2\alpha^2 \int \frac{p_{\underline{k}}^2(w_{\underline{k}})}{(F'(Q(\alpha|x_k, w_{\underline{k}})|x_k, w_{\underline{k}}))^2 p(x_k, w_{\underline{k}})} dw_{\underline{k}}\right)$$

The terms $\tilde{\Delta}_n^{(1.1)}(\alpha|x_k)$ and $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$ are uncorrelated which gives by the Cramér-Wold device

$$\sqrt{nh_1}(\tilde{\Delta}_n^{(1.1)}(\alpha|x_k) + \tilde{\Delta}_n^{(1.3)}(\alpha|x_k) + \tilde{b}_2(\alpha|x_k)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{s}_k^2(\alpha|x_k)).$$

Actually, the terms $\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)$ and $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$ are correlated, but the covariance can be estimated by $o\left(\frac{1}{nh_1}\right)$ using the Cauchy-Schwarz inequality, since the variance of $\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)$ is of order $O\left(\frac{h_1^4}{n} + \frac{(h_1+h_2)^2}{n^2 h_1 h_2^{d-1}}\right) = o\left(\frac{1}{nh_1}\right)$. Finally, combing all the results for $\tilde{\Delta}_n^{(1.1)}(\alpha|x_k)$, $\tilde{\Delta}_n^{(1.2)}(\alpha|x_k)$, and $\tilde{\Delta}_n^{(1.3)}(\alpha|x_k)$ the assertion of the Theorem follows. □

3.6 Finite sample properties and data analysis

In this section, we compare the distinct estimates of the marginal effects of additive conditional quantiles and of the additive component itself, respectively, in terms of finite sample properties. In particular, we contrast our estimator $\hat{q}_k(\alpha|x_k)$ with a procedure proposed by De Gooijer and Zerom (2003). The estimate of De Gooijer and Zerom (2003), called an average quantile estimator, works in a similar manner as $\hat{q}_k(\alpha|x_k)$ and uses the reweighted Nadaraya-Watson estimator for a conditional distribution function

$$\hat{F}_{RW}(y|\mathbf{x}) = \frac{\sum_{j=1}^n p_j(\mathbf{x}) K_{h_1}(x_k - X_{jk}) L_{H_2}(x_{\underline{k}} - X_{j\underline{k}}) I\{Y_j \leq y\}}{\sum_{j=1}^n p_j(\mathbf{x}) K_{h_1}(x_k - X_{jk}) L_{H_2}(x_{\underline{k}} - X_{j\underline{k}})}$$

with high-dimensional covariates proposed by Hall et al. (1999). The weights $p_j(\mathbf{x})$ are constructed as in section 3.3, but the weights solve the equation

$$\sum_{j=1}^n p_j(\mathbf{x})(X_{jk} - x_k) K_{h_1}(x_k - X_{jk}) L_{H_2}(x_{\underline{k}} - X_{j\underline{k}}) = 0$$

with respect to the kth covariate in order to construct an estimate for the kth additive component. Since this estimate is positive and increasing with respect to y provided the kernel functions K and L are positive, the inverse is easily obtained, which marks a huge advantage of this estimate in contrast to other non-increasing estimates of the conditional distribution function. On the other hand, this method works only as long as the dimension of the covariates $d < 5$. When $d \geq 5$ the bias of the estimate has to be reduced by using negative kernels, which unfortunately destroys the monotonicity property of the reweighted estimator $\hat{F}_{RW}(y|\mathbf{x})$. The estimate of De Gooijer and Zerom (2003) is calculated by reversing the reweighted Nadaraya-Watson estimate and applying

the marginal integration method as for $\hat{q}_k(\alpha|x_k)$. We denote this estimate for the marginal effect of the conditional quantile function by $\breve{q}_k(\alpha|x_k)$ and the estimate of the additive component $\breve{Q}_k(\alpha|x_k)$, respectively.
In the additive quantile regression setting, the estimate of the additive component can be derived from the estimator of the corresponding marginal effect by computing

$$\hat{Q}_k(\alpha|x_k) = \hat{q}_k(\alpha|x_k) - \frac{1}{n}\sum_{j=1}^{n} \hat{q}_k(\alpha|X_{jk}).$$

The same works to construct $\tilde{Q}_k(\alpha|x_k)$. In the following, we investigate the finite sample properties of the estimates:

- $\hat{q}_k(\alpha|x_k)$ with $\hat{F}_{LL}(y|\mathbf{x})$ as the preliminary estimate for the conditional distribution function, and the additive component estimate $\hat{Q}_k(\alpha|x_k)$ as defined above;
- the computational interesting internalized estimate $\tilde{q}_k(\alpha|x_k)$ with additive component estimate $\tilde{Q}_k(\alpha|x_k)$;
- the estimate of De Gooijer and Zerom (2003), $\breve{q}_k(\alpha|x_k)$, and the corresponding estimate of the additive component $\breve{Q}_k(\alpha|x_k)$.

For the sake of convenience, we use a uniform distribution function on the interval

$$[\min(X_{j1}), \max(X_{j1})] \times \ldots \times [\min(X_{jd}), \max(X_{jd})]$$

for the function G in (3.4) to transform the data onto $[0,1]^d$ as a simple and practical choice of G denoted by G_{uni}. Sometimes we also use a normal transformation for G called G_{norm} which is adjusted by an estimated mean and variance through the data.

3.6.1 Simulation studies

The goal of the following studies is to analyze the performance of the proposed methods to estimate an additive conditional quantile function. Some of the models are investigated by other authors as well and seemed to us as an appropriate simulation study. In general, we used Epanechnikov kernels to estimate the conditional distribution function either by local constant or linear techniques. In higher dimensional problems, this means we use a product kernel of several univariate Epanechnikov kernels. In all cases, we apply the Epanechnikov kernel for the monotonizing inversion as well.

Example 3.11 We consider the two-dimensional model

$$Y = 0.75X_1 + 1.5\sin(0.5\pi X_2) + \varepsilon, \tag{3.21}$$

where $\varepsilon \sim \mathcal{N}(0, 0.25^2)$. We assume that the covariates $(X_1, X_2)^T$ are bivariate normal with mean 0, variance 1, and correlation ρ. For the correlation, we distinguish two cases: a weak correlation $\rho = 0.2$ and a strong correlation $\rho = 0.8$. This experiment was originally carried out by De Gooijer and Zerom (2003). The Epanechikov kernel is used to estimate the local linear estimate and the internalized estimate of the conditional distribution function [$\hat{F}_{LL}(y|\mathbf{x})$ and $\hat{F}_{INW}(y|\mathbf{x})$, respectively] as well as to compute the monotonizing inversion, i.e.

$$K(x) = L(x) = K_m(x) = \frac{3}{4}(1 - x^2)I_{[-1,1]}(x).$$

In order to make our results comparable to De Gooijer and Zerom (2003), we chose the bandwidths $h_1 = 3\hat{\sigma}_1 n^{-1/5}$ and $h_2 = \hat{\sigma}_2 n^{-1/5}$, where $\hat{\sigma}_i$ is the standard deviation of the corresponding covariate. The quantile estimates are computed for $\alpha = 0.5$ and sample sizes $n = 100, 200,$ and 400. 41 simulation runs are performed and for each scenario the mean absolute deviation error (MADE) is collected, whereas the observations outside the square $[-2, 2]^2$ are disregarded to avoid boundary effects.

In Table 3.1, we display the results for the finite sample study of the model (3.21) for two cases of G. The results of the performance of $\breve{Q}_k(.5|x_k)$ for $k = 1, 2$ are extracted from De Gooijer and Zerom (2003).

We observe that the internalized marginal integration estimate yields a larger MADE than the local linear approach in all cases. A comparison with the estimates of De Gooijer and Zerom (2003) shows only advantages for the internalized marginal integration estimate, if the second (more oscillating) component is estimated and the data is strongly correlated. In all other cases the estimate of De Gooijer and Zerom (2003) yields a smaller MADE. On the other hand the local linear estimate has a smaller MADE than the estimate of De Gooijer and Zerom (2003), except in the case $\rho = 0.8$, $n = 100, 200$ and 400 for $Q_1(.5|x_1) = 0.75x_1$. The different choices of the transformation function G do not yield large differences. The larger the sample size the smaller the difference. Still for small sample sizes G_{norm} is slightly better than G_{uni}.

Example 3.12 To discuss the impact of the choice of the function G, we simulated a data set with $n = 250$ observations from the following model

$$Y = X_1^2 - 1 + \frac{X_2}{2} + \varepsilon_j,$$

where $X_1, X_2 \sim \mathcal{N}(0, 1)$ and $\mathrm{Cov}(X_1, X_2) = 0.2$. The error variable ε is generated by a $\mathcal{N}(0, 0.5^2)$ distribution. In the first step, the conditional distribution function $F(y|x_1, x_2)$ is estimated by local linear techniques with the bandwidth $h_1 = 0.5$ and $h_2 = 1.5$. For the conditional quantile function estimate defined in (3.4), we tried two different choices for H. In the first scenario, we use a uniform distribution function; in the second scenario, the normal distribution is used with $\mu = \bar{Y}$ and $\sigma^2 = \mathrm{Var}(Y)$. For each scenario,

		Approaches	local linear		internalized		De Gooijer/Zerom
			$\hat{Q}_k(.5\|x_1)$		$\tilde{Q}_k(.5\|x_k)$		$\breve{Q}_k(.5\|x_k)$
ρ	n	Component k=1,2	G_{uni}	G_{norm}	G_{uni}	G_{norm}	
.2	100	$.75x_1$	0.1176	0.0905	0.2661	0.1865	0.1374
		$1.5\sin(.5\pi x_2)$	0.2112	0.2059	0.3543	0.3457	0.1818
	200	$.75x_1$	0.0630	0.0580	0.1971	0.1617	0.1066
		$1.5\sin(.5\pi x_2)$	0.0969	0.1545	0.1849	0.2637	0.1272
	400	$.75x_1$	0.0474	0.0378	0.1570	0.1013	0.0734
		$\sin(.5\pi x_2)$	0.1169	0.1191	0.2138	0.1868	0.0936
.8	100	$.75x_1$	0.1939	0.1867	0.4145	0.3902	0.1365
		$1.5\sin(.5\pi x_2)$	0.2801	0.2692	0.4611	0.4514	0.4865
	200	$.75x_1$	0.1882	0.1628	0.4385	0.3781	0.1272
		$1.5\sin(.5\pi x_2)$	0.2305	0.2151	0.3646	0.4108	0.4350
	400	$.75x_1$	0.1829	0.1646	0.4207	0.3735	0.0985
		$\sin(.5\pi x_2)$	0.2152	0.2045	0.3871	0.3803	0.4009

Table 3.1: *The mean absolute deviation error of the different approaches.*

we estimated each component for 450 replications and calculated the absolute deviation error for the observations restricted to the square $[-2,2]^2$. The results of this simulation are displayed in Table 3.2. In Figure 3.1, a typical picture of the additive conditional quantile estimates is displayed based on three simulation runs. Figure 3.2 shows the mean squared error of 100 simulation runs evaluated in 20 points between -0.5 and 0.5 and confirms the results from Table 3.2, which gives the average over the different evaluation points. The normal distribution as transformation function does a better job for the first component, whereas the uniform distribution is slightly better for the second component.

G	$Q_1(0.5\|\cdot)$	$Q_2(0.5\|\cdot)$
Normal	0.2416	0.0961
Uniform	0.2960	0.3107

Table 3.2: The average absolute deviation error of two different choices of G evaluated over 20 points between -0.5 and 0.5.

Example 3.13 As a demonstration of the applicability of the presented method to estimate additive conditional quantile function in higher dimension than $d = 2$, we consider the model

$$Y = \sum_{k=1}^{4} \sin(X_k) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,1), \mathbf{X} \sim \mathcal{N}(0,\Sigma) \tag{3.22}$$

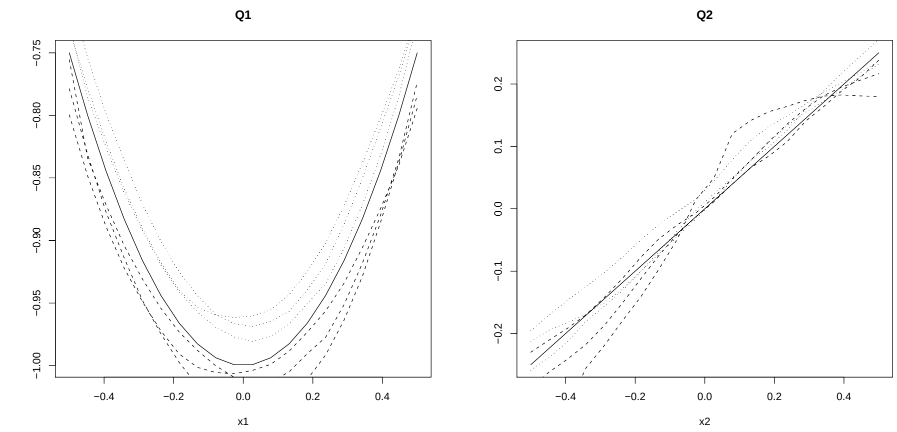

Figure 3.1: The additive conditional quantile function and their estimates. The additive component $Q_k(.5|x_k)$ (solid line), the estimates using the normal distribution (dotted line), and the estimates using the uniform distribution (dashed line).

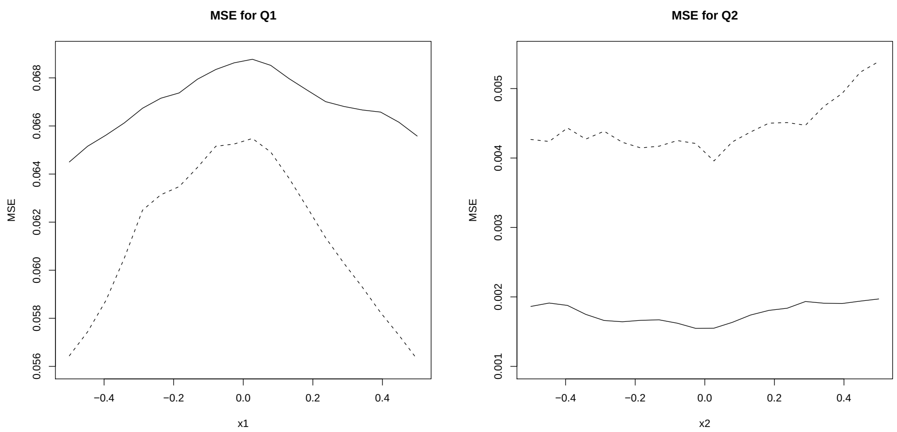

Figure 3.2: The MSE of the two different estimates for each additive component. The estimate using the normal distribution (dashed line), and the estimate using the uniform distribution (dotted line).

with two choices for Σ (for low and high correlation among the variables)

$$\Sigma_1 = \begin{pmatrix} 1.0 & 0.3 & 0.5 & 0.1 \\ 0.3 & 1.0 & 0.3 & 0.5 \\ 0.5 & 0.3 & 1.0 & 0.3 \\ 0.1 & 0.5 & 0.3 & 1.0 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \begin{pmatrix} 1.0 & 0.5 & 0.8 & 0.3 \\ 0.5 & 1.0 & 0.5 & 0.8 \\ 0.8 & 0.5 & 1.0 & 0.5 \\ 0.3 & 0.8 & 0.5 & 1.0 \end{pmatrix},$$

which was originally discussed by Hengartner and Sperlich (2005) in the context of traditional additive regression models. $n = 250$ observations are generated from this model for each of the 250 replications. Since the additive components and the marginal distributions are the same, we can average over all components at the same time. In the following table, the mean absolute deviation error and the mean squared error of the new estimate $\hat{Q}_k(.5|x_k)$ and the estimate $\breve{Q}_k(.5|x_k)$ proposed by De Gooijer and Zerom (2003) is recorded for the observations restricted to the square $[-2, 2]^4$. Note that the estimates

	AADE($\hat{Q}.(0.5\|x_k)$)	AADE($\breve{Q}.(0.5\|x_k)$)	MSE($\hat{Q}.(0.5\|x_k)$)	MSE($\breve{Q}.(0.5\|x_k)$)
Σ_1	0.08443	0.15315	0.01019	0.05624
Σ_2	0.08384	0.15989	0.01272	0.06094

Table 3.3: *AADE and MSE averaged over the four components* $\hat{Q}_1(0.5|x_k), \ldots, \hat{Q}_4(0.5|x_k)$ *and* $\breve{Q}_1(0.5|x_k), \ldots, \breve{Q}_4(0.5|x_k)$, *respectively, in the low and high correlation model (3.22).*

behave slightly better in the model with low correlation among the covariates. Furthermore, a comparison of the two estimates with respect to both criteria shows that the new estimate $\hat{Q}_k(.5|x_k)$ performs substantially better than the estimate $\breve{Q}_k(.5|x_k)$ suggested by De Gooijer and Zerom (2003).

3.6.2 Data examples

Example 3.14 To illustrate the performance of this method, we estimated the marginal effects in a real data example. The Boston housing data contains the housing values of suburbs of Boston and 13 variables/criteria, which might have an influence on the housing prices like pollution, crime, and urban amenities. This data set has been analyzed by several authors, also in the context of quantile regression [see e.g. De Gooijer and Zerom (2003)]. We focus on four covariates

- per capita crime rate (crime),
- average number of rooms per dwelling (rooms),
- weighted mean of distance to five Boston employment centers (distance),

- lower status of the population (econstatus),

and fit an additive conditional quantile model. We applied cross validation to determine the bandwidth for the four different variables. To simplify this problem, we set

$$h_1 = \hat{\sigma}_{\text{crime}} k, \;\; h_2 = \hat{\sigma}_{\text{rooms}} k, \;\; h_3 = \hat{\sigma}_{\text{distance}} k, \;\; h_4 = \hat{\sigma}_{\text{econstatus}} k,$$

minimized the cross validation criteria for $k \in [1/11, 30/11]$ for each marginal effect separately. Since the values are quite similar, we set $k = 1$ for all covariates. In Figure 3.3, we display five different curves of the marginal effects $q_k(\alpha|X_k)$ for fixed $\alpha = 0.05, 0.25, 0.5, 0.75, 0.95$. Note that the marginal effects are monotone in α.

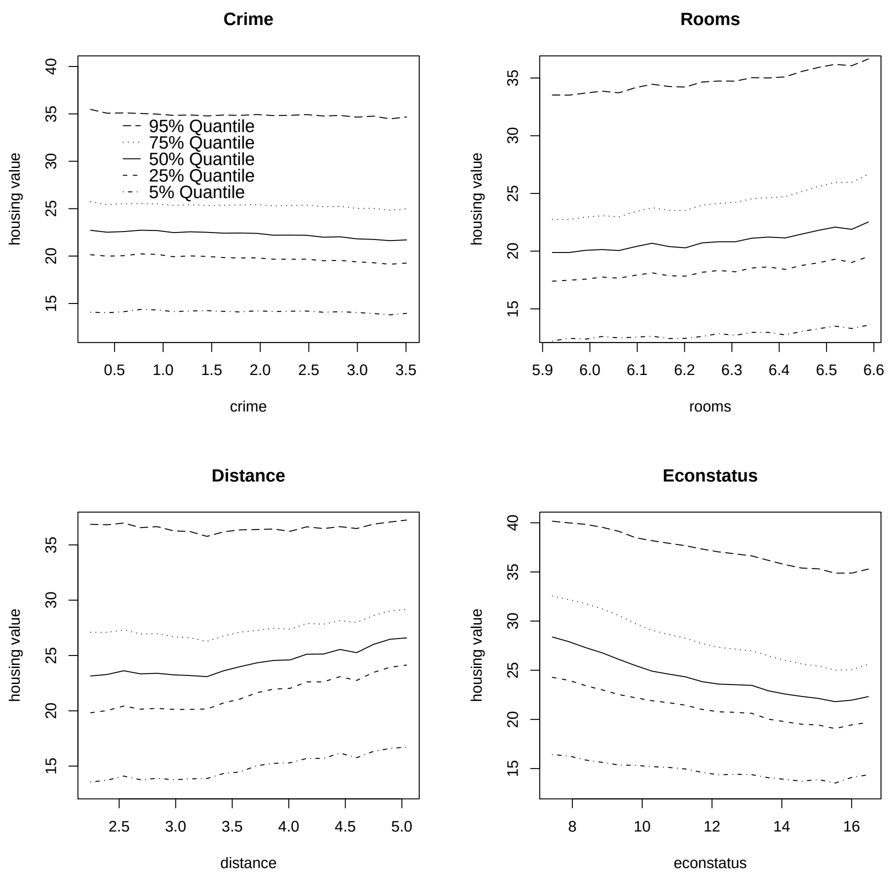

Figure 3.3: Boston housing data set: The marginal effects of four covariates at five different levels of α.

Example 3.15 The `baseball` data set in the package `corrgram` in `R` contains data from 322 major league baseball regular and substitute hitters in 1968. The data set covers 22 variables. For this example, we analyze the salary on the opening day 1987 conditioned on four covariates by fitting an additive quantile model. The covariate `years` collects the number of years in the major leagues, which measures the experience of the athlete. The covariate `runs` contains the number of runs in the career of the hitter until 1986. In a similar manner, the variables `hits` and `homeruns` describe the number of hits and home runs, respectively, in career. To determine the bandwidth of the local linear estimator for the conditional distribution function a cross validation is performed. As in the last example, we set

$$h_1 = \hat{\sigma}_{\text{years}} k, \;\; h_2 = \hat{\sigma}_{\text{runs}} k, \;\; h_3 = \hat{\sigma}_{\text{hits}} k, \;\; h_4 = \hat{\sigma}_{\text{homeruns}} k,$$

and minimize over $(0, 2)$. In this way, we obtain $k = 1.326733$. The following Figure describes the marginal effects for different levels of α for each of the four covariates.

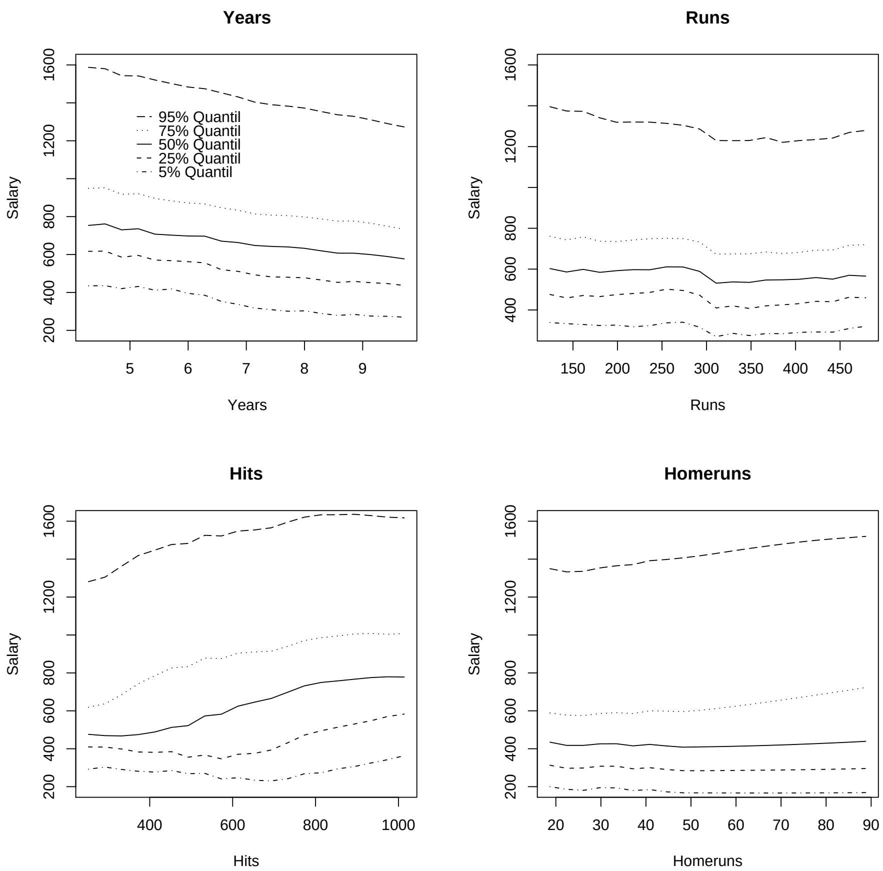

Figure 3.4: Baseball data set: The marginal effects of four covariates at five different levels of α.

List of Symbols

Scalars, Vectors, and Functions

X_j	jth univariate covariate	8
$\mathbf{X}_j = (X_{j1}, \dots, X_{jd})^T$	jth d-dimensional covariate	8
Y_j	j-th response variable	8
$X_{(j)}$	ordered statistic	4
$Y_{[j]}$	ordered statistic relative to $X_{(j)}$	4
n	sample size	8
ε_j	error variable	8
δ_{ij}	Kronecker delta	8
$m(\cdot)$	regression function	11, 29
$\sigma(\cdot)$	variance function	11, 29
$p(\cdot)$	density function of the random variable X_j	15
∇m	gradient of m	31
$\mathcal{H}(m)$	Hessian matrix of m	31
$x_{\underline{k}}$	notation for all variables but x_k	36
$m_k^{-1}(z\|x_{\underline{k}})$	inverse with respect to x_k	36
$\frac{\partial}{\partial x_k} m(x_k, x_{\underline{k}})$	derivative with respect to x_k	36
$Q(\alpha\|\mathbf{x})$	α-quantile of Y given $\mathbf{X} = \mathbf{x}$	77
$\rho_\alpha(u)$	check-function	78
$F(y\|\mathbf{X})$	conditional distribution function of Y given $\mathbf{X} = \mathbf{x}$	79
$\sigma(y\|\mathbf{X})$	conditional variance function	79
$G(\cdot)$	distribution function to transform data onto the compact interval $[0, 1]$	81
$g_k(x_k)$	additive component of the kth covariate in the additive regression model	82
$p_k(x_k)$	marginal density of X_k	82
$Q_k(\alpha\|x_k)$	additive component of the kth covariate in the additive quantile model	82
c, $c(\alpha)$	constant term in the additive regression model and quantile model, respectively	82, 82
$p_{\underline{k}}(x_{\underline{k}})$	marginal density of $X_{\underline{k}}$	85
$\gamma_W(x_k)$, $\gamma_W(\alpha\|x_k)$	contrast with respect to the measure W	85, 86
$q_k(\alpha\|x_k)$	marginal effect of the kth variable	90
$D^{\nu_{\underline{k}}}$	the differential operator multi-indexed by $\nu_{\underline{k}}$	97

Estimates

$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^T$	estimated coefficient for the polynomial regression and the local polynomial regression, respectively	9, 14
$\hat{m}_{NW}(\cdot)$	Nadaraya-Watson estimator	14, 30
$\hat{m}_{LL}(\cdot)$	local linear estimator	14, 30
$\hat{m}_E$, $\hat{m}_I$	external and internal regression estimator	17, 17
$\hat{p}$	kernel density estimator	17
$\hat{m}_{INW}$	internalized Nadaraya-Watson estimator	17
g_I^{-1} and g_A^{-1}	isotonic and antitonic distribution function	22, 23
g, $g_I^{-1}(\cdot, h_m)$, $g_I(\cdot, h_m)$	function g, isotonized inverse and isotonized g	26
$\hat{m}_{I_1}$, $\hat{m}_{I_{1,2}}$, $\hat{m}_{I_{1,\ldots,d}}$	isotonized estimate with respect to the first, the first and the second, and all d variables, respectively	36, 36, 36
$\hat{F}_{NW}(y\|\mathbf{x})$	Nadaraya-Watson estimate of the conditional distribution function	79
$\hat{F}_{RW}(y\|\mathbf{x})$	reweighted Nadaraya-Watson estimate of the conditional distribution function	80
$\hat{F}_{INW}(y\|\mathbf{x})$	internalized Nadaraya-Watson estimate of the conditional distribution function	80
$\hat{G}_I(\alpha\|\mathbf{x})$, $\hat{Q}_I(\alpha\|\mathbf{x})$	isotonized estimate of the quantile function	81
$G_N(\alpha\|\mathbf{x})$, $Q_N(\alpha\|\mathbf{x})$	approximation of the conditional quantile	82
$\hat{p}$, $\hat{p}_k$	empirical pdf of $\mathbf{X}$ and X_k, respectively	84
$\hat{\gamma}_W(\alpha\|x_k)$	estimate of the contrast $\gamma_W(\alpha\|x_k)$	86
$\gamma_{W,N}(\alpha\|x_k)$	approximation of the contrast $\gamma(\alpha\|x_k)$	87
$\hat{q}_k(\alpha\|x_k)$	estimated kth marginal effect	90
$\hat{Q}_k(\alpha\|\mathbf{x})$	estimate of the kth additive component	91
$\hat{Q}_{\text{add}}(\alpha\|\mathbf{x})$	estimated additive quantile function	91
$\tilde{q}_k(\alpha\|x_k)$	internalized estimator of the kth marginal effect	92
$\tilde{Q}_k(\alpha\|x_k)$	internalized estimate of the kth additive quantile component	123
$\breve{q}_k(\alpha\|x_k)$	kth marginal effect estimate of De Gooijer and Zerom (2003)	123
$\breve{Q}_k(\alpha\|x_k)$	estimate of the kth additive component by De Gooijer and Zerom (2003)	123

Convergence

$\xrightarrow{\mathcal{D}}$	convergence in distribution	15
$O(\cdot)$	Landau symbol *Big O*, i.e. $a = O(b)$ iff $a/b \to$ constant as $n \to \infty$ or $h \to 0$	16
$o(\cdot)$	Landau symbol *Small O*, i.e. $a = o(b)$ iff $a/b \to 0$ as $n \to \infty$ or $h \to 0$	19
$o_p(\cdot)$	stochastic order symbol, i.e. $X = o_p(Y)$ iff for all $\varepsilon > 0$ holds $P(\lvert X/Y \rvert > \varepsilon) \to 0$	19
$O_p(\cdot)$	stochastic order symbol, i.e. $X = O_p(Y)$ iff for all $\varepsilon > 0$ exists $c > 0$ such that $P(\lvert X/Y \rvert > c) < \varepsilon$ as n is sufficiently large or h is sufficiently small	47

Kernels and Bandwidths

K	kernel function	11, 13, 30, 80
$\lvert\lvert K \rvert\rvert_2$	L_2-norm of the kernel function	13
K_h	scaled kernel function, i.e. $K_h(\cdot) = K(\cdot/h)/h$	13
h	bandwidth	13
h^*	optimal bandwidth in terms of AMSE	16
$\kappa_s(K)$	constant depending on the sth moment of the kernel K	15
K_m and h_m	kernel and bandwidth of the monotonizing operator	21
I_d	$d \times d$ identity matrix	30
H	bandwidth matrix	30
K_H	scaled multivariate kernel function, i.e. $K_H(\cdot) = \frac{1}{\det(H)} K(H^{-1} \cdot)$	30
$h_1, \ldots, h_d$	bandwidths for the corresponding variables $x_1, \ldots, x_d$	40

Bibliography

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.

Andrews, D. W. K. and Whang, Y.-J. (1990). Additive interactive regression models: circumvention of the curse of dimensionality. *Econometric Theory*, 6(4):466–479.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.*, 26:641–647.

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

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression*. John Wiley & Sons, London-New York-Sydney. Wiley Series in Probability and Mathematical Statistics.

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

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

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, 80(391):580–619. With discussion and with a reply by the authors.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.*, 17(2):453–555.

Chen, R., Härdle, W., Linton, O. B., and Severance-Lossin, E. (1996). Nonparametric estimation of additive separable regression models. In *Statistical theory and computational aspects of smoothing (Semmering, 1994)*, Contrib. Statist., pages 247–265. Physica, Heidelberg.

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2007). Quantile and probability curves without crossing. http://www.citebase.org/abstract?id=oai:arXiv.org:0704.3649.

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.

De Gooijer, J. G. and Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates. *J. Amer. Statist. Assoc.*, 98(461):135–146.

Dette, H., Neumeyer, N., and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3):469–490.

Dette, H. and Volgushev, S. (2007). Non-crossing nonparametric estimates of quantile curves. http://www.ruhr-uni-bochum.de/mathematik3. Preprint, Ruhr-Universität Bochum.

Doksum, K. and Koo, J.-Y. (2000). On spline estimators and prediction intervals in nonparametric regression. *Comput. Statist. Data Anal.*, 35(1):67–82.

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

Fan, J. (2000). Prospects of nonparametric modeling. *J. Amer. Statist. Assoc.*, 95(452):1296–1300.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Gallant, A. R. and Golub, G. H. (1984). Imposing curvature restrictions on flexible functional forms. *Journal of Econometrics*, 26(3):295–321. available at http://ideas.repec.org/a/eee/econom/v26y1984i3p295-321.html.

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

Hall, P., Huang, L.-S., Gifford, J. A., and Gijbels, I. (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *J. Comput. Graph. Statist.*, 10(3):592–614.

Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.*, 94(445):154–163.

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.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.

He, X. (1997). Quantile curves without crossing. *Amer. Statist.*, 51(2):186–192.

Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *J. Multivariate Anal.*, 95(2):246–272.

Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.*, 100(472):1238–1249.

Jones, M. C., Davies, S. J., and Park, B. U. (1994). Versions of kernel-type regression estimators. *J. Amer. Statist. Assoc.*, 89(427):825–832.

Kim, W., Linton, O. B., and Hengartner, N. W. (1999). A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. *J. Comput. Graph. Statist.*, 8(2):278–297.

Koenker, R. (2005). *Quantile regression*, volume 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Koenker, R. and Bassett, Jr., G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

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.

Li, Q. (2000). Efficient estimation of additive partially linear models. *Internat. Econom. Rev.*, 41(4):1073–1092.

Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82(1):93–100.

Mack, Y. P. and Müller, H.-G. (1989). Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā Ser. A*, 51(1):59–72.

Mammen, E., Linton, O., and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.*, 27(5):1443–1490.

Masry, E. (1996). Multivariate regression estimation—local polynomial fitting for time series. *Stochastic Process. Appl.*, 65(1):81–101.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.*, 24(2):165–179.

Matzkin, R. L. (1986). Restrictions of economic theory in nonparametric methods. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 42, pages 2523–2558. Elsevier. available at http://ideas.repec.org/h/eee/ecochp/4-42.html.

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

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9(1):141–142.

Nielsen, J. P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(1):43–61.

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.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization*, volume 49 of *Nonconvex Optim. Appl.*, pages 272–281. Kluwer Acad. Publ., Dordrecht.

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

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.*, 42:43–47.

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

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

Scheder, R. (2007). *monoProc: strictly monotone smoothing procedure*. R package version 1.0-5.

Scheffé, H. (1999). *The analysis of variance*. Wiley Classics Library. John Wiley & Sons Inc., New York. Reprint of the 1959 original, A Wiley Publication in Mathematical Statistics.

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

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, 13(2):689–705.

Tjøstheim, D. and Auestad, B. H. (1994). Nonparametric identification of nonlinear time series: projections. *J. Amer. Statist. Assoc.*, 89(428):1398–1409.

Umantsev, L. and Chernozhukov, V. (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, 26(1):271–292. available at http://ideas.repec.org/a/spr/empeco/v26y2001i1p271-292.html.

van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Nederl. Akad. Wetensch. Proc. Ser. A.* **59** = *Indag. Math.*, 18:444–455.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372.

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

Yu, K. and Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics & Data Analysis*, 25(2):159–166. available at http://ideas.repec.org/a/eee/csdana/v25y1997i2p159-166.html.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.*, 93(441):228–237.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *The Statistician*, 52(3):331–350.

Zygmund, A. (2002). *Trigonometric series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition. With a foreword by Robert A. Fefferman.

Acknowledgments

First of all, it is a pleasure to me to thank my supervisor Prof. Dr. Holger Dette for great support and the friendly and fruitful atmosphere in his research group. Special thanks to my long-term officemate Dr. Eva Nagel who moved with me from office to office, and to my companion from the first day at the Fakultät für Mathematik in Bochum Nagat Karroum.
I wish to thank all the people for stimulating discussions, I was able to meet while I worked on this thesis. In particular, I like to thank Prof. Dr. Wolfgang Polonik, who gave me the opportunity to visit the University of California, Davis, where I learned a lot about statistics and met amazing people. Warm thanks to Sheila Eghbali, Elcin Tan, and Arthur and Virginia Gressel who made me feel at home in Davis, California.
Furthermore, I would like to acknowledge the financial support of the DAAD, the German Academic Exchange Service, during the stay at the Department of Statistics at the University of California, Davis. The author is grateful to Merja Maier and Angela Lohberg for reading parts of this thesis beyond the mathematics. Warm thanks to my friends in Bochum, Stuttgart, and the rest of the world. My life would be incomplete without them.
Last but not least, I would like to thank my parents and my sister for supporting me in my decisions and encouraging me over all the years.

Lebenslauf

Persönliche Daten

Name	Regine Scheder
Geburtsdatum	29. Februar 1980
Geburtsort	Bietigheim-Bissingen

Ausbildung

09/86 - 07/90	Schillerschule Bietigheim-Bissingen
09/90 - 06/99	Ellental Gymnasien Bietigheim-Bissingen
06/99	Abitur
10/99 - 03/04	Diplomstudiengang Mathematik mit Nebenfach Wirtschaftswissenschaften, Ruhr-Universität Bochum
09/01 - 03/04	Stipendiatin der Studienstiftung des deutschen Volkes
09/01	Vordiplom in Mathematik
03/04	Diplom in Mathematik *Thema der Diplomarbeit:* Minimax Schätzung bei standardisiertem Verlust

Beschäftigung

04/04 - 08/04	Wissenschaftliche Mitarbeiterin an der Fakultät für Mathematik, Ruhr-Universität Bochum
09/04 - 08/05	Forschungsaufenthalt an der University of California, Davis, gefördert durch ein Doktorandenstipendium des DAAD
seit 09/05	Wissenschaftliche Mitarbeiterin an der Fakultät für Mathematik, Ruhr-Universität Bochum