Technische Universität München

Department of Mathematics

Kernel Estimation of Conditional Copula Densities

Master's Thesis

by

Alexej Brauer

Supervisor:	Prof. Claudia Czado, Ph.D.
Advisor:	Thomas Nagler
Advisor:	Prof. Claudia Czado, Ph.D.
Submission date:	08 December 2016

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

City, Day Month Year

Dedication

This thesis is dedicated to my parents. They sacrificed their prosperity so that I could pursue an excellent education.

Acknowledgments

I would first like to thank Prof. Claudia Czado for giving me the opportunity to work on this topic and especially for her excellent supervision. In addition, I am truly grateful to my advisor Thomas Nagler for his exceptional guidance. His door was always open whenever I had a question about my research or writing. In joyful discussions they both consistently allowed this thesis to be my own work, yet steered me in the right direction wherever they thought I needed it. The writing of this thesis further encouraged me in my passion for mathematical statistics, for which I am deeply grateful.

Finally and most importantly, I want to thank my parents, siblings, friends and girlfriend. Your friendship, love, and unyielding support during my life helped me to become the person I am today.

Abstract

In many cases, the dependence structure between random variables varies according to the values of measured covariates. A natural approach studying these type of models is the usage of conditional copulas. Here we provide an introduction to this concept and propose a novel fully nonparametric method for the estimation of conditional copula densities. Our procedure is based on transformation kernels combined with local linear regression. The asymptotic properties are studied and a bandwidth selection procedure is suggested. We analyze the performance of this estimator in a simulation study containing a variety of scenarios, using univariate as well as multivariate covariates. In all scenarios, the method proved to work well and improves upon an estimator that does not take the covariates into account.

Zusammenfassung

In vielen Fällen variiert die Stärke der Abhängigkeit zwischen Zufallsvariablen anhand von Bedingungsgrößen. Ein mittlerweile natürlicher Ansatz, der diese Art von Modellen untersucht, ist der, der bedingten Copulas. Zunächst geben wir eine Einführung in das Konzept der bedingten Copulas. Wir stellen einen neuartigen, gänzlich nichtparametrischen Ansatz zur Schätzung solcher bedingten Copuladichten vor. Dieses Verfahren basiert auf Transformationskernels kombiniert mit lokal linearer Regression. Des Weiteren werden die asymptotischen Eigenschaften untersucht und ein Auswahlverfahren für die Bandbreite vorgeschlagen. Schließlich analysieren wir die Performance dieses Schätzers in einer Simulationsstudie, die eine Vielzahl von Szenarien mit ein- und mehrdimensionalen Bedingungen enthält. Hierbei überzeugte die Methode in allen Szenarien und stellte eine Verbesserung zu einem Schätzer dar, welcher die Bedingungsgrößen nicht in Betracht zieht.

Contents

1. Introduction

Accurate modeling of dependence structures was and will always be a mainstay in multivariate statistics. Copulas have become a standard tool for this task. Sklar (1959) provided the theoretical foundation to separate the problem of modeling the dependence (via copulas) from the problem of modeling the marginal distributions. In the last two decades, copulas have been extensively studied in the statistical literature. Among others, we find the usage of copulas in the fields of risk management, insurance, econometrics, and hydrology.

A more recent concept are conditional copulas, where the influence of covariates on the dependence structure is studied. Since the seminal work of Patton (2006), a variety of estimation methods for these conditional copulas have been proposed.

Gijbels et al. (2011) presented two intuitive nonparametric estimators for conditional copulas. The asymptotic properties of these estimators were studied in Veraverbeke et al. (2011), the cases of multivariate and functional covariates are considered in Gijbels et al. (2012). However, the majority of the literature on conditional copula estimation is of semiparametric nature. One considers a parametric family of copulas and assumes that the dependence of the copula on the covariate is fully determined by the covariates influence on the copula parameter. Acar et al. (2011) proposed a semiparametric approach where the conditional copula parameter estimation is based on local likelihood and the selection of an appropriate copula family uses cross-validated prediction errors. Building upon this work, Lin et al. (2016) provided an extension to the multivariate parameter case. Here the authors assume known or parametrically specified marginals. To account for this Abegaz et al. (2012) suggested estimating the conditional copula by using a semiparametric method in which the marginal distributions are assumed to be unknown. Vatter and Chavez-Demoulin (2015) modeled conditional copula parameters via a generalized additive model and proposed a maximum penalized log-likelihood estimator. Further, the case where only the marginal distributions are affected by the covariates is considered in Gijbels et al. (2015).

Since parametric as well as semiparametric estimators rely on the correct specification of the model, a fully nonparametric procedure would be an appealing alternative. To the best of this author's knowledge, such a method was never considered in the context of conditional copula densities. The main contribution of this thesis is to study a fully nonparametric procedure. It is based on transformation kernels and local linear regression.

The remainder of this thesis is organized as follows. Chapter 2 recalls briefly the theoretical background. The concept of conditional and partial copulas will be introduced in Chapter 3. Chapter 4 concentrates on the kernel estimation of conditional copula densities. Here we discuss at first the estimation based on transformation. Section 4.2 will discuss a general estimation method for multivariate conditional densities, based on local linear regression. Chapter 5 deals with the asymptotic behavior of this estimator. Section 5.1 contains the main theorem of this thesis. It states that under regularity conditions the conditional density estimator is asymptotically normally distributed. We prove this theorem in Section 5.2. In Chapter 6 we propose a method for choosing the smoothing parameters. The performance of the estimator is analyzed by a simulation study in Chapter 7. Chapter 8 concludes. Detailed proofs are given in the Appendix (Appendix A).

2. Theoretical background

In this chapter we will mention some basic definitions and theorems from probability theory. The chapter will then introduce briefly the copula theory and discuss the weighted least squares approach as a method for the estimation of parameter in a linear model. We continue with some basic notations concerning kernels and giving an introduction in the local linear regression approach. We end this chapter by introducing some error measures that will be used throughout this thesis.

Remark 2.1. *Throughout this thesis we make the assumption that all mentioned random quantities are absolutely continuous, therefore one has in all situations the respective densities available.*

2.1. Basic theorems

Since we will use the Taylor theorem so often in this thesis, we will state a version suitable for our specific setting (see, Klemelä, 2009, p. 503).

Theorem 2.1. *(Multivariate Taylor theorem)*
Let the function $g(\boldsymbol{z}) : \mathbb{R}^d \to \mathbb{R}$ have bounded continuous first, second and third order derivatives with respect to $\boldsymbol{z}$. Furthermore let $\bar{\boldsymbol{z}}, \boldsymbol{z} \in \mathbb{R}^d$ and M be a symmetric, positive definite matrix $M \in \mathbb{R}^{d\times d}$ with $M = o(1)$. By Taylor expansion it holds that

$$g(M\bar{\boldsymbol{z}} + \boldsymbol{z}) = g(\boldsymbol{z}) + R_0, \tag{2.1}$$

$$g(M\bar{\boldsymbol{z}} + \boldsymbol{z}) = g(\boldsymbol{z}) + (M\bar{\boldsymbol{z}})^\top \nabla_{\boldsymbol{z}} g(\boldsymbol{z}) + R_1, \tag{2.2}$$

$$g(M\bar{\boldsymbol{z}} + \boldsymbol{z}) = g(\boldsymbol{z}) + (M\bar{\boldsymbol{z}})^\top \nabla_{\boldsymbol{z}} g(\boldsymbol{z}) + \frac{1}{2}(M\bar{\boldsymbol{z}})^\top \mathcal{H}_{\boldsymbol{z}}(M\bar{\boldsymbol{z}}) + R_2, \tag{2.3}$$

where R_0, R_1 and R_2 are the remainder terms, $\nabla_{\boldsymbol{z}} g(\boldsymbol{z}) := \left(\frac{\partial}{\partial z_1} g(\boldsymbol{z}), \ \dots \ , \frac{\partial}{\partial z_d} g(\boldsymbol{z})\right)^\top$ are the partial derivatives and $\mathcal{H}_{\boldsymbol{z}} \in \mathbb{R}^{d\times d}$ is the Hessian of $g(\boldsymbol{z})$ with respect to $\boldsymbol{z}$. For the remainder terms it holds that

$$R_0 = \sum_{\substack{(k,r)\\ \in\{1,\dots,d\}^2}} O\Big([M]_{k,r}\bar{z}_r\Big), \tag{2.4}$$

$$R_1 = \sum_{\substack{(k_1,k_2,r_1,r_2)\\ \in\{1,\dots,d\}^4}} O\Big([M]_{k_1,r_1}[M]_{k_2,r_2}\bar{z}_{r_1}\bar{z}_{r_2}\Big), \tag{2.5}$$

$$R_2 = \sum_{\substack{(k_1,k_2,k_3,r_1,r_2,r_3)\\ \in\{1,\dots,d\}^6}} O\Big([M]_{k_1,r_1}[M]_{k_2,r_2}[M]_{k_2,r_3}\bar{z}_{r_1}\bar{z}_{r_2}\bar{z}_{r_3}\Big), \tag{2.6}$$

where $[M]_{k,l}$ is the element in the k'th row and l'th column of M and $\bar{z}_l$ is the l'th element of the vector $\bar{\boldsymbol{z}}$.

To see why the remainders can be expressed in such a way, we will derive the expression (2.6) in detail. One can express the remainder in the mean value form by

$$R_2 = \frac{1}{6} \sum_{\substack{(k_1,k_2,k_3)\\ \in\{1,\dots,d\}^3}} \frac{\partial^3}{\partial z_{k_1}\partial z_{k_2}\partial z_{k_3}} g(\mathbf{z} + c\cdot M\bar{\mathbf{z}})[M\bar{\mathbf{z}}]_{k_1}[M\bar{\mathbf{z}}]_{k_2}[M\bar{\mathbf{z}}]_{k_3},$$

where $c \in (0,1)$ and $[M\bar{\mathbf{z}}]_j$ is the j'th component of the vector $M\bar{\mathbf{z}}$. By using that $g(\mathbf{z})$ has bounded continuous third order derivatives with respect to $\mathbf{z}$ we

see that

$$
\begin{aligned}
R_2 &= O\Big(\sum_{\substack{(k_1,k_2,k_3)\\ \in\{1,\dots,d\}^3}} [M\bar{\mathbf{z}}]_{k_1}[M\bar{\mathbf{z}}]_{k_2}[M\bar{\mathbf{z}}]_{k_3}\Big)\\
&= O\Big(\sum_{\substack{(k_1,k_2,k_3)\\ \in\{1,\dots,d\}^3}} \Big(\sum_{r_1=1}^{d}[M]_{k_1,r_1}\bar{z}_{r_1}\Big)\Big(\sum_{r_2=1}^{d}[M]_{k_2,r_2}\bar{z}_{r_2}\Big)\Big(\sum_{r_3=1}^{d}[M]_{k_3,r_3}\bar{z}_{r_3}\Big)\Big)\\
&= O\Big(\sum_{\substack{(k_1,k_2,k_3)\\ \in\{1,\dots,d\}^3}} \sum_{\substack{(r_1,r_2,r_3)\\ \in\{1,\dots,d\}^3}} [M]_{k_1,r_1}[M]_{k_2,r_2}[M]_{k_2,r_3}\bar{z}_{r_1}\bar{z}_{r_2}\bar{z}_{r_3}\Big)\\
&= \sum_{\substack{(k_1,k_2,k_3,r_1,r_2,r_3)\\ \in\{1,\dots,d\}^6}} O\Big([M]_{k_1,r_1}[M]_{k_2,r_2}[M]_{k_2,r_3}\bar{z}_{r_1}\bar{z}_{r_2}\bar{z}_{r_3}\Big).
\end{aligned}
$$

The derivation of the other remainders is similar. We also state Taylor's theorem for the special case $M = b \cdot I$, where $b \in \mathbb{R}$ and I is the identity matrix in $\mathbb{R}^{d\times d}$.

Corollary 2.1. *Let the function $g(\boldsymbol{z}) : \mathbb{R}^d \to \mathbb{R}$ has bounded continuous first, second and third order derivatives with respect to $\boldsymbol{z}$. Furthermore let $\bar{\boldsymbol{z}}, \boldsymbol{z} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ converge to zero. By Taylor expansion it holds that*

$$
\begin{aligned}
g(b\bar{\boldsymbol{z}} + \boldsymbol{z}) &= g(\boldsymbol{z}) + R_0,\\
g(b\bar{\boldsymbol{z}} + \boldsymbol{z}) &= g(\boldsymbol{z}) + b\bar{\boldsymbol{z}}^\top \nabla_{\boldsymbol{z}} g(\boldsymbol{z}) + R_1,\\
g(b\bar{\boldsymbol{z}} + \boldsymbol{z}) &= g(\boldsymbol{z}) + b\bar{\boldsymbol{z}}^\top \nabla_{\boldsymbol{z}} g(\boldsymbol{z}) + \frac{b^2}{2}\bar{\boldsymbol{z}}^\top \mathcal{H}_{\boldsymbol{z}}\bar{\boldsymbol{z}} + R_2,
\end{aligned}
$$

where R_0, R_1 and R_2 are the remainder, $\nabla_{\boldsymbol{z}} g(\boldsymbol{z}) := \big(\frac{\partial}{\partial z_1} g(\boldsymbol{z}), \quad \dots \quad , \frac{\partial}{\partial z_d} g(\boldsymbol{z})\big)^\top$ are the partial derivatives and $\mathcal{H}_{\boldsymbol{z}} \in \mathbb{R}^{d\times d}$ is the Hessian of $g(\boldsymbol{z})$ with respect to $\boldsymbol{z}$. For the remainder it holds that

$$
\begin{aligned}
R_0 &= \sum_{r\in\{1,\dots,d\}} O\Big(b\bar{z}_r\Big),\\
R_1 &= \sum_{\substack{(r_1,r_2)\\ \in\{1,\dots,d\}^2}} O\Big(b^2\bar{z}_{r_1}\bar{z}_{r_2}\Big),\\
R_2 &= \sum_{\substack{(r_1,r_2,r_3)\\ \in\{1,\dots,d\}^3}} O\Big(b^3\bar{z}_{r_1}\bar{z}_{r_2}\bar{z}_{r_3}\Big),
\end{aligned}
$$

where $\bar{z}_l$ is the l'th element of the vector $\bar{\boldsymbol{z}}$.

Motivated by finding the density of transformed random variables, the following transformation theorem will be useful.

Theorem 2.2 (Transformation theorem)**.** *Let* $\boldsymbol{Y} := (Y_1, \dots, Y_d)^\top$ *be a random vector in* $\mathbb{R}^d$ *and*

$$Z_j := T_j(Y_1, \dots, Y_d), \text{ for } j \in \{1, \dots, d\},$$

be some random variables resulting from applying functions $T_j : \mathbb{R}^d \to \mathbb{R}$ *with* $j \in \{1, \dots, d\}$ *on the random vector* $\boldsymbol{Y}$*. Assume that the inverse of the functions* T_j *exist such that*

$$Y_j = T_j^{-1}(Z_1, \dots, Z_d), \text{ for } j \in \{1, \dots, d\},$$

and are continuous with continuous partial derivatives.
Let T *and* T^{-1} *denote the following functions:*

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad s.t. \quad \boldsymbol{y} := (y_1, \dots, y_d)^\top \mapsto (T_1(\boldsymbol{y}), T_2(\boldsymbol{y}), \dots, T_d(\boldsymbol{y}))^\top,$$

$$T^{-1} : \mathbb{R}^d \to \mathbb{R}^d, \quad s.t. \quad \boldsymbol{z} := (z_1, \dots, z_d)^\top \mapsto \left(T_1^{-1}(\boldsymbol{z}), T_2^{-1}(\boldsymbol{z}), \dots, T_d^{-1}(\boldsymbol{z})\right)^\top.$$

Use $dT(\boldsymbol{y})$ *as the notation for the Jacobian matrix of* T*, and* $dT^{-1}(\boldsymbol{z})$ *as the notation for the Jacobian matrix of* T^{-1}*:*

$$dT(\boldsymbol{y}) = \begin{pmatrix} \frac{\partial T_1}{\partial y_1}(\boldsymbol{y}) & \dots & \frac{\partial T_1}{\partial y_d}(\boldsymbol{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial T_d}{\partial y_1}(\boldsymbol{y}) & \dots & \frac{\partial T_d}{\partial y_d}(\boldsymbol{y}) \end{pmatrix}, \quad dT^{-1}(\boldsymbol{z}) = \begin{pmatrix} \frac{\partial T_1^{-1}}{\partial z_1}(\boldsymbol{z}) & \dots & \frac{\partial T_1^{-1}}{\partial z_d}(\boldsymbol{z}) \\ \vdots & \ddots & \vdots \\ \frac{\partial T_d^{-1}}{\partial z_1}(\boldsymbol{z}) & \dots & \frac{\partial T_d^{-1}}{\partial z_d}(\boldsymbol{z}) \end{pmatrix}.$$

Then the joint probability density function of $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ *is*

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{\boldsymbol{Z}}(z_1, \dots, z_d) = |\det(dT^{-1}(\boldsymbol{z}))| f_{\boldsymbol{Y}}(T^{-1}(\boldsymbol{z})),$$

and the joint probability density function of $\boldsymbol{Y}$ *can be written as*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{Y}}(y_1, \dots, y_d) = |\det(dT(\boldsymbol{y}))| f_{\boldsymbol{Z}}(T(\boldsymbol{y})).$$

Concerning the asymptotics of our estimators, an very important result will be the following Liapunov central limit theorem (see, Li and Racine, 2006, Appendix).

Theorem 2.3. *(Liapunov Central Limit Theorem) Let* $\{T_{n,i}\}$ *be a sequence of independent (double array) random variable such that* $\mathrm{E}(T_{n,i}) = \mu_{n,i}$ *and* $\mathrm{Var}(T_{n,i}) = \sigma_{n,i}^2$*, with* $\mathrm{E}[|T_{n,i}|^{2+\delta}] < \infty$ *for some* $\delta > 0$*. Let* $S_n = \sum_{i=1}^n T_{n,i}$*, and let* $\sigma_n^2 = \mathrm{Var}(S_n) = \sum_{i=1}^n \sigma_{n,i}^2$*.*

If $\sigma_n^2 = \sigma^2 + o(1)$ *(*σ^2 *is a constant), and*

$$\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}[|T_{n,i} - \mu_{n,i}|^{2+\delta}] = 0 \text{ for some } \delta > 0, \tag{2.7}$$

$$\text{then } \sigma_n^{-1}(S_n - \mathrm{E}(S_n)) = \sigma_n^{-1} \sum_{i=1}^n [T_{n,i} - \mathrm{E}(T_{n,i})] \xrightarrow{d} N(0,1).$$

2.2. Copulas

In this section we will briefly introduce copulas and state an important theorem.

Definition 2.1 (Copula function)**.** *A function $C : [0,1]^d \to [0,1]$ is called a d-dimensional* ***copula*** *if there exists a random vector $(U_1, \dots, U_d)$ with $U_i \sim U[0,1]$, $i \in \{1, \dots, d\}$, such that*

$$P(U_1 \leq u_1, \dots, U_d \leq u_d) = C(u_1, \dots, u_d),$$

i.e., C is the cumulative distribution function of $(U_1, \dots, U_d)$.

Theorem 2.4 (Sklar's Theorem)**.** *Let F be a continuous d-dimensional distribution function with marginal distributions $F_1, \dots, F_d$. Then there exits a unique d-dimensional copula C such that for all $(x_1, \dots, x_d) \in \mathbb{R}^d$ is holds*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \tag{2.8}$$

Conversely, if C is a d-dimensional copula and $F_1, \dots, F_d$ are univariate distribution functions, F as defined in (2.8) is a d-dimensional distribution function.

Definition 2.2 (Copula density)**.** *A copula density $c(u_1, \dots, u_d)$ is defined as*

$$c(u_1, \dots, u_d) := \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d},$$

where $C(u_1, \dots, u_d)$ is a differentiable copula function.

Remark 2.2. *We can now construct a copula density by using the inversion of equation (2.8) in Sklar's theorem:*

$$c(u_1, \dots, u_d) = \frac{f\left(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\right)}{\prod_{i=1}^d f_i(F_i^{-1}(u_i))}, \tag{2.9}$$

where F is a d-dimensional cdf, the F_i's, $i \in \{1, \dots, d\}$ are the marginal cdf's and $f, f_1, \dots, f_d$ are the densities corresponding to $F, F_1, \dots, F_d$.

2.3. Weighted least squares approach

Suppose that we have the model

$$Y = \mathbf{x}^\top \boldsymbol{\beta} + \epsilon,$$

for a response random variable Y with a random error variable ϵ, a covariate vector $\mathbf{x}$ and a parameter vector $\boldsymbol{\beta}$ in $\mathbb{R}^{\tilde{p}}$. The regression parameters $\boldsymbol{\beta}$ are unknown and

have to be estimated by using n observations $(y_i, \mathbf{x}_i) := (y_i, x_{i,1}, \ldots, x_{i,\tilde{p}})$ for $i \in \{1, \ldots, n\}$. We can rewrite this model with the following matrix notation:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,\tilde{p}} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,\tilde{p}} \\ \vdots & & & \vdots \\ x_{n,1} & x_{n,2} & \ldots & x_{n,\tilde{p}} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

The weighted least squares approach estimates the parameters $\boldsymbol{\beta}$ by minimizing the weighted residual sum of squares

$$\sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \beta)^2 w_i. \tag{2.10}$$

The standard least squares approach (minimizing the standard residual sum of squares $\sum_{i=1}^{n} (y_i - \mathbf{x}_i \beta)^2$) is just a special case of the weighted least squares approach, where the weights are $w_i = 1$ for $i \in \{1, \ldots, n\}$.

There are several situations, where it is appropriate to use the weighted least squares instead of the standard least squares. For example if we assume that some data points are of poor quality we can use the weights to regulate the influence of those points on our estimation. Low-quality data points would than have lower weights than a good-quality data point. In general, the weights are determining the influence of the data points on the estimation.

Define the matrices $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{W}$ as follows:

$$\mathcal{X} := \begin{pmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,\tilde{p}} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,\tilde{p}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \ldots & x_{n,\tilde{p}} \end{pmatrix}, \quad \mathcal{Y} := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathcal{W} := \begin{pmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & w_n \end{pmatrix},$$

where $\mathcal{X}$ is called the design matrix, the $\mathcal{Y}$ is the response vector and $\mathcal{W}$ is the weight matrix. An important result is, that under the condition that the matrix $(\mathcal{X}^\top \mathcal{W} \mathcal{X})$ is invertible, the result of the of minimization problem (2.10) is:

$$\hat{\boldsymbol{\beta}} = (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y}.$$

2.4. Kernels

The purpose of this section is to introduce some basic notions for kernels that we will use throughout this thesis. In general, a kernel $K(\cdot)$ is just a probability density function.

A univariate kernel $K_h(z)$ with bandwidth h is defined for $z \in \mathbb{R}$ as

$$K_h(z) := \frac{1}{h} K\left(\frac{z}{h}\right), \text{ for } h > 0.$$

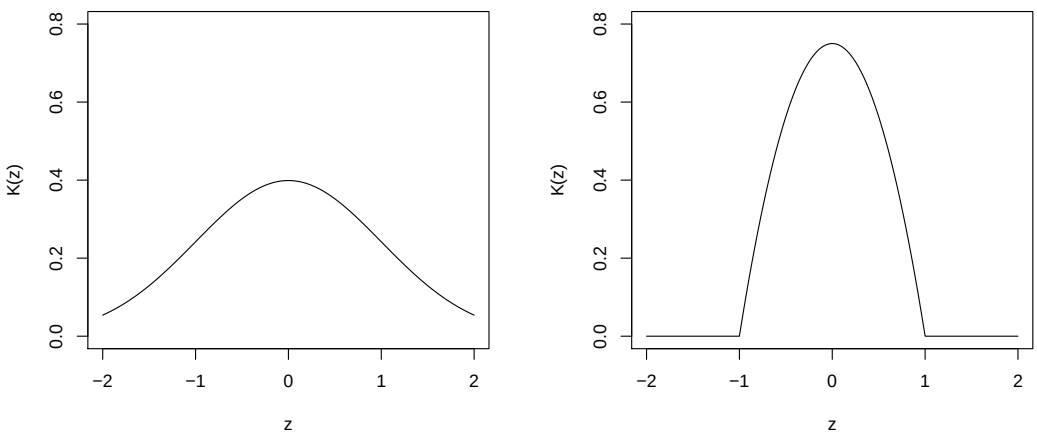
$K(\cdot)$ is here a bounded around zero symmetric univariate density function that fulfills the condition $\int_{\mathbb{R}} |z^5| K(z) dz < \infty$. Typical examples for K(z) are here the Gaussian kernel

$$K(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right)$$

and the Epanechnikov kernel

$$K(z) := \begin{cases} \frac{3}{4}(1 - z^2), \text{ for } z \in [-1, 1] \\ 0 \text{ otherwise} \end{cases},$$

which are shown in Figure 2.1.



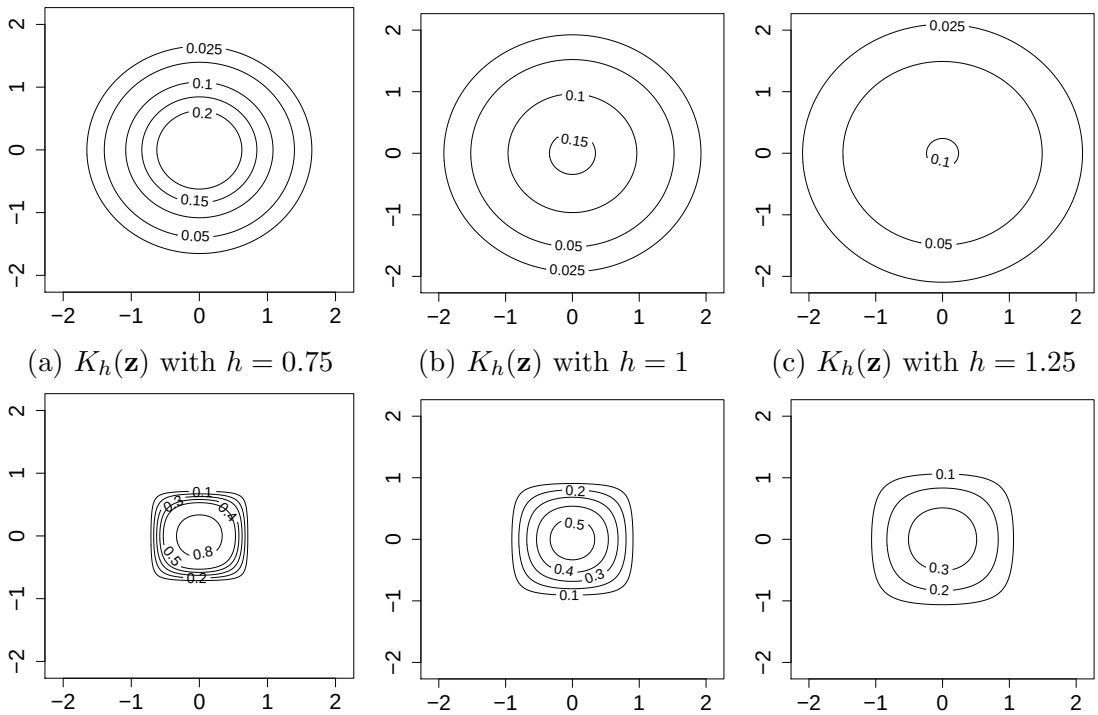
(a) Gaussian kernel (b) Epanechnikov kernel

Figure 2.1.: Gaussian kernels (a) and Epanechnikov kernel (b).

For multivariate kernels, we will use two different notations. The first one is the so called product kernel $K_h(\mathbf{z})$: For $\mathbf{z} = (z_1, \ldots, z_d)^\top \in \mathbb{R}^d$,

$$K_h(\mathbf{z}) := \prod_{j=1}^{d} K_h(z_j) = \frac{1}{h^d} \prod_{j=1}^{d} K\left(\frac{z_j}{h}\right), \text{ for } h > 0,$$

where $K_h(z)$ is a univariate kernel with bandwidth h. Figure 2.2 shows the contour plots resulting for two dimensional product kernels using Gaussian and Epanechnikov univariate kernels with different bandwidths h.



(a) $K_h(\mathbf{z})$ with $h = 0.75$ (b) $K_h(\mathbf{z})$ with $h = 1$ (c) $K_h(\mathbf{z})$ with $h = 1.25$

(d) $K_h(\mathbf{z})$ with $h = 0.75$ (e) $K_h(\mathbf{z})$ with $h = 1$ (f) $K_h(\mathbf{z})$ with $h = 1.25$

Figure 2.2.: Contour plots of bivariate product kernels using univariate Gaussian kernels (a)-(c) and univariate Epanechnikov kernels (d)-(f).

We will also use the following notation, which we will call the general kernel:

$$K_H(\mathbf{z}) := \frac{1}{\sqrt{|H|}} \mathcal{K}(H^{-\frac{1}{2}}\mathbf{z}),$$

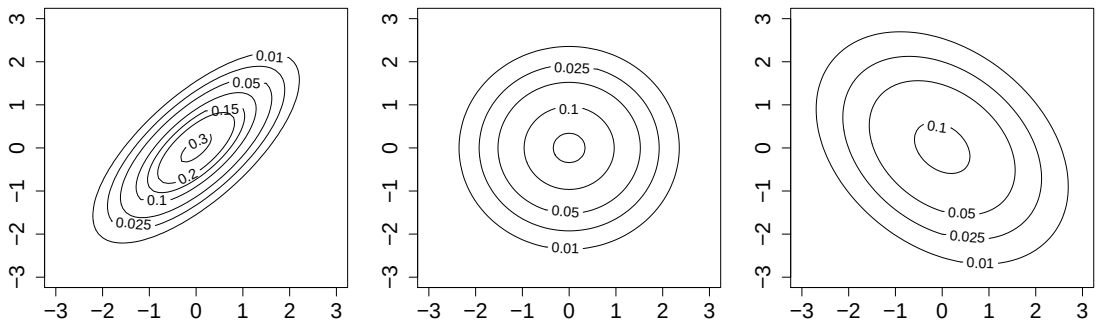
where $|H|$ is denoting the determinant of a matrix $H \in \mathbb{R}^{d\times d}$ and $\mathcal{K}(\cdot)$ is a around zero symmetric multivariate density satisfying

$$\int_{\mathbb{R}^d} \mathbf{z}\mathbf{z}^\top \mathcal{K}(\mathbf{z}) d\mathbf{z} = \mu_{\mathcal{K}} I, \tag{2.11}$$

where the integral is taken componentwise. Here $\mu_{\mathcal{K}}$ is a scalar and I is the identity matrix. H is a $d \times d$ Matrix which is symmetric and positive definite. We call H the bandwidth matrix. $H^{-\frac{1}{2}}$ is the inverse of the square root of H, which is defined such that $H = H^{\frac{1}{2}} H^{\frac{1}{2}}$. Note that because H is positive definite and symmetric $H^{\frac{1}{2}}$ is symmetric and unique. Furthermore for $\mathcal{K}(\cdot)$ should hold that $\int_{\mathbb{R}^d} |z_j z_l z_k| \mathcal{K}(\mathbf{z}) d\mathbf{z} < \infty$, where z_j, z_l, z_k are components of $\mathbf{z}$ with $j, l, k \in \{1, \ldots, d\}$.

Remark 2.3. *Note that the product kernel $K_h(\boldsymbol{z})$ is a special case of the general kernel $K_H(\boldsymbol{z})$. This is easy to see, since we can use for the multivariate density of $K_H(\boldsymbol{z})$ the product of univariate densities $K(\cdot)$ and for H a diagonal matrix where every diagonal entry corresponds to h^2.*

Figure 2.3 shows the contour plots resulting for two dimensional general kernels using for $\mathcal{K}(\cdot)$ the standard bivariate normal density and different bandwidth matrices H.



(a) $\mathcal{K}_H(\mathbf{z})$ with $H = \begin{pmatrix} 0.7 & 0.5 \\ 0.5 & 0.7 \end{pmatrix}$

(b) $\mathcal{K}_H(\mathbf{z})$ with $H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

(c) $\mathcal{K}_H(\mathbf{z})$ with $H = \begin{pmatrix} 1.5 & -0.5 \\ -0.5 & 1.5 \end{pmatrix}$

Figure 2.3.: Contour plots of bivariate general kernels using the multivariate standard normal density.

2.5. Local linear regression

This chapter is based on the introduction of the paper Ruppert and Wand (1994) and Chapter 2.3.1 of the book Fan and Gijbels (1996).

Suppose that we have $i.i.d$ $(p+1)$-dimensional observations

$$(y_i, \mathbf{x}_i) = (y_i, x_{i1}, \ldots, x_{ip}) \text{ for } i \in \{1, \ldots, n\},$$

which arise from the $(p+1)$-dimensional random vectors

$$(Y_i, \mathbf{X}_i) = (Y_i, X_{i1}, \ldots, X_{ip}) \; i.i.d \text{ for } i \in \{1, \ldots, n\}.$$

The main purpose of multivariate local linear regression is to estimate

$$m(\mathbf{x}) := E(Y|\mathbf{X} = \mathbf{x}), \text{ for } \mathbf{x} \in \mathbb{R}^p,$$

where Y is a univariate random variable and $\mathbf{X}$ is a p-dimensional random vector having the same distributions as Y_i and $\mathbf{X}_i$ for $i \in \{1, \ldots, n\}$.

For fixed $\mathbf{x} \in \mathbb{R}^p$, we approximate the function $m(\cdot)$ locally by its first order Taylor expansion,

$$m(\tilde{\mathbf{x}}) \approx m(\mathbf{x}) + \sum_{j=1}^{p} \frac{\partial m(\mathbf{x})}{\partial x_j}(\tilde{x}_j - x_j) = m(\mathbf{x}) + \nabla m(\mathbf{x})^\top (\tilde{\mathbf{x}} - \mathbf{x}), \tag{2.12}$$

for every $\tilde{\mathbf{x}} \in \mathbb{R}^p$ in a neighborhood of $\mathbf{x}$, where $\nabla m(\mathbf{x}) := (\frac{\partial m(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial m(\mathbf{x})}{\partial x_p})$ is the gradient of $m(\mathbf{x})$. Notice that $\mathbf{x}$ (and thereby $m(\mathbf{x})$) in (2.12) stays the same for different $\tilde{\mathbf{x}}$. Since $m(\tilde{\mathbf{x}}) = E(Y|\mathbf{X} = \tilde{\mathbf{x}})$ is a (conditional) expectation of $\mathbf{Y}$ and $\tilde{\mathbf{x}}$ is free to choose the following linear model is suggested:

$$\begin{aligned} Y_i &= m(\mathbf{x}) + \nabla m(\mathbf{x})^\top (\mathbf{x}_i - \mathbf{x}) + \epsilon_i, \\ &:= \beta_0 + \boldsymbol{\beta}_1^\top (\mathbf{x}_i - \mathbf{x}) + \epsilon_i, \end{aligned}$$

with $\beta_0 := m(\mathbf{x})$, $\boldsymbol{\beta}_1 := \nabla m(\mathbf{x})$ and random ϵ_i. The fact that in this model the parameter $m(\mathbf{x})$ stays the same for all $i \in \{1, \ldots, n\}$, leads to an approach for estimating $m(\mathbf{x})$ based on local least squares fitting using kernel weights. To see this one has to interpret Y_i as the response random variable, ϵ_i as the error random variable, $(1, (\mathbf{x}_i - \mathbf{x})^\top)^\top$ as the covariate vector and $\boldsymbol{\beta}^\top := (\beta_0, \boldsymbol{\beta}_1^\top)^\top = (m(\mathbf{x}), \nabla m(\mathbf{x})^\top)^\top$ as the parameter vector. The estimator of $m(\mathbf{x})$ will then be the solution $\hat{\beta}_0$ of the following problem:

$$\min_{\beta_0, \boldsymbol{\beta}_1} \sum_{i=1}^{n} \{y_i - \beta_0 - \boldsymbol{\beta}_1^\top (\mathbf{x}_i - \mathbf{x})\}^2 W_b(\mathbf{x}_i - \mathbf{x}), \tag{2.13}$$

with weights $W_b(\mathbf{x}_i - \mathbf{x})$ for $i \in \{1, \ldots, n\}$, where $W_b(\cdot)$ is a p-dimensional product kernel with bandwidth b, given by

$$W_b(\mathbf{x}) := \frac{1}{b^p} \prod_{j=1}^{p} K\left(\frac{x_j}{b}\right), \text{ for } b > 0, \ \mathbf{x} \in \mathbb{R}^p.$$

The reason for using these weights is the following: Since we used a Taylor expansion before, the observations which have $\mathbf{x}_i$-values $i \in \{1, \ldots, n\}$ close to $\mathbf{x}$ should have a bigger impact on the estimation, whereas the observation with values far away from $\mathbf{x}$ should have a smaller impact. One possibility to achieve this behavior is to use the symmetric multidimensional kernels $W_b(\mathbf{x})$. The mode of those kernels is at 0 and the function values decrease, when moving away from 0. Therefore, by inserting the the vectors $(\mathbf{x}_i - \mathbf{x})$ for $i \in \{1, \ldots, n\}$ in the symmetric multidimensional kernels $W_b(\cdot)$, we get the desired properties for our weights.

The problem (2.13) is a weighted least squares problem (see Section 2.3). Using the notations,

$$\mathcal{Y} := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

$$\mathcal{X} := \begin{pmatrix} 1 & (x_{1,1} - x_1) & \ldots & (x_{1,p} - x_p) \\ \vdots & \vdots & & \vdots \\ 1 & (x_{n,1} - x_1) & \ldots & (x_{n,p} - x_p) \end{pmatrix},$$

$$\mathcal{W} := \begin{pmatrix} W_b(\mathbf{x}_1 - \mathbf{x}) & & 0 \\ & \ddots & \\ 0 & & W_b(\mathbf{x}_i - \mathbf{x}) \end{pmatrix},$$

it has the following solution

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_1 \end{pmatrix} = (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y}.$$

The estimator for $m(\mathbf{x})$ is then

$$\hat{m}(\mathbf{x}) := \hat{\beta}_0 = e_1^\top (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y},$$

where $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{p+1}$. Note also that $\hat{\boldsymbol{\beta}}_1$ can be used as an estimator for $\nabla m(\mathbf{x})$.

2.6. Error measures

In this section we will introduce error measures that quantify the quality of estimators. Let for this purpose θ be an unknown quantity that one has to estimate based on n observations $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,r})^\top$, which arise from i.i.d. random vectors $\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{i,r})^\top$ with $i \in \{1, \ldots, n\}$. These random vectors follow a certain model depending on θ. We will denote the matrix containing (in its rows) all this random vectors by $\mathcal{Z} := (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top)^\top$ and the matrix containing the corresponding observations by $\mathscr{Z} := (\mathbf{z}_1^\top, \ldots, \mathbf{z}_n^\top)^\top$. Throughout this thesis we will denote estimators with a hat. We use the notation $\hat{\theta}$ for the random variable $\hat{\theta}(\mathcal{Z})$ as well as for the real value $\hat{\theta}(\mathscr{Z})$, which is the realization in the case of $\mathcal{Z} = \mathscr{Z}$ (see, Czado and Schmidt, 2011, p. 72).

The estimator of an unknown probability density function $f_{\mathbf{Y}}(\mathbf{y})$ at $\mathbf{y} \in \mathbb{R}^d$ of a random vector $\mathbf{Y}$ is denoted by $\hat{f}_{\mathbf{Y}}(\mathbf{y})$. For the following definitions $\hat{f}_{\mathbf{Y}}(\mathbf{y})$ will represent the random variable $\hat{f}_{\mathbf{Y}}(\mathbf{y})(\mathcal{Z})$.

The definition of the mean squared error of such an estimator $\hat{f}_{\mathbf{Y}}(\mathbf{y})$ at $\mathbf{y}$ is:

$$\mathrm{MSE}\left[\hat{f}_{\mathbf{Y}}, \mathbf{y}\right] := \mathrm{E}\left[\left(\hat{f}_{\mathbf{Y}}(\mathbf{y}) - f_{\mathbf{Y}}(\mathbf{y})\right)^2\right] = \mathrm{Bias}\left[\hat{f}_{\mathbf{Y}}(\mathbf{y})\right]^2 + \mathrm{Var}\left[\hat{f}_{\mathbf{Y}}(\mathbf{y})\right].$$

The integrated squared error of the estimator is defined as follows:

$$\mathrm{ISE}\left[\hat{f}_{\mathbf{Y}}\right] := \int_{\mathbb{R}^d} \left(\hat{f}_{\mathbf{Y}}(\mathbf{y}) - f_{\mathbf{Y}}(\mathbf{y})\right)^2 d\mathbf{y}.$$

We will also use the mean integrated squared error:

$$\mathrm{MISE}\left[\hat{f}_{\mathbf{Y}}\right] := \mathrm{E}\left[\mathrm{ISE}\left[\hat{f}_{\mathbf{Y}}\right]\right].$$

By using the famous Fubini Tonelli theorem we get in our case that the MISE is equal to the integrated mean squared error:

$$\mathrm{MISE}\left[\hat{f}_{\mathbf{Y}}\right] = \int_{\mathbb{R}^d}\left[\mathrm{Bias}\left[\hat{f}_{\mathbf{Y}}(\mathbf{y})\right]^2 + \mathrm{Var}\left[\hat{f}_{\mathbf{Y}}(\mathbf{y})\right]\right] d\mathbf{y}.$$

Another important error measure is the integrated absolute error:

$$\mathrm{IAE}\left[\hat{f}_{\mathbf{Y}}\right] := \int_{\mathbb{R}^d}\left|\hat{f}_{\mathbf{Y}}(\mathbf{y}) - f_{\mathbf{Y}}(\mathbf{y})\right| d\mathbf{y}.$$

Further we will use the mean integrated absolute error:

$$\mathrm{MIAE}\left[\hat{f}_{\mathbf{Y}}\right] := \mathrm{E}\left[\mathrm{IAE}\left[\hat{f}_{\mathbf{Y}}\right]\right].$$

Remark 2.4. $\mathrm{MSE}\left[\hat{f}_{\boldsymbol{Y}}, \boldsymbol{y}\right]$ *is a deterministic functions of* $\boldsymbol{y}$. *Further* $\mathrm{MISE}\left[\hat{f}_{\boldsymbol{Y}}\right]$ *and* $\mathrm{MIAE}\left[\hat{f}_{\boldsymbol{Y}}\right]$ *are real values, while* $\mathrm{ISE}\left[\hat{f}_{\boldsymbol{Y}}\right]$ *and* $\mathrm{IAE}\left[\hat{f}_{\boldsymbol{Y}}\right]$ *are random variables which depend on the random matrix* $\mathcal{Z}$.

3. Conditional and partial copulas

In this chapter we will first state some basic notations that will be used in this thesis. We will furthermore define and analyze conditional and partial copulas and illustrate these with examples.

3.1. Notations

To avoid misunderstandings, this section will introduce the basic notations and definitions.

Let $\mathbf{Y} := (Y_1, \ldots, Y_d)^\top$ be a random vector and $\mathbf{y} := (y_1, \ldots, y_d)^\top \in \mathbb{R}^d$ be a realization of this random vector. (We will later assume that our response-data will be distributed like $\mathbf{Y}$.) Let $\mathbf{X} := (X_1, \ldots, X_p)^\top$ be a random vector and $\mathbf{x} := (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ be a realization of this random vector. (We will later assume that our covariate-data will be distributed like $\mathbf{X}$.)

Denote the conditional marginal distribution function of Y_j with $j \in \{1, \ldots, d\}$, given $\mathbf{X} = \mathbf{x}$ as

$$F_{Y_j|\mathbf{X}}(y_j|\mathbf{x}) = P(Y_j \leq y_j|\mathbf{X} = \mathbf{x}) = \frac{\int_{-\infty}^{y_j} f_{Y_j,\mathbf{X}}(\tilde{y_j}, \mathbf{x}) d\tilde{y_j}}{f_{\mathbf{X}}(\mathbf{x})}, \text{ for } j \in \{1, \ldots, d\},$$

and the conditional marginal density of Y_j given $\mathbf{X} = \mathbf{x}$ as

$$f_{Y_j|\mathbf{X}}(y_j|\mathbf{x}) = \frac{f_{Y_j,\mathbf{X}}(y_j, \mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} = \frac{\partial}{\partial y_j} F_{Y_j|\mathbf{X}}(y_j|\mathbf{x}).$$

Here $f_{Y_j,\mathbf{X}}(y_j, \mathbf{x})$ is the joint density of the random vector $(Y_j, X_1, \ldots, X_p)^\top$ and $f_{\mathbf{X}}(\mathbf{x})$ is the joint density of the random vector $\mathbf{X} = (X_1, \ldots, X_p)^\top$, with $f_{\mathbf{X}}(\mathbf{x}) > 0$. The conditional joint distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ will be denoted as

$$\begin{aligned} F_{\mathbf{Y}|\mathbf{X}}(y_1, \ldots, y_d|\mathbf{x}) &= P(Y_1 \leq y_1, \ldots, Y_d \leq y_d|\mathbf{X} = \mathbf{x}) = \\ &= \frac{\int_{-\infty}^{y_d} \ldots \int_{-\infty}^{y_1} f_{\mathbf{Y},\mathbf{X}}(\tilde{y_1}, \ldots, \tilde{y_d}, \mathbf{x}) d\tilde{y_1} \ldots d\tilde{y_d}}{f_{\mathbf{X}}(\mathbf{x})}, \end{aligned} \tag{3.1}$$

and the conditional joint density of $\mathbf{Y}$, given $\mathbf{X} = \mathbf{x}$, as

$$f_{\mathbf{Y}|\mathbf{X}}(y_1, \ldots, y_d|\mathbf{x}) = \frac{f_{\mathbf{Y},\mathbf{X}}(y_1, \ldots, y_d, \mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} = \frac{\partial^d}{\partial y_1 \ldots \partial y_d} F_{\mathbf{Y}|\mathbf{X}}(y_1, \ldots, y_d|\mathbf{x}),$$

where $f_{\mathbf{Y},\mathbf{X}}(\mathbf{y}, \mathbf{x})$ is the joint density of the random vector $(Y_1, \ldots, Y_d, X_1, \ldots, X_p)^\top$ and $f_{\mathbf{X}}(\mathbf{x})$ is the joint density of the random vector $(X_1, \ldots, X_p)^\top$, with $f_{\mathbf{X}}(\mathbf{x}) > 0$.

Remark 3.1. *We will assume that all densities functions $f_{\boldsymbol{X}}(\boldsymbol{x})$ and all conditional density functions $f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})$ used in this thesis, have bounded continuous first, second and third order derivatives with respect to $\boldsymbol{x}$ and $\boldsymbol{z}$.*

By using random variables and random vector as an input of the conditional distributions we denote/define the random variables in Table 3.1.

Table 3.1.: Summary of different distribution situations.

	For $j \in \{1, \dots, d\}$	For $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$ and $\mathbf{y} = (y_1, \dots, y_d)^\top$
Usual cond. cdf		
Notation/Definition:	$F_{Y_j\|\mathbf{X}}(y_j\|\mathbf{x}) := P(Y_j \leq y_j\|\mathbf{X} = \mathbf{x})$	$F_{Y\|\mathbf{X}}(\mathbf{y}\|\mathbf{x}) := P(Y_1 \leq y_1, \dots, Y_d \leq y_d\|\mathbf{X} = \mathbf{x})$
other formulation	$= \frac{\int_{-\infty}^{y_j} f_{Y_j,\mathbf{X}}(\tilde{y}_j,\mathbf{x})d\tilde{y}_j}{f_{\mathbf{X}}(\mathbf{x})}$ (nonrandom function of $(y_j, \mathbf{x})$)	$= \frac{\int_{-\infty}^{y_d} \dots \int_{-\infty}^{y_1} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1,\dots,\tilde{y}_d,\mathbf{x})d\tilde{y}_1 \dots d\tilde{y}_d}{f_{\mathbf{X}}(\mathbf{x})}$ (nonrandom function of $(\mathbf{y}, \mathbf{x})$)
Cond. cdf with random condition		
Notation/Definition:	$F_{Y_j\|\mathbf{X}}(y_j\|\mathbf{X}) := \frac{\int_{-\infty}^{y_j} f_{Y_j,\mathbf{X}}(\tilde{y}_j,\mathbf{X})d\tilde{y}_j}{f_{\mathbf{X}}(\mathbf{X})}$ (random function of y_j)	$F_{Y\|\mathbf{X}}(\mathbf{y}\|\mathbf{X}) := \frac{\int_{-\infty}^{y_d} \dots \int_{-\infty}^{y_1} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1,\dots,\tilde{y}_d,\mathbf{X})d\tilde{y}_1 \dots d\tilde{y}_d}{f_{\mathbf{X}}(\mathbf{X})}$ (random function of $\mathbf{y}$)
Cond. cdf with random response		
Notation/Definition:	$F_{Y_j\|\mathbf{X}}(Y_j\|\mathbf{x}) := \frac{\int_{-\infty}^{Y_j} f_{Y_j,\mathbf{X}}(\tilde{y}_j,\mathbf{x})d\tilde{y}_j}{f_{\mathbf{X}}(\mathbf{x})}$ (random function of $\mathbf{x}$)	$F_{Y\|\mathbf{X}}(\mathbf{Y}\|\mathbf{x}) := \frac{\int_{-\infty}^{Y_d} \dots \int_{-\infty}^{Y_1} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1,\dots,\tilde{y}_d,\mathbf{x})d\tilde{y}_1 \dots d\tilde{y}_d}{f_{\mathbf{X}}(\mathbf{x})}$ (random function of $\mathbf{x}$)
Cond. cdf with rand. res. & cond.		
Notation/Definition:	$F_{Y_j\|\mathbf{X}}(Y_j\|\mathbf{X}) := \frac{\int_{-\infty}^{Y_j} f_{Y_j,\mathbf{X}}(\tilde{y}_j,\mathbf{X})d\tilde{y}_j}{f_{\mathbf{X}}(\mathbf{X})}$ (random variable)	$F_{Y\|\mathbf{X}}(\mathbf{Y}\|\mathbf{X}) := \frac{\int_{-\infty}^{Y_d} \dots \int_{-\infty}^{Y_1} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1,\dots,\tilde{y}_d,\mathbf{X})d\tilde{y}_1 \dots d\tilde{y}_d}{f_{\mathbf{X}}(\mathbf{X})}$ (random variable)
Cond. probability		
Notation/Definition:	$P(Y_j \leq y_j\|\mathbf{X}) := \mathrm{E}[1_{\{Y_j \leq y_j\}}\|\mathbf{X}]$ (random function of y_i)	$P(\mathbf{Y} \leq \mathbf{y}\|\mathbf{X}) := \mathrm{E}[1_{\{Y_1 \leq y_1,\dots,Y_d \leq y_d\}}\|\mathbf{X}]$ (random function of $\mathbf{y}$)

Note that we can **not** simply write the defined random functions $F_{Y_j|\mathbf{X}}(y_j|\mathbf{X})$, $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$, $F_{Y_j|\mathbf{X}}(Y_j|\mathbf{X})$, $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})$ as follows:

$$\begin{aligned}
F_{Y_j|\mathbf{X}}(y_j|\mathbf{X}) &\neq P(Y_j \leq y_j|\tilde{\mathbf{X}} = \mathbf{X}) && for\ j \in \{1,\dots,d\},\\
F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X}) &\neq P(Y_1 \leq y_1,\dots,Y_d \leq y_d|\tilde{\mathbf{X}} = \mathbf{X}),\\
F_{Y_j|\mathbf{X}}(Y_j|\mathbf{x}) &\neq P(\tilde{Y}_j \leq Y_j|\mathbf{X} = \mathbf{x}) && for\ j \in \{1,\dots,d\},\\
F_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{x}) &\neq P(\tilde{Y}_1 \leq Y_1,\dots,\tilde{Y}_d \leq Y_d|\tilde{\mathbf{X}} = \mathbf{x}),\\
F_{Y_j|\mathbf{X}}(Y_j|\mathbf{X}) &\neq P(\tilde{Y}_j \leq Y_j|\tilde{\mathbf{X}} = \mathbf{X}) && for\ j \in \{1,\dots,d\},\\
F_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X}) &\neq P(\tilde{Y}_1 \leq Y_1,\dots,\tilde{Y}_d \leq Y_d|\tilde{\mathbf{X}} = \mathbf{X}),
\end{aligned}$$

where $\tilde{Y}_j$ $j \in \{1,\dots,d\}$ are independent copies of the random variables Y_j and $\tilde{X}$ is a independent copy of the random vector X.

This is easy to see, because the terms on the left sides are random variables whereas the terms on the right side are real values. For example $P(\tilde{Y}_1 \leq Y_1|\tilde{\mathbf{X}} = \mathbf{X})$ can be rewritten as $P(\tilde{Y}_1 - Y_1 \leq 0|\tilde{\mathbf{X}} - \mathbf{X} = 0)$, which is just a probability of the random variable $(\tilde{Y}_1 - Y_1)$ being smaller then zero under the condition that a random variable $(\tilde{\mathbf{X}} - \mathbf{X})$ is equal to zero.

We can however state in terms of probability theory the following $\forall\omega \in \Omega$:

$$\begin{aligned}
F_{Y_j|\mathbf{X}}(y_j|\mathbf{X})(\omega) &= P(Y_j \leq y_j|\tilde{\mathbf{X}} = \mathbf{X}(\omega)) && for\ j \in \{1,\dots,d\},\\
F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})(\omega) &= P(Y_1 \leq y_1,\dots,Y_d \leq y_d|\tilde{\mathbf{X}} = \mathbf{X}(\omega)),\\
F_{Y_j|\mathbf{X}}(Y_j|\mathbf{x})(\omega) &= P(\tilde{Y}_j \leq Y_j(\omega)|\mathbf{X} = \mathbf{x}) && for\ j \in \{1,\dots,d\},\\
F_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{x})(\omega) &= P(\tilde{Y}_1 \leq Y_1(\omega),\dots,\tilde{Y}_d \leq Y_d(\omega)|\tilde{\mathbf{X}} = \mathbf{x}),\\
F_{Y_j|\mathbf{X}}(Y_j|\mathbf{X})(\omega) &= P(\tilde{Y}_j \leq Y_j(\omega)|\tilde{\mathbf{X}} = \mathbf{X}(\omega)) && for\ j \in \{1,\dots,d\},\\
F_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})(\omega) &= P(\tilde{Y}_1 \leq Y_1(\omega),\dots,\tilde{Y}_d \leq Y_d(\omega)|\tilde{\mathbf{X}} = \mathbf{X}(\omega)).
\end{aligned}$$

The following theorem provides us a link between the notations $F_{Y_j|\mathbf{X}}(y_j|\mathbf{X})$ and $P(Y_j \leq y_j|\mathbf{X})$ as well as the link between $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$ and $P(\mathbf{Y} \leq \mathbf{y}|\mathbf{X})$.

Theorem 3.1. *By using the notations from Table 3.1 and assuming that the densities $f_{Y_j,\boldsymbol{X}}(\cdot,\cdot)$, $f_{\boldsymbol{Y},\boldsymbol{X}}(\cdot,\cdot)$, $f_{\boldsymbol{X}}(\cdot)$ exist, we can state:*

$$F_{Y_j|\boldsymbol{X}}(y_j|\boldsymbol{X}) = P(Y_j \leq y_j|\boldsymbol{X})\ for\ j \in \{1,\dots,d\}\ a.s., \tag{3.2}$$
$$F_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X}) = P(\boldsymbol{Y} \leq \boldsymbol{y}|\boldsymbol{X})\ a.s.. \tag{3.3}$$

Proof. (Theorem 3.1)
Since (3.2) can be considered as a special case of (3.3) we will only proof the second statement in detail. By definition, $P(\mathbf{Y} \leq \mathbf{y}|\mathbf{X}) := E[1_{\{Y_1 \leq y_1,\dots,Y_d \leq y_d\}}|\mathbf{X}]$. Remembering the definition of the conditional expectation we have to show the following:

(i) $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$ is measurable w.r.t. the σ-algebra $\sigma(\mathbf{X})$.
(ii) For every bounded random variable Z which is measurable w.r.t. $\sigma(\mathbf{X})$, we have $E[Z \cdot F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})] = E[Z \cdot 1_{\{Y_1 \leq y_1, \ldots, Y_d \leq y_d\}}]$.

Statement (i) is clear because we assume that $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$ is a continuous function in $\mathbf{X}$, so only statement (ii) is remaining. At first we use the following theorem from measure theory: If Z is $\sigma(\mathbf{X})$ measurable then a function $g(.) : \mathbb{R}^p \to \mathbb{R}$ exists such that $g(\mathbf{X}) = Z$ almost surely (see, Billingsley, 1995, p. 255). Therefore we can rewrite statement (ii) as follows:

$$
\begin{aligned}
(ii) \iff & E[g(\mathbf{X}) \cdot F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})] = E[g(\mathbf{X}) \cdot 1_{\{Y_1 \leq y_1, \ldots, Y_d \leq y_d\}}] \\
\iff & \int_{\mathbb{R}^p} g(\mathbf{x}) F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
& \qquad = \int_{\mathbb{R}^p} \int_{\mathbb{R}^d} g(\mathbf{x}) \cdot 1_{\{\tilde{y}_1 \leq y_1, \ldots, \tilde{y}_d \leq y_d\}} f_{\mathbf{Y},\mathbf{X}}(\tilde{\mathbf{y}}, \mathbf{x}) d\tilde{\mathbf{y}} d\mathbf{x} \\
\iff & \int_{\mathbb{R}^p} g(\mathbf{x}) \int_{-\infty}^{y_d} \cdots \int_{-\infty}^{y_1} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1, \ldots, \tilde{y}_d, \mathbf{x}) d\tilde{\mathbf{y}} d\mathbf{x} \\
& \qquad = \int_{\mathbb{R}^p} \int_{-\infty}^{y_d} \cdots \int_{-\infty}^{y_1} g(\mathbf{x}) f_{\mathbf{Y},\mathbf{X}}(\tilde{\mathbf{y}}, \mathbf{x}) d\tilde{\mathbf{y}} d\mathbf{x},
\end{aligned}
$$

where the last step uses the definition of the condition distribution (3.1). □

3.2. Conditional copulas

As mentioned in Gijbels et al. (2011) and Gijbels et al. (2012), we can define a conditional copula as follows:

Definition 3.1. *(Conditional copula)*
If $F_{Y_j|\boldsymbol{X}}(y_j|\boldsymbol{x})$ for $j \in \{1, \ldots, d\}$ are continuous, we know by Sklar's theorem (2.4) that under the condition $\boldsymbol{X} = \boldsymbol{x}$, there exists a unique copula $C_{\boldsymbol{U}|\boldsymbol{X}}(u_1, \ldots, u_d|\boldsymbol{x})$ such that

$$
F_{\boldsymbol{Y}|\boldsymbol{X}}(y_1, \ldots, y_d|\boldsymbol{x}) = C_{\boldsymbol{U}|\boldsymbol{X}}\left(F_{Y_1|\boldsymbol{X}}(y_1|\boldsymbol{x}), \ldots, F_{Y_d|\boldsymbol{X}}(y_d|\boldsymbol{x})\middle|\boldsymbol{x}\right). \tag{3.4}
$$

We call $C_{\boldsymbol{U}|\boldsymbol{X}}(u_1, \ldots, u_d|\boldsymbol{x})$ the conditional copula at $\boldsymbol{X} = \boldsymbol{x}$.

This notation was introduced in the case of a bivariate random vector $\mathbf{Y}$ by Patton (2006). Our goal will later be to estimate the conditional copula density $c_{\mathbf{U}|\mathbf{X}}(u_1, \ldots, u_d|\mathbf{x})$.

We define the random vector $\mathbf{U}_{\mathbf{x}}$ by

$$
\mathbf{U}_{\mathbf{x}} := (F_{Y_1|\mathbf{X}}(Y_1|\mathbf{x}), \ldots, F_{Y_d|\mathbf{X}}(Y_d|\mathbf{x}))^\top, \text{ for all } \mathbf{x} \in \mathbb{R}^p,
$$

as well as the random variables U_j

$$
U_j := F_{Y_j|\mathbf{X}}(Y_j|\mathbf{x}), \text{ for } j \in \{1, \ldots, d\}, \text{ for all } \mathbf{x} \in \mathbb{R}^p.
$$

Lemma 3.1. *For all $\boldsymbol{x} \in \mathbb{R}^p$, given $\boldsymbol{X} = \boldsymbol{x}$: $U_j \sim U(0,1)$ $\forall j \in \{1, \dots, d\}$.*

Proof. (Lemma 3.1)
Let $j \in \{1, \dots, d\}$. Then

$$\begin{aligned} P(U_j \leq t | \mathbf{X} = \mathbf{x}) &= P(F_{Y_j|\mathbf{X}}(Y_j|\mathbf{x}) \leq t | \mathbf{X} = \mathbf{x}) \\ &= P(Y_j \leq F^{-1}_{Y_j|\mathbf{X}}(t|\mathbf{x}) | \mathbf{X} = \mathbf{x}) = F_{Y_j|\mathbf{X}}(F^{-1}_{Y_j|\mathbf{X}}(t|\mathbf{x})|\mathbf{x}) = t, \end{aligned} \tag{3.5}$$

where $F^{-1}_{Y_j|\mathbf{X}}(t|\mathbf{x})$ is the quantile function of Y_j given $\mathbf{X} = \mathbf{x}$. □

Note that the (marginal-) distributions of the $U_j's$ are independent of the conditioning vector $\mathbf{X}$. This is the reason why we leave out the $\mathbf{x}$ in the notation of U_j for $j \in \{1, \dots, d\}$.

Remark 3.2. *The joint conditional distribution (under the condition $\boldsymbol{X} = \boldsymbol{x}$) of $\boldsymbol{U_x}$ is the conditional copula $C_{\boldsymbol{U}|\boldsymbol{X}}(u_1, \dots, u_d|\boldsymbol{x})$ and which (in contrast to the marginal distributions) depends on the value of $\boldsymbol{x}$.*

Lemma 3.2. *It holds $U_{Y_j|\boldsymbol{X}} := F_{Y_j|\boldsymbol{X}}(Y_j|\boldsymbol{X}) \sim U(0,1)$.*

Proof. (Lemma 3.2)
For all $j \in \{1, \dots, d\}$ it holds that:

$$\begin{aligned} P(U_{Y_j|\mathbf{X}} < t) &= P(F_{Y_j|\mathbf{X}}(Y_j|\mathbf{X}) < t) \\ &= \int_{\mathbb{R}^p} P(F_{Y_j|\mathbf{X}}(Y_j|\mathbf{X}) < t | \mathbf{X} = \mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^p} P(F_{Y_j|\mathbf{X}}(Y_j|\mathbf{x}) < t | \mathbf{X} = \mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^p} t f_X(\mathbf{x}) d\mathbf{x} = t, \end{aligned}$$

where we used (3.5). □

By defining the random vector $U_{\mathbf{X}}$ as:

$$\mathbf{U_X} := (U_{Y_1|\mathbf{X}}, \dots, U_{Y_d|\mathbf{X}})^\top = (F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}), \dots, F_{Y_d|\mathbf{X}}(Y_d|\mathbf{X}))^\top,$$

we state the following lemma:

Lemma 3.3.

$$P(F_{Y_1|\boldsymbol{X}}(Y_1|\boldsymbol{X}) \leq u_1, \dots, F_{Y_d|\boldsymbol{X}}(Y_d|\boldsymbol{X}) \leq u_d | \boldsymbol{X} = \boldsymbol{x}) = C_{\boldsymbol{U}|\boldsymbol{X}}(u_1, \dots, u_d|\boldsymbol{x}),$$

where $C_{\boldsymbol{U}|\boldsymbol{X}}(u_1, \dots, u_d|\boldsymbol{x})$ is the conditional copula as defined in equation (3.4).

Proof. (Lemma 3.3)

$$\begin{aligned}
P(\mathbf{U_X} \leq \mathbf{u}|\mathbf{X} = \mathbf{x}) &= P(F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \leq u_1, \ldots, F_{Y_d|\mathbf{X}}(Y_d|\mathbf{X}) \leq u_d|\mathbf{X} = \mathbf{x}) \\
&= P(F_{Y_1|\mathbf{X}}(Y_1|\mathbf{x}) \leq u_1, \ldots, F_{Y_d|\mathbf{X}}(Y_d|\mathbf{x}) \leq u_d|\mathbf{X} = \mathbf{x}) \\
&= P(Y_1 \leq F^{-1}_{Y_1|\mathbf{X}}(u_1|\mathbf{x}), \ldots, Y_d \leq F^{-1}_{Y_d|\mathbf{X}}(u_d|\mathbf{x})|\mathbf{X} = \mathbf{x}) \\
&= F_{\mathbf{Y}|\mathbf{X}}(F^{-1}_{Y_1|\mathbf{X}}(u_1|\mathbf{x}), \ldots, F^{-1}_{Y_d|\mathbf{X}}(u_d|\mathbf{x})|\mathbf{x}) = C_{\mathbf{U}|\mathbf{X}}(u_1, \ldots, u_d|\mathbf{x}).
\end{aligned}$$

□

This shows that the joint conditional distribution (under the condition $\mathbf{X} = \mathbf{x}$) of $\mathbf{U_X}$ is (like for $\mathbf{U_x}$) also the conditional copula $C_{\mathbf{U}|\mathbf{X}}(u_1, \ldots, u_d|\mathbf{x})$.

To avoid some boundary issues later we define the random vector $\mathbf{Z}$ as a transformation of $\mathbf{U_X}$ by applying the inverse univariate standard normal distribution function $\Phi^{-1}(\cdot)$ to each component of $\mathbf{U_X}$:

$$\mathbf{Z} := (Z_1, \ldots, Z_d)^\top := (\Phi^{-1}(U_{Y_1|\mathbf{X}}), \ldots, \Phi^{-1}(U_{Y_d|\mathbf{X}}))^\top.$$

Lemma 3.4. *The joint density of $\boldsymbol{Z}$ given $\boldsymbol{X} = \boldsymbol{x}$ is*

$$f_{\boldsymbol{Z}|\boldsymbol{X}}(z_1, \ldots, z_d|\boldsymbol{x}) = \left(\prod_{j=1}^{d} \phi(z_j)\right) c_{\boldsymbol{U}|\boldsymbol{X}}(\Phi(z_1), \ldots, \Phi(z_d)|\boldsymbol{x}), \forall x \in \mathbb{R}^p. \tag{3.6}$$

Proof. (Lemma 3.4)
Here we use Theorem 2.2 for conditional densities (by applying the definition of conditional densities it is easy to see that this theorem also holds for those),

$$f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) = f_{\mathbf{Z}|\mathbf{x}}(z_1, \ldots, z_d|\mathbf{x}) = |det(dT^{-1}(\mathbf{z}))| f_{\mathbf{U}|\mathbf{X}}(T^{-1}(\mathbf{z})|\mathbf{x}). \tag{3.7}$$

We have in our case the following inverse transformation

$$U_{Y_j|\mathbf{X}} = T_j^{-1}(Z_1, \ldots, Z_d) = \Phi(Z_j), \text{ for } j \in \{1, \ldots, d\},$$

therefore:

$$dT^{-1}(\mathbf{z}) = \begin{pmatrix} \frac{\partial T_1^{-1}}{\partial z_1}(\mathbf{z}) & \ldots & \frac{\partial T_1^{-1}}{\partial z_d}(\mathbf{z}) \\ \vdots & \ddots & \vdots \\ \frac{\partial T_d^{-1}}{\partial z_1}(\mathbf{z}) & \ldots & \frac{\partial T_d^{-1}}{\partial z_d}(\mathbf{z}) \end{pmatrix} = \begin{pmatrix} \phi(z_1) & 0 & \ldots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & \phi(z_d) \end{pmatrix},$$

$$\Rightarrow \det(dT^{-1}(\mathbf{z})) = \left(\prod_{j=1}^{d} \phi(z_j)\right).$$

By using 3.3, (that $f_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x}) = c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$) and inserting the determinant in equation (3.7) the statement is proven. □

3.3. Partial copulas and conditional copulas with random condition

We now define of a partial copula, which is the joint (unconditional) distribution function of $\mathbf{U}_{\mathbf{X}}$.

Definition 3.2. *(Partial copula)*
Let

$$\boldsymbol{U}_{\boldsymbol{X}} := (F_{Y_1|\boldsymbol{X}}(Y_1|\boldsymbol{X}), \ldots, F_{Y_d|\boldsymbol{X}}(Y_d|\boldsymbol{X}))^\top.$$

The partial copula is the joint distribution function of $\boldsymbol{U}_{\boldsymbol{X}}$ *and is denoted by* $C^P_{\boldsymbol{U}|\boldsymbol{X}}(u_1, \ldots, u_d)$.

This definition has been introduced and mentioned by Bergsma (2011),Gijbels et al. (2015), Spanhel and Kurz (2015a) and Spanhel and Kurz (2015b). To provide another interpretation of partial copulas, we have at first to define conditional copulas with random conditions.

Since we know that a conditional copula (Definition 3.1) is a conditional distribution function, $C_{U|X}(u_1, \ldots, u_d|\mathbf{X})$ can be defined in the same way as $F_{Y|\mathbf{X}}(\mathbf{y}|\mathbf{X})$ in Table 3.1. One can find a summary of the conditional copulas in Table 3.2.

In Table 3.2 we use the notation of the inverse of the conditional distribution $F^{-1}_{Y_j|\mathbf{X}}(u_j|\mathbf{x})$ for $j \in \{1, \ldots, d\}$ and $\mathbf{x} \in \mathbb{R}^p$ which is defined by

$$F^{-1}_{Y_j|\mathbf{X}}(F_{Y_j|\mathbf{X}}(y_j|\mathbf{x})|\mathbf{x}) = y_j, \ \forall y_j \in \mathbb{R} \text{ and } F_{Y_j|\mathbf{X}}(F^{-1}_{Y_j|\mathbf{X}}(u_j|\mathbf{x})|\mathbf{x}) = u_j, \ \forall u_j \in [0, 1].$$

Note that we therefore assume that:

$$F^{-1}_{Y_j|\mathbf{X}}(F_{Y_j|\mathbf{X}}(Y_j|\mathbf{X})|\mathbf{X}) = Y_j \text{ for } j \in \{1, \ldots, d\} \text{ a.s. } . \tag{3.8}$$

Table 3.2.: Summary of different copula situations.

	For $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$, $\mathbf{X} = (X_1, \dots, X_p)^\top$ and $\mathbf{y} = (y_1, \dots, y_d)^\top$, $\mathbf{x} = (x_1, \dots, x_p)^\top$
Usual cond. copula	
Notation/Definition	$C_{\mathbf{U}\|\mathbf{X}}(u_1, \dots, u_d\|\mathbf{x}) \quad := P(F_{Y_1\|\mathbf{X}}(Y_1\|\mathbf{x}) \leq u_1, \dots, F_{Y_d\|\mathbf{X}}(Y_d\|\mathbf{x}) \leq u_d\|\mathbf{X} = \mathbf{x})$
1st formulation:	$= F_{\mathbf{Y}\|\mathbf{X}}(F^{-1}_{Y_1\|\mathbf{X}}(u_1\|\mathbf{x}), \dots, F^{-1}_{Y_d\|\mathbf{X}}(u_1\|\mathbf{x})\|\mathbf{x})$
2nd formulation:	$= \frac{\int_{-\infty}^{F^{-1}_{Y_d\|\mathbf{X}}(u_1\|\mathbf{x})} \dots \int_{-\infty}^{F^{-1}_{Y_1\|\mathbf{X}}(u_1\|\mathbf{x})} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1,\dots,\tilde{y}_d,\mathbf{x}) d\tilde{y}_1 \dots d\tilde{y}_d}{f_{\mathbf{X}}(\mathbf{x})}$
3rd formulation:	$= P(F_{Y_1\|\mathbf{X}}(Y_1\|\mathbf{X}) \leq u_1, \dots, F_{Y_d\|\mathbf{X}}(Y_d\|\mathbf{X}) \leq u_d\|\mathbf{X} = \mathbf{x})$ (nonrandom function of $(\mathbf{u}, \mathbf{x})$)
Cond. copula with random condition	
Notation/Definition:	$C_{\mathbf{U}\|\mathbf{X}}(u_1, \dots, u_d\|\mathbf{X}) \quad := F_{\mathbf{Y}\|\mathbf{X}}(F^{-1}_{Y_1\|\mathbf{X}}(u_1\|\mathbf{X}), \dots, F^{-1}_{Y_d\|\mathbf{X}}(u_1\|\mathbf{X})\|\mathbf{X})$
1st formulation:	$= \frac{\int_{-\infty}^{F^{-1}_{Y_d\|\mathbf{X}}(u_1\|\mathbf{X})} \dots \int_{-\infty}^{F^{-1}_{Y_1\|\mathbf{X}}(u_1\|\mathbf{X})} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1,\dots,\tilde{y}_d,\mathbf{X}) d\tilde{y}_1 \dots d\tilde{y}_d}{f_{\mathbf{X}}(\mathbf{X})}$ (random function of $\mathbf{u}$)
Notation/Definition:	$P(F_{Y_1\|\mathbf{X}}(Y_1\|\mathbf{X}) \leq u_1, \dots, F_{Y_d\|X}(Y_d\|\mathbf{X}) \leq u_d\|\mathbf{X})$ $:= E(1_{\{F_{Y_1\|\mathbf{X}}(Y_1\|\mathbf{X}) \leq u_1, \dots, F_{Y_d\|X}(Y_d\|\mathbf{X}) \leq u_d\}}\|\mathbf{X})$ (random function of $\mathbf{u}$)
Partial copula	
Notation/Definition:	$C^P_{\mathbf{U}\|\mathbf{X}}(u_1, \dots, u_d) \quad := P(F_{Y_1\|\mathbf{X}}(Y_1\|\mathbf{X}) \leq u_1, \dots, F_{Y_d\|X}(Y_d\|\mathbf{X}) \leq u_d)$ (nonrandom function of $\mathbf{u}$)

The following lemma provides us a link between the partial copula $C^P_{\mathbf{U}|\mathbf{X}}(u_1,\dots,u_d)$ and the random variable $C_{\mathbf{U}|\mathbf{X}}(u_1,\dots,u_d|\mathbf{X})$:

Lemma 3.5. *The partial copula can be written as the expected conditional copula with random condition:*

$$C^P_{\boldsymbol{U}|\boldsymbol{X}}(u_1,\dots,u_d) = \mathrm{E}(C_{\boldsymbol{U}|\boldsymbol{X}}(u_1,\dots,u_d|\boldsymbol{X})).$$

Proof. (Lemma 3.5)
By definition of the expectation it hold that:

$$\mathrm{E}\left[C_{\mathbf{U}|\mathbf{X}}\ (u_1,\dots,u_d|\mathbf{X})\right] = \int_{\mathbb{R}^p} C_{\mathbf{U}|\mathbf{X}}(u_1,\dots,u_d|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

By using Lemma 3.3 this is equal to

$$= \int_{\mathbb{R}^p} P(F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \le u_1,\dots,F_{Y_d|\mathbf{X}}(Y_d|\mathbf{X}) \le u_d|\mathbf{X}=\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Remember that $\mathbf{U}_{\mathbf{X}} := (F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}),\dots,F_{Y_d|\mathbf{X}}(Y_d|\mathbf{X}))^\top$ therefore:

$$\begin{aligned}
&= \int_{\mathbb{R}^p} \left\{ \int_{-\infty}^{u_d} \dots \int_{-\infty}^{u_1} \frac{f_{\mathbf{U}_{\mathbf{X}},\mathbf{X}}(t_1,\dots,t_d,\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} dt_1 \dots dt_d \right\} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^p} \int_{-\infty}^{u_d} \dots \int_{-\infty}^{u_1} f_{\mathbf{U}_{\mathbf{X}},\mathbf{X}}(t_1,\dots,t_d,\mathbf{x}) dt_1 \dots dt_d d\mathbf{x} \\
&= \int_{-\infty}^{u_d} \dots \int_{-\infty}^{u_1} f_{\mathbf{U}_{\mathbf{X}}}(t_1,\dots,t_d) dt_1 \dots dt_d \\
&= P(F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \le u_1,\dots,F_{Y_d|\mathbf{X}}(Y_d|\mathbf{X}) \le u_d) = C^P_{\mathbf{U}|\mathbf{X}}(u_1,\dots,u_d).
\end{aligned}$$

□

Note that there is also a link between the notations $C_{\mathbf{U}|\mathbf{X}}(u_1,\dots,u_d|\mathbf{X})$ and $P(F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \le u_1,\dots,F_{Y_d|X}(Y_d|\mathbf{X}) \le u_d|\mathbf{X})$.

Lemma 3.6. *By using the notations from Table 3.2 and assuming that the densities $f_{\boldsymbol{Y},\boldsymbol{X}}(\cdot,\cdot)$, $f_{\boldsymbol{X}}(\cdot)$ and the inverse conditional distribution functions $F^{-1}_{Y_j|\boldsymbol{X}}(u_j|\boldsymbol{x})$ for $j \in \{1,\dots,d\}$ and $\boldsymbol{x} \in \mathbb{R}^p$ exist we can state:*

$$C_{\boldsymbol{U}|\boldsymbol{X}}(u_1,\dots,u_d|\boldsymbol{X}) = P(F_{Y_1|\boldsymbol{X}}(Y_1|\boldsymbol{X}) \le u_1,\dots,F_{Y_d|X}(Y_d|\boldsymbol{X}) \le u_d|\boldsymbol{X}) \ \textit{a.s.}\ .$$

Proof. (Lemma 3.6)
This proof is similar to the one of Theorem 3.1. By definition

$$\begin{aligned}
&P(F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \le u_1,\dots,F_{Y_d|X}(Y_d|\mathbf{X}) \le u_d|\mathbf{X}) \\
&\quad := E(1_{\{F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \le u_1,\dots,F_{Y_d|X}(Y_d|\mathbf{X}) \le u_d\}}|\mathbf{X}).
\end{aligned}$$

We have to show the following:

(i) $C_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{X})$ is measurable w.r.t. the σ-algebra $\sigma(\mathbf{X})$.
(ii) For every bounded random variable Z which is measurable w.r.t. $\sigma(\mathbf{X})$, we have $E[Z \cdot C_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{X})] = E[Z \cdot 1_{\{F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \le u_1, \dots, F_{Y_d|X}(Y_d|\mathbf{X}) \le u_d\}}]$.

Statement (i) is clear because we assume that $C_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{X})$ is a continuous function in X, so only statement (ii) is remaining. We can rewrite statement (ii) as follows:

$$
\begin{aligned}
(ii) \iff & E[g(\mathbf{X}) \cdot C_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{X})] = E[g(\mathbf{X}) \cdot 1_{\{F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \le u_1, \dots, F_{Y_d|X}(Y_d|\mathbf{X}) \le u_d\}}] \\
\iff & \int_{\mathbb{R}^p} g(\mathbf{x}) C_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
& = E[g(\mathbf{X}) \cdot 1_{\{Y_1 \le F^{-1}_{Y_1|\mathbf{X}}(u_1|\mathbf{X}), \dots, Y_d \le F^{-1}_{Y_d|\mathbf{X}}(u_d|\mathbf{X})\}}] \\
& = \int_{\mathbb{R}^p} \int_{\mathbb{R}^d} g(\mathbf{x}) \cdot 1_{\{\tilde{y}_1 \le F^{-1}_{Y_1|\mathbf{X}}(u_1|\mathbf{x}), \dots, \tilde{y}_d \le F^{-1}_{Y_d|\mathbf{X}}(u_d|\mathbf{x})\}} f_{\mathbf{Y},\mathbf{X}}(\tilde{\mathbf{y}}, \mathbf{x}) d\tilde{\mathbf{y}} d\mathbf{x} \\
\iff & \int_{\mathbb{R}^p} g(\mathbf{x}) \int_{-\infty}^{F^{-1}_{Y_d|\mathbf{X}}(u_1|\mathbf{x})} \dots \int_{-\infty}^{F^{-1}_{Y_1|\mathbf{X}}(u_1|\mathbf{x})} f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1, \dots, \tilde{y}_d, \mathbf{x}) d\tilde{y}_1 \dots d\tilde{y}_d d\mathbf{x} \\
& = \int_{\mathbb{R}^p} \int_{-\infty}^{F^{-1}_{Y_d|\mathbf{X}}(u_1|\mathbf{x})} \dots \int_{-\infty}^{F^{-1}_{Y_1|\mathbf{X}}(u_1|\mathbf{x})} g(\mathbf{x}) f_{\mathbf{Y},\mathbf{X}}(\tilde{y}_1, \dots, \tilde{y}_d, \mathbf{x}) d\tilde{y}_1 \dots d\tilde{y}_d d\mathbf{x},
\end{aligned}
$$

where $g(\mathbf{X}) = Z$ (see proof of Theorem 3.1), for the second equality we used equation (3.8) and in the last step we applied the 2nd formulation of the condition copula distribution from Table 3.2. □

We illustrate the different copulas situations by an example, where we calculate the partial and conditional copula directly.

Example 3.1. *Assume that we have the following model:*

$$
\begin{gathered}
Y_1 = \alpha_0^1 + \alpha_1^1 X + \epsilon_1, \\
Y_2 = \alpha_0^2 + \alpha_1^2 X + \epsilon_2, \\
\text{with } X \sim N(0,1),\ \epsilon_1 \sim N(0,1),\ \epsilon_2 \sim N(0,1), \\
\epsilon_1 \perp X,\ \epsilon_2 \perp X, \\
\text{and } \alpha_0^1, \alpha_1^1, \alpha_0^2, \alpha_1^2 \in \mathbb{R}
\end{gathered}
$$

Define σ_{12} as the covariance of ϵ_1 and ϵ_2,

$$\sigma_{12} := \mathrm{Cov}(\epsilon_1, \epsilon_2) = \mathrm{Cor}(\epsilon_1, \epsilon_2).$$

Calculation of joint distributions:
Since every linear combination of Y_1, Y_2 and X is normal, it follows that the joint distributions of (Y_1, X), (Y_2, X) and (Y_1, Y_2, X) are multivariate normal distributions, which we can determine by calculating the means and covariances.

By calculating those we get:

$$
\begin{aligned}
(Y_1, X) &\sim N_2\left(\begin{pmatrix}\alpha_0^1\\ 0\end{pmatrix}, \begin{pmatrix}(\alpha_1^1)^2+1 & \alpha_1^1\\ \alpha_1^1 & 1\end{pmatrix}\right),\\
(Y_2, X) &\sim N_2\left(\begin{pmatrix}\alpha_0^2\\ 0\end{pmatrix}, \begin{pmatrix}(\alpha_1^2)^2+1 & \alpha_1^2\\ \alpha_1^2 & 1\end{pmatrix}\right),\\
(Y_1, Y_2, X) &\sim N_3\left(\begin{pmatrix}\alpha_0^1\\ \alpha_0^2\\ 0\end{pmatrix}, \begin{pmatrix}(\alpha_1^1)^2+1 & \alpha_1^2\alpha_1^1+\sigma_{12} & \alpha_1^1\\ \alpha_1^2\alpha_1^1+\sigma_{12} & (\alpha_1^2)^2+1 & \alpha_1^2\\ \alpha_1^1 & \alpha_1^2 & 1\end{pmatrix}\right).
\end{aligned}
$$

Therefore the joint density of (Y_1, Y_2, X) *is:*

$$
f_{\boldsymbol{Y},X}(y_1, y_2, x) = \frac{1}{\sqrt{(2\pi)^3|\Sigma_{\boldsymbol{Y},X}|}} exp\left(-0.5\begin{pmatrix}y_1-\alpha_0^1\\ y_2-\alpha_0^2\\ x\end{pmatrix}^\top \left(\Sigma_{\boldsymbol{Y},X}\right)^{-1}\begin{pmatrix}y_1-\alpha_0^1\\ y_2-\alpha_0^2\\ x\end{pmatrix}\right), \tag{3.9}
$$

where

$$
\Sigma_{\boldsymbol{Y},X} := \begin{pmatrix}(\alpha_1^1)^2+1 & \alpha_1^2\alpha_1^1+\sigma_{12} & \alpha_1^1\\ \alpha_1^2\alpha_1^1+\sigma_{12} & (\alpha_1^2)^2+1 & \alpha_1^2\\ \alpha_1^1 & \alpha_1^2 & 1\end{pmatrix}, \tag{3.10}
$$

$$
\Rightarrow (\Sigma_{\boldsymbol{Y},X})^{-1} = \begin{pmatrix}\frac{1}{1-\sigma_{12}^2} & \frac{-\sigma_{12}}{1-\sigma_{12}^2} & \frac{\alpha_1^2\sigma_{12}-\alpha_1^1}{1-\sigma_{12}^2}\\ \frac{-\sigma_{12}}{1-\sigma_{12}^2} & \frac{1}{1-\sigma_{12}^2} & \frac{\alpha_1^1\sigma_{12}-\alpha_1^2}{1-\sigma_{12}^2}\\ \frac{\alpha_1^2\sigma_{12}-\alpha_1^1}{1-\sigma_{12}^2} & \frac{\alpha_1^1\sigma_{12}-\alpha_1^2}{1-\sigma_{12}^2} & \frac{(\alpha_1^1)^2-2\alpha_1^1\alpha_1^2\sigma_{12}}{1-\sigma_{12}^2}+1\end{pmatrix}, \tag{3.11}
$$

$$
|\Sigma_{\boldsymbol{Y},X}| = 1-\sigma_{12}^2, \tag{3.12}
$$

$$
\begin{aligned}
\Rightarrow f_{\boldsymbol{Y},X}(y_1, y_2, x) = &\frac{1}{\sqrt{(2\pi)^3(1-\sigma_{12}^2)}}\exp\Bigg(\frac{0.5}{\sigma_{12}^2-1}\\
&\cdot\big\{-2\sigma_{12}(y_2-\alpha_1^1)(y_1-\alpha_0^1)+(y_2-\alpha_1^1)^2\\
&\quad+(y_1-\alpha_0^1)^2+[2\alpha_1^1\sigma_{12}-2\alpha_1^2](y_2-\alpha_1^1)x\\
&\quad+[2\alpha_1^2\sigma_{12}-2\alpha_1^1](y_1-\alpha_0^1)x\\
&\quad+[(\alpha_1^1)^2+(\alpha_1^2)^2-\sigma_{12}^2-2\alpha_1^1\alpha_1^2\sigma_{12}+1]x^2\ \big\}\Bigg).
\end{aligned}
$$

Calculation of conditional distributions:
By using the formula of the conditional multivariate normal distributions we can also calculate the following conditional distributions:

$$
\begin{aligned}
(Y_1|X=x) &\sim N\left(\alpha_0^1+\alpha_1^1x, 1\right),\\
(Y_2|X=x) &\sim N\left(\alpha_0^2+\alpha_1^2x, 1\right),\\
(Y_1, Y_2|X=x) &\sim N_2\left(\begin{pmatrix}\alpha_0^1+\alpha_1^1x\\ \alpha_0^2+\alpha_1^2x\end{pmatrix}, \begin{pmatrix}1 & \sigma_{12}\\ \sigma_{12} & 1\end{pmatrix}\right),
\end{aligned}
$$

$$
\begin{aligned}
F_{\boldsymbol{Y}|X}(y_1, y_2|x) &= P(Y_1 < y_1, Y_2 < y_2|X = x) \\
&= \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} f_{\boldsymbol{Y}|X}(\tilde{y}_1, \tilde{y}_2|x) d\tilde{y}_1 d\tilde{y}_2 \\
&= \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} \frac{1}{\sqrt{(2\pi)^2 |\Sigma_{\boldsymbol{Y}|X}|}} \\
&\qquad exp\left(-0.5 \begin{pmatrix} \tilde{y}_1 - \mu_1 \\ \tilde{y}_2 - \mu_2 \end{pmatrix}^\top (\Sigma_{\boldsymbol{Y}|X})^{-1} \begin{pmatrix} \tilde{y}_1 - \mu_1 \\ \tilde{y}_2 - \mu_2 \end{pmatrix}\right) d\tilde{y}_1 d\tilde{y}_2,
\end{aligned} \tag{3.13}
$$

where

$$
\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} := \begin{pmatrix} \alpha_0^1 + \alpha_1^1 x \\ \alpha_0^2 + \alpha_1^2 x \end{pmatrix} \text{ and } \Sigma_{\boldsymbol{Y}|X} := \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}.
$$

After calculating the inverse matrix, the determinate and expanding the matrix product we get:

$$
\begin{aligned}
F_{\boldsymbol{Y}|X}(y_1, y_2|x) = \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} \frac{1}{2\pi\sqrt{1-\sigma_{12}^2}} \exp\Bigg(&- \frac{1}{2(1-\sigma_{12}^2)} \\
&\cdot \Big\{ (\tilde{y}_1 - \alpha_0^1 + \alpha_1^1 x)^2 + (\tilde{y}_2 - \alpha_0^2 + \alpha_1^2 x)^2 \\
&- 2\rho(\tilde{y}_1 - \alpha_0^1 + \alpha_1^1 x)(\tilde{y}_2 - \alpha_0^2 + \alpha_1^2 x) \Big\} \Bigg) d\tilde{y}_1 d\tilde{y}_2.
\end{aligned}
$$

Calculation of the conditional copula:

We know by the definition of the conditional copula that:

$$
C_{\boldsymbol{U}|X}(u_1, u_2|x) = F_{\boldsymbol{Y}|X}(F_{Y_1|X}^{-1}(u_1|x), F_{Y_2|X}^{-1}(u_2|x)|x),
$$

so by applying equation (3.13), the conditional copula in our example is:

$$
\begin{aligned}
C_{\boldsymbol{U}|X}(u_1, u_2|x) = &\int_{-\infty}^{F_{Y_2|X}^{-1}(u_2|x)} \int_{-\infty}^{F_{Y_1|X}^{-1}(u_1|x)} \frac{1}{\sqrt{(2\pi)^2 |\Sigma_{\boldsymbol{Y}|X}|}} \\
&\cdot \exp\left(-0.5 \begin{pmatrix} \tilde{y}_1 - \mu_1 \\ \tilde{y}_2 - \mu_2 \end{pmatrix}^\top (\Sigma_{\boldsymbol{Y}|X})^{-1} \begin{pmatrix} \tilde{y}_1 - \mu_1 \\ \tilde{y}_2 - \mu_2 \end{pmatrix}\right) d\tilde{y}_1 d\tilde{y}_2.
\end{aligned}
$$

By using that

$$
\begin{aligned}
F_{Y_1|X}^{-1}(u_1|x) &= (\alpha_0^1 + \alpha_1^1 x) + \Phi^{-1}(u_1), \\
F_{Y_2|X}^{-1}(u_2|x) &= (\alpha_0^2 + \alpha_1^2 x) + \Phi^{-1}(u_2),
\end{aligned}
$$

and then substitution with $\bar{y}_1 = \tilde{y}_1 - (\alpha_0^1 + \alpha_1^1 x)$ *and* $\bar{y}_2 = \tilde{y}_2 - (\alpha_0^2 + \alpha_1^2 x)$ *we get:*

$$C_{\boldsymbol{U}|X}(u_1,u_2|x) = \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_1)} \frac{1}{\sqrt{(2\pi)^2|\Sigma_{\boldsymbol{Y}|X}|}} \cdot \exp\left(-0.5 \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}^\top \left(\Sigma_{\boldsymbol{Y}|X}\right)^{-1} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}\right) d\bar{y}_1 d\bar{y}_2.$$

Using that

$$\Sigma_{\boldsymbol{Y}|X} = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix},$$

we can observe that $C_{\boldsymbol{Y}|X}(u_1,u_2|x)$ is the Gaussian bivariate copula with correlation parameter σ_{12}.

In our case $C_{\boldsymbol{U}|X}(u_1,u_2|x)$ is independent of x, so $C_{\boldsymbol{U}|X}(u_1,u_2|X)$ is a degenerate constant random variable.

Calculation of the partial copula:
By applying Lemma 3.5 we can easily see that the partial copula $C^p_{\boldsymbol{U}|X}(u_1,u_2)$ is also the Gaussian bivariate copula with correlation parameter σ_{12}.

However to show the basic idea of how to calculate a partial copula and also to check our result, we will calculate the partial copula by hand. By definition the partial copula is:

$$C^p_{\boldsymbol{U}|X}(u_1,u_2) = P(F_{Y_1|X}(Y_1|X) \le u_1, F_{Y_2|X}(Y_2|X) \le u_2).$$

The main idea of calculating the partial copula is to use Theorem 2.2 (densities of transformed random variables) and then integrate the third variable out.
Let

$$\begin{aligned} Z_1 &= T_1(Y_1,Y_2,X) = F_{Y_1|X}(Y_1|X) = \Phi(Y_1 - (\alpha_0^1 + \alpha_1^1 X)), \\ Z_1 &= T_2(Y_1,Y_2,X) = F_{Y_2|X}(Y_2|X) = \Phi(Y_2 - (\alpha_0^2 + \alpha_1^2 X)), \\ Z_3 &= T_3(Y_1,Y_2,X) = X, \end{aligned}$$

with

$$\begin{aligned} Y_1 &= T_1^{-1}(Z_1,Z_2,Z_3) = F^{-1}_{Y_1|X}(Z_1|Z_3) = (\alpha_0^1 + \alpha_1^1 Z_3) + \Phi^{-1}(Z_1), \\ Y_2 &= T_2^{-1}(Z_1,Z_2,Z_3) = F^{-1}_{Y_2|X}(Z_2|Z_3) = (\alpha_0^2 + \alpha_1^2 Z_3) + \Phi^{-1}(Z_2), \\ X &= T_3^{-1}(Z_1,Z_2,Z_3) = Z_3, \end{aligned}$$

and

$$dT^{-1}(\boldsymbol{z}) = \begin{pmatrix} \frac{\partial F^{-1}_{Y_1|X}(Z_1|Z_3)}{\partial z_1} & 0 & \frac{\partial F^{-1}_{Y_1|X}(Z_1|Z_3)}{\partial z_3} \\ 0 & \frac{\partial F^{-1}_{Y_2|X}(Z_2|Z_3)}{\partial z_2} & \frac{\partial F^{-1}_{Y_2|X}(Z_2|Z_3)}{\partial z_2} \\ 0 & 0 & 1 \end{pmatrix},$$

$$det[dT^{-1}(\boldsymbol{z})] = \frac{\partial F^{-1}_{Y_1|X}(z_1|z_3)}{\partial z_1}\frac{\partial F^{-1}_{Y_2|X}(z_2|z_3)}{\partial z_2} = \frac{\partial \Phi^{-1}(z_1)}{\partial z_1}\frac{\partial \Phi^{-1}(z_2)}{\partial z_2},$$

By Theorem 2.2 we get then that:

$$\begin{aligned} f_{Z_1,Z_2,Z_3} &= |det[dT^{-1}(\boldsymbol{z})]| f_{\boldsymbol{Y},X}(F^{-1}_{Y_1|X}(z_1|z_3), F^{-1}_{Y_2|X}(z_2|z_3), z_3), \\ &= \frac{\partial \Phi^{-1}(z_1)}{\partial z_1}\frac{\partial \Phi^{-1}(z_2)}{\partial z_2} \\ &\quad \cdot f_{\boldsymbol{Y},X}\left((\alpha_0^1 + \alpha_1^1 z_3) + \Phi^{-1}(z_1), (\alpha_0^2 + \alpha_1^2 z_3) + \Phi^{-1}(z_2), z_3\right). \end{aligned}$$

Integrating the z_3 variable yields the partial copula:

$$\begin{aligned} C^p_{\boldsymbol{U}|X}(u_1, u_2) &= F_{Z_1,Z_2}(u_1, u_2) \\ &= \int_{-\infty}^{u_2}\int_{-\infty}^{u_1}\int_{\mathbb{R}} \frac{\partial \Phi^{-1}(z_1)}{\partial z_1}\frac{\partial \Phi^{-1}(z_2)}{\partial z_2} \\ &\quad \cdot f_{\boldsymbol{Y},X}\left((\alpha_0^1 + \alpha_1^1 z_3) + \Phi^{-1}(z_1), (\alpha_0^2 + \alpha_1^2 z_3) + \Phi^{-1}(z_2), z_3\right) dz_3 dz_1 dz_2. \end{aligned}$$

Using equation (3.9), plugging in $\Sigma_{\boldsymbol{Y},X}$ from (3.10), (3.11), and (3.12), simplifying, and finally integrating z_3 out, yields:

$$\begin{aligned} C^p_{\boldsymbol{U}|X}(u_1, u_2) &= \int_{-\infty}^{u_2}\int_{-\infty}^{u_1} \frac{\partial \Phi^{-1}(z_1)}{\partial z_1}\frac{\partial \Phi^{-1}(z_2)}{\partial z_2} \\ &\qquad \cdot \frac{\exp\left(-0.5\frac{\Phi^{-1}(z_1)^2 - 2\sigma_{12}\Phi^{-1}(z_1)\Phi^{-1}(z_2) + \Phi^{-1}(z_2)^2}{(1-\sigma_{12}^2)}\right)}{\sqrt{(2\pi)^2(1-\sigma_{12}^2)}} dz_1 dz_2 \\ &= \int_{-\infty}^{u_2}\int_{-\infty}^{u_1} \frac{\partial \Phi^{-1}(z_1)}{\partial z_1}\frac{\partial \Phi^{-1}(z_2)}{\partial z_2} \phi_{2,\sigma_{12}}(\Phi^{-1}(z_1), \Phi^{-1}(z_2)) dz_1 dz_2, \end{aligned}$$

where $\phi_{2,\sigma_{12}}(\cdot,\cdot)$ is the density of a two dimensional normal distribution with mean $(0,0)^\top$ and covariance matrix

$$\Sigma_{\tilde{y}_1,\tilde{y}_2} = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}.$$

By the inverse function theorem we know that

$$\frac{\partial \Phi^{-1}(z)}{\partial z} = \frac{1}{\phi(\Phi^{-1}(z))}.$$

Applying this leads then to

$$C^p_{\boldsymbol{U}|X}(u_1, u_2) = \int_{-\infty}^{u_2}\int_{-\infty}^{u_1} \frac{\phi_{2,\sigma_{12}}(\Phi^{-1}(z_1), \Phi^{-1}(z_2))}{\phi(\Phi^{-1}(z_1))\phi(\Phi^{-1}(z_2))} dz_1 dz_2.$$

So we proved that in our case the partial copula is a Gaussian copula with correlation parameter σ_{12}.

4. Kernel estimation of a conditional copula density

In this chapter we will at first present a concept for estimation of a conditional copula density. We will then continue with a section, where we discuss the estimation of conditional densities using a method based on the a locally linear regression estimation mentioned in Section 2.5.

4.1. Conditional copula estimation based on transformation

Our goal is to estimate the conditional copula density

$$c_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{x}) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} P(F_{Y_1|\mathbf{X}}(Y_1|\mathbf{X}) \leq u_1, \dots, F_{Y_d|\mathbf{X}}(Y_d|\mathbf{X}) \leq u_d|\mathbf{X} = \mathbf{x}),$$

of a corresponding random vector $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. Suppose that we observe n *i.i.d.* copies of $(\mathbf{Y}, \mathbf{X})$,

$$(\mathbf{Y}_i, \mathbf{X}_i) = (Y_{i,1}, \dots Y_{i,d}, X_{i,1}, \dots, X_{i,p})^\top, \text{ for } i \in \{1, \dots, n\},$$

following the $(d+p)$-dimensional probability density function $f_{\mathbf{Y},\mathbf{X}}(\mathbf{y}, \mathbf{x})$. We denote the associated samples by

$$(\mathbf{y}_i, \mathbf{x}_i) = (y_{i,1}, \dots y_{i,d}, x_{i,1}, \dots, x_{i,p})^\top, \text{ for } i \in \{1, \dots, n\}.$$

Using estimated marginal conditional distributions $\hat{F}_{Y_j|\mathbf{X}}(\cdot|\cdot)$, for $j \in \{1, \dots, d\}$ we can transform our observations into pseudo conditional copula observations

$$(u_{i,1}, \dots u_{i,d}, x_{i,1}, \dots, x_{i,p})^\top := (\hat{F}_{Y_1|\mathbf{X}}(y_{i,1}|\mathbf{x}_i), \dots, \hat{F}_{Y_d|\mathbf{X}}(y_{i,d}|\mathbf{x}_i), x_{i,1}, \dots, x_{i,p})^\top,$$

for $i \in \{1, \dots, n\}$.

It is important to be aware of the fact that the $(u_{i,1}, \dots u_{i,d})^\top$ are not true samples from the conditional copula $C_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{x}_i)$, but estimated conditional copula samples.

At this point we have observations $(u_{i,1}, \dots u_{i,d}, x_{i,1}, \dots, x_{i,p})^\top$, where $i \in \{1, \dots, n\}$, to estimate the conditional copula density $c_{\mathbf{U}|\mathbf{X}}(u_1, \dots, u_d|\mathbf{x})$. Notice that by Lemma 3.3 the functions $\hat{F}_{Y_j|\mathbf{X}}(\cdot|\cdot)$ do not have to depend on the specific $\mathbf{x}$.

By using Lemma 3.4, we can write the conditional copula density in the following way

$$c_{\mathbf{U}|\mathbf{X}}(u_1, \ldots, u_d|\mathbf{x}) = \frac{f_{\mathbf{Z}|\mathbf{X}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)|\mathbf{x})}{\prod_{j=1}^{d} \phi(\Phi^{-1}(u_j))}, \tag{4.1}$$

where $\Phi^{-1}(\cdot)$ and $\phi(\cdot)$ denote the quantile and density functions of an univariate standard normal random variable. By estimating the density function in the enumerator of equation (4.1) by a nonparametric estimator $\hat{f}_{\mathbf{Z}|\mathbf{X}}(z_1, \ldots, z_d|\mathbf{x})$ using the samples

$$(\mathbf{z}_i, \mathbf{x}_i) = (z_{i,1}, \ldots z_{i,d}, x_{i,1}, \ldots, x_{i,p})^\top := (\Phi^{-1}(u_{i,1}), \ldots, \Phi^{-1}(u_{i,d}), x_{i,1}, \ldots, x_{i,p})^\top,$$

for $i = 1 \ldots n$, we get an estimator for the conditional copula density

$$\hat{c}_{\mathbf{U}|\mathbf{X}}(u_1, \ldots, u_d|\mathbf{x}) = \frac{\hat{f}_{\mathbf{Z}|\mathbf{X}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)|\mathbf{x})}{\prod_{j=1}^{d} \phi(\Phi^{-1}(u_j))}. \tag{4.2}$$

For the estimator $\hat{f}_{\mathbf{Z}|\mathbf{X}}(z_1, \ldots, z_d|\mathbf{x})$ we use a locally linear regression approach, which will be the subject of the next section.

4.2. Estimation of conditional densities using local linear regression

In this section, we will provide a method for estimating the conditional density function $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ of a random vector $\mathbf{Z}$ given $\mathbf{X} = \mathbf{x}$. Suppose that we observe n *i.i.d.* copies of $(\mathbf{Z}, \mathbf{X})$,

$$(\mathbf{Z}_i^\top, \mathbf{X}_i^\top)^\top = (Z_{i,1}, \ldots Z_{i,d}, X_{i,1}, \ldots, X_{i,p})^\top \text{ } i.i.d. \text{ for } i \in \{1, \ldots, n\},$$

following the $(d+p)$-dimensional probability density function $f_{\mathbf{Z},\mathbf{X}}(\mathbf{z}, \mathbf{x})$. We denote the corresponding realizations by

$$(\mathbf{z}_i^\top, \mathbf{x}_i^\top)^\top = (z_{i,1}, \ldots z_{i,d}, x_{i,1}, \ldots, x_{i,p})^\top, \text{ for } i \in \{1, \ldots, n\}.$$

The following approach is based on Fan et al. (1996), where they estimated the univariate conditional density and their derivative by a local quadratic regression method. Because it is not in our interest to estimate the derivative, a locally linear regression approach is sufficient. The next lemma will provide a motivation for the locally linear regression approach.

Lemma 4.1. *Let $f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})$ be the conditional density of $\boldsymbol{Z}$ given $\boldsymbol{X} = \boldsymbol{x}$, assumed smooth in both $\boldsymbol{z} \in \mathbb{R}^d$ and $\boldsymbol{x} \in \mathbb{R}^p$. Then:*

$$\lim_{H \to \boldsymbol{0}_{d\times d}} \mathrm{E}\left(K_H(\boldsymbol{Z} - \boldsymbol{z}) \middle| \boldsymbol{X} = \boldsymbol{x}\right) = f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}),$$

where $\boldsymbol{0}_{d\times d}$ denotes a $d \times d$ zero matrix.

Note here that the following proof will be very detailed. The reason for this is that in many further proofs we will use similar steps for the handling of the remainder in the Taylor expansions. We will omit the details in the further proofs, but show these here (see Equation 4.3) so that this proof is a detailed example for how it could be done.
The following proof will be too elaborate for the purpose of proving Lemma 4.1, because we will use some intermediary results in future proofs of the thesis.

Proof. (Lemma 4.1)
The construction of this proof is similar to the proof of the expected value of general multivariate kernel estimation for non-conditional densities in Wand (1992). It holds,

$$\begin{aligned}\mathrm{E}\Big(K_H(\mathbf{Z} - \mathbf{z}) \Big| \mathbf{X} = \mathbf{x}\Big) &= \int_{\mathbb{R}^d} K_H(\tilde{\mathbf{z}} - \mathbf{z}) f_{\mathbf{Z}|\mathbf{X}}(\tilde{\mathbf{z}}|\mathbf{x}) d\tilde{\mathbf{z}} \\ &= \frac{1}{\sqrt{|H|}} \int_{\mathbb{R}^d} \mathcal{K}(H^{-\frac{1}{2}}(\tilde{\mathbf{z}} - \mathbf{z})) f_{\mathbf{Z}|\mathbf{X}}(\tilde{\mathbf{z}}|\mathbf{x}) d\tilde{\mathbf{z}},\end{aligned}$$

using the change of variable $\tilde{\mathbf{z}} = (H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z})$

$$= \frac{1}{\sqrt{|H|}} \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) f_{\mathbf{Z}|\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z}|\mathbf{x}) |H^{\frac{1}{2}}| d\bar{\mathbf{z}}.$$

Since $|H| = |H^{\frac{1}{2}}H^{\frac{1}{2}}| = |H^{\frac{1}{2}}||H^{\frac{1}{2}}| = |H^{\frac{1}{2}}|^2$ it follows that $|H^{\frac{1}{2}}| = \sqrt{|H|}$.

$$= \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) f_{\mathbf{Z}|\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z}|\mathbf{x}) d\bar{\mathbf{z}}.$$

Using here that we assume that $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ has bounded continuous third order derivatives with respect to $\mathbf{z}$ (see Remark 3.1) and applying the d-dimensional Taylor expansion of $f_{\mathbf{Z}|\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z}|\mathbf{x})$ around $\mathbf{z} \in \mathbb{R}^d$ (see Theorem 2.1) leads to

$$\begin{aligned}&= \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) \left\{ f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + (H^{\frac{1}{2}}\bar{\mathbf{z}})^\top \nabla_{\mathbf{z}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \frac{1}{2}(H^{\frac{1}{2}}\bar{\mathbf{z}})^\top \mathcal{H}_{\mathbf{z}}(H^{\frac{1}{2}}\bar{\mathbf{z}}) \right\} d\bar{\mathbf{z}} \\ &+ \sum_{\substack{(k_1,k_2,k_3,r_1,r_2,r_3) \\ \in \{1,\dots,d\}^6}} O\left([H^{\frac{1}{2}}]_{k_1,r_1}[H^{\frac{1}{2}}]_{k_2,r_2}[H^{\frac{1}{2}}]_{k_2,r_3}\right) \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) O\left(\bar{z}_{r_1}\bar{z}_{r_2}\bar{z}_{r_3}\right) d\bar{\mathbf{z}},\end{aligned} \tag{4.3}$$

where $\nabla_{\mathbf{z}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) := \left(\frac{\partial}{\partial z_1} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}), \quad \dots \quad , \frac{\partial}{\partial z_d} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\right)^\top$ are the partial derivatives and $\mathcal{H}_{\mathbf{z}} \in \mathbb{R}^{d\times d}$ is the Hessian of $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ with respect to $\mathbf{z}$. Furthermore since we assume that $\int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})|\bar{z}_{r_1}\bar{z}_{r_2}\bar{z}_{r_3}|d\bar{\mathbf{z}} < \infty$ we get

$$\begin{aligned}
&\mathrm{E}\Big(K_H(\mathbf{Z}-\mathbf{z})\big|\mathbf{X}=\mathbf{x}\Big) \\
&\quad = \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) \left\{ f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + (H^{\frac{1}{2}}\bar{\mathbf{z}})^\top \nabla_{\mathbf{z}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \frac{1}{2}(H^{\frac{1}{2}}\bar{\mathbf{z}})^\top \mathcal{H}_{\mathbf{z}}(H^{\frac{1}{2}}\bar{\mathbf{z}}) \right\} d\bar{\mathbf{z}} \\
&\qquad + O(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))).
\end{aligned}$$

Here $\mathrm{tr}(\cdot)$ is denoting the trace function, $\otimes$ is the Kronecker product and $\mathbf{1}_{d^3\times d^3}$ is the matrix in $d^3 \times d^3$, where every entry equals one. Simplifying leads to:

$$\begin{aligned}
&= f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) d\bar{\mathbf{z}} + \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})(H^{\frac{1}{2}}\bar{\mathbf{z}})^\top \nabla_{\mathbf{z}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) d\bar{\mathbf{z}} \\
&\quad + \frac{1}{2}\int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})(H^{\frac{1}{2}}\bar{\mathbf{z}})^\top \mathcal{H}_{\mathbf{z}}(H^{\frac{1}{2}}\bar{\mathbf{z}}) d\bar{\mathbf{z}} + O(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))),
\end{aligned}$$

using that $\mathrm{K}(\mathbf{z})$ is a d-dimensional density function and again simplifying (note that $H^{\frac{1}{2}}$ is also symmetric) leads to

$$\begin{aligned}
&= f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})\bar{\mathbf{z}}^\top H^{\frac{1}{2}} \nabla_{\mathbf{z}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) d\bar{\mathbf{z}} \\
&\quad + \frac{1}{2}\int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})\bar{\mathbf{z}}^\top H^{\frac{1}{2}} \mathcal{H}_{\mathbf{z}} H^{\frac{1}{2}} \bar{\mathbf{z}} d\bar{\mathbf{z}} + O(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))). \quad (4.4)
\end{aligned}$$

Let $\mathbf{v} := (v_1, \dots, v_d) = H^{\frac{1}{2}} \nabla_{\mathbf{z}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ and the matrix elements $m_{l,k} := (H^{\frac{1}{2}} \mathcal{H}_{\mathbf{z}} H^{\frac{1}{2}})_{l,k}$. By using the definition of the (scalar product and quadratic form) we can rewrite the term above as

$$\begin{aligned}
&= f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \sum_{j=1}^{d} v_j \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})\bar{z}_j d\bar{\mathbf{z}} \\
&\quad + \frac{1}{2}\sum_{l=1}^{d}\sum_{k=1}^{d} m_{l,k} \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})\bar{z}_l \bar{z}_k d\bar{\mathbf{z}} + O(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))),
\end{aligned}$$

we conclude then by applying that the kernel is symmetric around zero and (2.11),

$$= f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \frac{1}{2}\mu_{\mathcal{K}} \sum_{l=1}^{d} m_{l,l} + O(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))),$$

finally by noticing that $\mathrm{tr}(H^{\frac{1}{2}} \mathcal{H}_{\mathbf{z}} H^{\frac{1}{2}}) = \mathrm{tr}(H^{\frac{1}{2}} H^{\frac{1}{2}} \mathcal{H}_{\mathbf{z}}) = \mathrm{tr}(H\mathcal{H}_{\mathbf{z}})$,

$$= f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \frac{1}{2}\mu_{\mathcal{K}} \mathrm{tr}(H\mathcal{H}_{\mathbf{z}}) + O(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))).$$

Since H converges to the zero matrix, we get our requested result. □

The following procedure will be a local linear regression, where an introduction can be found in Section 2.5.

Using the p-dimensional Taylor expansion of first order of $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{a})$ around $\mathbf{x} = (x_1, ..., x_p)^\top \in \mathbb{R}^p$ we get

$$\mathrm{E}\left(K_H(\mathbf{Z}-\mathbf{z})\middle|\mathbf{X}=\mathbf{a}\right) \approx f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{a}) \approx \underbrace{f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}_{=:\beta_0} + (\mathbf{a}-\mathbf{x})^\top \underbrace{\nabla_{\mathbf{x}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}_{=:\boldsymbol{\beta}_1^\top},$$

where $\nabla_{\mathbf{x}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) := \left(\frac{\partial}{\partial x_1} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}), \quad \ldots \quad , \frac{\partial}{\partial x_p} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\right)^\top$ are the partial derivatives of $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ with respect to $\mathbf{x}$.

This suggests the following regression model:

$$K_H(\mathbf{Z}-\mathbf{z}) = \underbrace{f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}_{=:\beta_0} + (\mathbf{a}-\mathbf{x})^\top \underbrace{\nabla_{\mathbf{x}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}_{=:\boldsymbol{\beta}_1^\top} + \epsilon,$$

with random variable ϵ. By interpreting $K_H(\mathbf{Z}-\mathbf{z})$ as the response random variable, ϵ as the error random variable, $(1, (\mathbf{a}-\mathbf{x})^\top)^\top$ as the covariate vector and

$$\boldsymbol{\beta}^\top = \left(f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}), \tfrac{\partial}{\partial x_1} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}), \ldots, \tfrac{\partial}{\partial x_p} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\right)^\top$$

as the parameter vector, we can use the weighted least squares approach. Therefore let $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ minimize:

$$\sum_{i=1}^{n} \left[K_H(\mathbf{z}_i - \mathbf{z}) - \beta_0 - \boldsymbol{\beta}_1^T(\mathbf{x}_i - \mathbf{x})\right]^2 w_i,$$

with weights $w_i = W_b(\mathbf{x}_i - \mathbf{x})$ for $i \in \{1, \ldots, n\}$ where $W_b(\cdot)$ is a p-dimensional product kernel with bandwidth b,

$$W_b(\mathbf{x}) := \frac{1}{b^p} \prod_{j=1}^{p} K\left(\frac{x_j}{b}\right), \text{ for } b > 0.$$

We can then estimate $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ by $\hat{\beta}_0$. As mentioned in Section 2.3

$$\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t\right)^\top = (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y}$$

and therefore using $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{p+1}$ our estimator is

$$\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) := e_1^\top (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y}. \tag{4.5}$$

Using $(\mathbf{z}_i, \mathbf{x}_i) := ((z_{i,1}, \dots, z_{i,d})^\top, (x_{i,1}, \dots, x_{i,p})^\top)$ for the i'th observation, we get for $\mathcal{Y}$, $\mathcal{X}$, $\mathcal{W}$:

$$\mathcal{Y} := \begin{pmatrix} K_H(\mathbf{z}_1 - \mathbf{z}) \\ K_H(\mathbf{z}_2 - \mathbf{z}) \\ \vdots \\ K_H(\mathbf{z}_n - \mathbf{z}) \end{pmatrix},$$

$$\mathcal{X} := \begin{pmatrix} 1 & (x_{1,1} - x_1) & \dots & (x_{1,p} - x_p) \\ 1 & (x_{2,1} - x_1) & \dots & (x_{2,p} - x_p) \\ \vdots & \vdots & & \vdots \\ 1 & (x_{n,1} - x_1) & \dots & (x_{n,p} - x_p) \end{pmatrix},$$

$$\mathcal{W} := \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_n \end{pmatrix} \qquad \begin{array}{l} w_i = W_b(\mathbf{x}_i - \mathbf{x}) \\ \text{for } i \in \{1, \dots, n\}. \end{array}$$

Note that we can also rewrite the problem by expanding the matrix-products: $\sum_{i=1}^n \sum_{k=1}^{p+1} \mathcal{X}_{ij} \mathcal{W}_{ii} \mathcal{X}_{ik} \hat{\boldsymbol{\beta}}_k = \sum_{i=1}^n \mathcal{X}_{ij} \mathcal{W}_{ii} \mathcal{Y}_i$ for $j \in \{1, \dots, (p+1)\}$.

Lemma 4.2. *In the special case of $p = 1$ (dimension of the conditional vector $\boldsymbol{x}$ is one) we can write:*

$$\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) = \frac{\sum_{i=1}^n K_H(\boldsymbol{z}_i - \boldsymbol{z}) \cdot R_i}{\sum_{i=1}^n R_i},$$

where

$$R_i := W_b(x_i - x) \left(\left\{ \sum_{l=1}^n (x_l - x)^2 \cdot W_b(x_l - x) \right\} - (\hat{x}_i - x) \left\{ \sum_{l=1}^n (x_l - x) \cdot W_b(x_l - x) \right\} \right).$$

Proof. (Lemma 4.2)
In the case of $p = 1$ we have

$$\begin{aligned} \mathcal{X}^\top \mathcal{W} \mathcal{X} &= \begin{pmatrix} 1 & \dots & 1 \\ (x_1 - x) & \dots & (x_n - x) \end{pmatrix} \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix} \begin{pmatrix} 1 & (x_1 - x) \\ \vdots & \vdots \\ 1 & (x_n - x) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i (x_i - x) \\ \sum_{i=1}^n w_i (x_i - x) & \sum_{i=1}^n w_i (x_i - x)^2 \end{pmatrix}. \end{aligned}$$

Using the inverse formula for 2×2 matrices leads to

$$(\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} = \frac{1}{|\mathcal{X}^\top \mathcal{W} \mathcal{X}|} \begin{pmatrix} \sum_{i=1}^n w_i (x_i - x)^2 & -\sum_{i=1}^n w_i (x_i - x) \\ -\sum_{i=1}^n w_i (x_i - x) & \sum_{i=1}^n w_i \end{pmatrix}.$$

Using that

$$\begin{aligned} |\mathcal{X}^\top \mathcal{W} \mathcal{X}| &= \left(\sum_{i=1}^n w_i \right) \left(\sum_{l=1}^n w_l (x_l - x)^2 \right) - \left(\sum_{i=1}^n w_i (x_i - x) \right) \left(\sum_{l=1}^n w_l (x_l - x) \right) \\ &= \sum_{i=1}^n R_i, \end{aligned}$$

yields

$$(\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} = \frac{1}{\sum_{i=1}^n R_i} \begin{pmatrix} \sum_{i=1}^n w_i (x_i - x)^2 & -\sum_{i=1}^n w_i (x_i - x) \\ -\sum_{i=1}^n w_i (x_i - x) & \sum_{i=1}^n w_i \end{pmatrix}.$$

Also note that

$$\begin{aligned} \mathcal{X}^\top \mathcal{W} \mathcal{Y} &= \begin{pmatrix} 1 & \dots & 1 \\ (x_1 - x) & \dots & (x_n - x) \end{pmatrix} \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix} \begin{pmatrix} K_H(\mathbf{z}_1 - \mathbf{z}) \\ \vdots \\ K_H(\mathbf{z}_n - \mathbf{z}) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n w_i K_H(\mathbf{z}_i - \mathbf{z}) \\ \sum_{i=1}^n w_i (x_i - x) K_H(\mathbf{z}_i - \mathbf{z}) \end{pmatrix}. \end{aligned}$$

Combining these results leads to

$$\begin{aligned} \hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) &= e_1^\top (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y} \\ &= \frac{1}{\sum_{i=1}^n R_i} \left\{ \left(\sum_{i=1}^n w_i K_H(\mathbf{z}_i - \mathbf{z}) \right) \left(\sum_{l=1}^n w_l (x_l - x)^2 \right) \right. \\ &\qquad \left. - \left(\sum_{i=1}^n w_i (x_i - x) K_H(\mathbf{z}_i - \mathbf{z}) \right) \left(\sum_{l=1}^n w_l (x_l - x) \right) \right\} \\ &= \frac{\sum_{i=1}^n K_H(\mathbf{z}_i - \mathbf{z}) \cdot R_i}{\sum_{i=1}^n R_i}. \end{aligned}$$

□

As in the case of an ordinary kernel density estimator the quality of our estimation is highly dependent on the choice of the bandwidth parameters. A method for the section of such smoothing parameters will be presented in Chapter 6.

5. Asymptotic behavior

This chapter contains the main theorem of this thesis (Theorem 5.1). It states that under certain assumptions the conditional density estimator $\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ (derived in Section 4.2) is asymptotically normal distributed. The following sections will then deal with the proof of this theorem.

5.1. Asymptotic properties

Concerning the asymptotic distribution of our estimator we can state the following theorem.

Theorem 5.1. *Assuming*

(i) $W_b(\cdot)$ *is a p-dimensional product kernel with bandwidth* b *and univariate density function* $K(\cdot)$ *and* $K_H(\cdot)$ *is a general kernel with bandwidth matrix* H *and d-dimensional density function* $\mathcal{K}(\cdot)$ *(see Section 2.4).*

(ii) $\{\boldsymbol{X}_i, \boldsymbol{Z}_i\}$ *are i.i.d for* $i \in \{1, \dots, n\}$

(iii) $b \to 0$ *and* $H \to \boldsymbol{0}_{d\times d}$ *as* $n \to \infty$*, where* $\boldsymbol{0}_{d\times d}$ *denotes* $d \times d$ *zero matrix.*

(iv) $\frac{\mathrm{tr}(\boldsymbol{1}_{d\times d}H)}{b} = O(1)$

(v) $nb^{p+2}|H|^{\frac{1}{2}} \to \infty$

(vi) $\sqrt{nb^p|H|^{\frac{1}{2}}}b^3 \to 0$

(vii) $\sqrt{n|H|^{\frac{1}{2}}b^p}\mathrm{tr}(\boldsymbol{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}})) \to 0$*, where* $\otimes$ *is the Kronecker product.*

it holds that:

$$\sqrt{n|H|^{\frac{1}{2}}b^p} \cdot \left(\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) - f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) - \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial x_j^2} - \frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\boldsymbol{z}}) \right) \to N\left(0, R_{\mathcal{K}}\nu_K^p \frac{f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})}\right),$$

where $\mu_K = \int t^2 K(t)dt$*,* $\nu_K = \int_{\mathbb{R}} K(t)^2 dt$*,* $R_{\mathcal{K}} = \int_{\mathbb{R}^d} \mathcal{K}(\boldsymbol{z})^2 d\boldsymbol{z}$*,* $\mu_{\mathcal{K}}$ *is defined such that* $\int_{\mathbb{R}^d} \boldsymbol{z}\boldsymbol{z}^\top \mathcal{K}(\boldsymbol{z})d\boldsymbol{z} = \mu_{\mathcal{K}} I$*,* $\mathcal{H}_{\boldsymbol{z}} \in \mathbb{R}^{d\times d}$ *is the Hessian of* $f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})$ *with respect to* $\boldsymbol{z}$ *and* $f_{\boldsymbol{X}}(\boldsymbol{x})$ *is the density function of* $\boldsymbol{X}$*.*

Remark 5.1. *Theorem 5.1 would also hold, if one would replace condition (iv) by the condition* $\sqrt{nb^p|H|^{\frac{1}{2}}}b^2\mathrm{tr}(\boldsymbol{1}_{d\times d}H) \to 0$.

The next section will provide the proof of this theorem. By choosing for $K_H(\mathbf{z})$ also a product kernel $K_h(\mathbf{z})$ with bandwidth $h > 0$, we get a more symmetric form, which can be seen in the following corollary.

Corollary 5.1. *Under the following conditions:*

(i) *The univariate kernels* $K(\cdot)$ *used in the product kernels* $W_b(\cdot)$ *and* $K_h(\cdot)$ *are equal.*

(ii) $K(\cdot)$ *is symmetric around zero and fulfills* $\int_{\mathbb{R}} |z^5|K(z)dz < \infty$

(iii) $\{\boldsymbol{X}_i, \boldsymbol{Z}_i\}$ *are i.i.d for* $i \in \{1, \dots, n\}$

(iv) $b \to 0$ *and* $h \to 0$ *as* $n \to \infty$

(v) $\frac{h^2}{b} = O(1)$

(vi) $nb^{p+2}h^d \to \infty$

(vii) $nb^{p+9}h^d \to 0$

(viii) $nb^p h^{d+9} \to 0$

it holds that:

$$\sqrt{nh^d b^p} \cdot \left(\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) - f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) - \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial x_j^2} - \mu_K \sum_{j=1}^{d} \frac{h^2}{2} \frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial z_j^2}\right) \to N\left(0, \nu_K^{d+p} \frac{f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})}\right),$$

where
$\mu_K = \int t^2 K(t)dt$, $\nu_K = \int K(t)^2 dt$, *and* $f_{\boldsymbol{X}}(\boldsymbol{x})$ *is the density function of* $\boldsymbol{X}$.

A similar theorem for the case of $d = 1$ and $p = 1$ is stated in Fan et al. (1996) in the remark after Theorem 1.

Corollary 5.2. *(* Bias *and variance of* $\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})$*)*
It follows from the proof of Theorem 5.1 (see Remark 5.2) that the leading term of the bias of the estimator $\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})$ *is*

$$\begin{aligned}\mathrm{Bias}\{\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})\} &= \mathrm{E}\left[\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})\right] - f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) \\ &\approx \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial x_j^2} + \frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\boldsymbol{z}}),\end{aligned}$$

and the leading term of the variance of the estimator $\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})$ *is*

$$\mathrm{Var}\{\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})\} \approx \frac{1}{n|H|^{\frac{1}{2}}b^p} R_{\mathcal{K}} \nu_K^p \frac{f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})}.$$

To give an interpretation of the bias and variance of a transformation estimator, we state the following lemma.

Lemma 5.1. *(*Bias *and variance under transformation)*
Let $T(\boldsymbol{U})$ be the following transformation of $\boldsymbol{U}$:

$$T(\boldsymbol{U}) := (T_1(U_1), T_2(U_2), \ldots, T_d(U_d))^\top.$$

If we use

$$\hat{g}(\boldsymbol{u}|\boldsymbol{x}) := \hat{f}_{T(\boldsymbol{u})|\boldsymbol{X}}(T(\boldsymbol{u})|\boldsymbol{x}) \cdot \prod_{j=1}^{d} \frac{\partial}{\partial u_j} T_j(u_j),$$

as the estimator for the conditional density $f_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})$, where $\hat{f}_{T(\boldsymbol{U})|\boldsymbol{X}}(T(\boldsymbol{u})|\boldsymbol{x})$ is the local linear regression estimator (4.5) (applied on $(T(\boldsymbol{U}_i), \boldsymbol{X}_i)$). The leading terms of the bias and variance of the estimator $\hat{g}(\boldsymbol{u}|\boldsymbol{x})$ are:

$$\begin{aligned} Bias\{\hat{g}(\boldsymbol{u}|\boldsymbol{x})\} &= E[\hat{g}(\boldsymbol{u}|\boldsymbol{x})] - f_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x}) \\ &\approx \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})}{\partial x_j^2} + \frac{L}{2} \mu_{\mathcal{K}} \mathrm{tr}(H \mathcal{H}_{T(\boldsymbol{u})}), \\ Var\{\hat{g}(\boldsymbol{u}|\boldsymbol{x})\} &\approx \frac{L}{n|H|^{\frac{1}{2}} b^p} R_{\mathcal{K}} \nu_K^p \frac{f_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})}, \end{aligned}$$

where $L := \prod_{j=1}^{d} \frac{\partial}{\partial u_j} T_j(u_j)$ and $\mathcal{H}_{T(\boldsymbol{u})}$ is the Hessian of $f_{T(\boldsymbol{U})|\boldsymbol{X}}(T(\boldsymbol{u})|\boldsymbol{x})$ with respect to $T(\boldsymbol{u})$.

Proof. (Lemma 5.1)
Note that by Theorem 2.2 (for conditional densities), we get

$$f_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x}) = |\det(dT(\mathbf{u}))| f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x}).$$

We have in our case the following transformation

$$T(\mathbf{U}) := (T_1(U_1), T_2(U_2), \ldots, T_d(U_d))^\top.$$

Therefore,

$$\begin{aligned} dT(\mathbf{u}) &= \begin{pmatrix} \frac{\partial T_1}{\partial u_1}(\mathbf{u}) & \ldots & \frac{\partial T_1}{\partial u_d}(\mathbf{u}) \\ \vdots & \ddots & \vdots \\ \frac{\partial T_d}{\partial u_1}(\mathbf{u}) & \ldots & \frac{\partial T_d}{\partial u_d}(\mathbf{u}) \end{pmatrix} = \begin{pmatrix} \frac{\partial T_1}{\partial u_1}(u_1) & & 0 \\ & \ddots & \\ 0 & & \frac{\partial T_d}{\partial u_d}(u_d) \end{pmatrix}, \\ \Rightarrow \det(dT(\mathbf{u})) &= \prod_{j=1}^{d} \frac{\partial}{\partial u_j} T_j(u_j) = L. \end{aligned}$$

Therefore we get that

$$f_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x}) = f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x}) \cdot \prod_{i=1}^{d} \frac{\partial}{\partial u_i} T_i(u_i) = L \cdot f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x}).$$

We conclude

$$\begin{aligned}
\mathrm{Bias}\{\hat{g}_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})\} &= \mathrm{E}\left[\hat{g}_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})\right] - f_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x}) \\
&= \mathrm{E}[\hat{f}_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x}) \cdot L] - f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x}) \cdot L, \\
\mathrm{Var}[\hat{g}_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})] &= L^2 \mathrm{Var}[\hat{f}_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x})].
\end{aligned}$$

By using now Corollary 5.2 we get

$$\begin{aligned}
\mathrm{Bias}\{\hat{g}_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})\} &\approx L \left\{ \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x})}{\partial x_j^2} + \frac{1}{2} \mu_{\mathcal{K}} \mathrm{tr}(H \mathcal{H}_{T(\mathbf{u})}) \right\}, \\
\mathrm{Var}\{\hat{g}_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})\} &\approx L^2 \left(\frac{1}{n|H|^{\frac{1}{2}} b^p} R_{\mathcal{K}} \nu_K^p \frac{f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right),
\end{aligned}$$

where $\mathcal{H}_{T(\mathbf{u})}$ is the Hessian of $f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x})$ with respect to $T(\mathbf{u})$. Finally since $f_{T(\mathbf{U})|\mathbf{X}}(T(\mathbf{u})|\mathbf{x}) = \frac{f_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})}{L}$ and L is independent of $\mathbf{x}$ we get our result for the bias and variance. □

The following corollary is providing the leading terms of the bias and variance of our conditional copula density estimator (4.2).

Corollary 5.3. *(*Bias *and variance of conditional copula density estimator) Let* $\boldsymbol{Z} = (\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_d))$*. The leading terms of the bias and variance of the estimator* $\hat{c}_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})$ *(4.2) are:*

$$\begin{aligned}
Bias\{\hat{c}_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})\} &= E[\hat{c}_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})] - c_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x}) \\
&\approx \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 c_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})}{\partial x_j^2} + \frac{L}{2} \mu_{\mathcal{K}} \mathrm{tr}(H \mathcal{H}_{\boldsymbol{z}}), \\
Var\{\hat{c}_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})\} &\approx \frac{L}{n|H|^{\frac{1}{2}} b^p} R_{\mathcal{K}} \nu_K^p \frac{c_{\boldsymbol{U}|\boldsymbol{X}}(\boldsymbol{u}|\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})},
\end{aligned}$$

where $L := \left[\prod_{j=1}^{d} \phi(\Phi^{-1}(u_j))\right]^{-1}$ *and* $\mathcal{H}_{\boldsymbol{z}}$ *is the Hessian of* $f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})$ *with respect to* $\boldsymbol{z}$*.*

Note that Corollary 5.3 follows immediately from Lemma 5.1 by using the transformation $\mathbf{Z} := T(\mathbf{U}) = (\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_d))$ and the inverse function theorem.

5.2. Proof of Theorem 5.1

The structure of the following proof is similar to the one in the proof of Theorem 2.7 in Li and Racine (2006), where they proof the asymptomatic normality of the standard local linear regression estimator. It is also noteworthy that in the unpublished report Fan et al. (1993) a proof for the quadratic regression approach can be found, for the case of $p = 1$ and $d = 1$, and a ρ-mixing condition.

We will define for notational reasons the conditional expectation function $g(\cdot)$ as follows:

$$g(\mathbf{a}) := E\left(K_H(\mathbf{Z} - \mathbf{z})|\mathbf{X} = \mathbf{a}\right) = \int_{\mathbb{R}^d} K_H(\tilde{\mathbf{z}} - \mathbf{z}) f_{\mathbf{Z}|\mathbf{X}}(\tilde{\mathbf{z}}|\mathbf{a}) d\tilde{\mathbf{z}}.$$

Note that the conditional expectation is defined such that the random vectors $\mathbf{Z}$ and $\mathbf{X}$ have the joint probability density function $f_{\mathbf{Z},\mathbf{X}}(\mathbf{z}, \mathbf{x})$.

Lemma 5.2. *It holds that:*

$$g(\boldsymbol{X}_i) = E\left(K_H(\boldsymbol{Z}_i - \boldsymbol{z})|\boldsymbol{X}_i\right).$$

It is important to notice, that the $\mathbf{Z}$ in the conditional expectation of Lemma 5.2 has now also the index i.

Proof. (Lemma 5.2)
This proof is similar to the one of Theorem 3.1. We have to show the following:

(i) $g(\mathbf{X}_i)$ is measurable w.r.t. the σ-algebra $\sigma(\mathbf{X}_i)$.

(ii) For every bounded random variable L which is measurable w.r.t. $\sigma(\mathbf{X}_i)$, we have $E[L \cdot g(\mathbf{X}_i)] = E[L \cdot K_H(\mathbf{Z}_i - \mathbf{z})]$.

Statement (i) is clear because we assume that $g(a)$ is a continuous function in a, so only statement (ii) is remaining. Since we know for $i \in \{1, \ldots, n\}$ that $(\mathbf{X}_i, \mathbf{Z}_i)$ has the joint probability density function $f_{\mathbf{Z},\mathbf{X}}(\mathbf{z}, \mathbf{x})$, we can rewrite statement (ii) as follows:

$$\begin{aligned}
(ii) \iff & E[h(\mathbf{X}_i) \cdot g(\mathbf{X}_i)] = E[h(\mathbf{X}_i) \cdot K_H(\mathbf{Z}_i - \mathbf{z})] \\
\iff & \int_{\mathbb{R}^p} h(\mathbf{x}) g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^p} \int_{\mathbb{R}^d} h(\mathbf{x}) \cdot K_H(\tilde{\mathbf{z}} - \mathbf{z}) f_{\mathbf{Z},\mathbf{X}}(\tilde{\mathbf{z}}, \mathbf{x}) d\tilde{\mathbf{z}} d\mathbf{x} \\
\iff & \int_{\mathbb{R}^p} h(\mathbf{x}) \int_{\mathbb{R}^d} K_H(\tilde{\mathbf{z}} - \mathbf{z}) f_{\mathbf{Z}|\mathbf{X}}(\tilde{\mathbf{z}}|\mathbf{x}) d\tilde{\mathbf{z}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
& = \int_{\mathbb{R}^p} \int_{\mathbb{R}^d} h(\mathbf{x}) \cdot K_H(\tilde{\mathbf{z}} - \mathbf{z}) f_{\mathbf{Z},\mathbf{X}}(\tilde{\mathbf{z}}, \mathbf{x}) d\tilde{\mathbf{z}} d\mathbf{x},
\end{aligned}$$

where $h(\mathbf{X}) = L$ (see proof of Theorem 3.1). We used in the last step the definition of the condition distribution (3.1). □

We will now use the definition of $g(\cdot)$ to proof Theorem 5.1.

Proof. (Theorem 5.1) We begin our proof by remembering that:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_1 \end{pmatrix} = \begin{pmatrix} \hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) \\ \hat{\boldsymbol{\beta}}_1 \end{pmatrix} = (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y}. \tag{5.1}$$

Note that $\mathcal{Y}$,$\mathcal{X}$ and $\mathcal{W}$ were defined as follows

$$\mathcal{Y} := \begin{pmatrix} K_H(\mathbf{Z}_1 - \mathbf{z}) \\ K_H(\mathbf{Z}_2 - \mathbf{z}) \\ \vdots \\ K_H(\mathbf{Z}_n - \mathbf{z}) \end{pmatrix},$$

$$\mathcal{X} := \begin{pmatrix} 1 & (X_{1,1} - x_1) & \dots & (X_{1,p} - x_p) \\ 1 & (X_{2,1} - x_1) & \dots & (X_{2,p} - x_p) \\ \vdots & \vdots & & \vdots \\ 1 & (X_{n,1} - x_1) & \dots & (X_{n,p} - x_p) \end{pmatrix} = \begin{pmatrix} 1, (\mathbf{X}_1 - \mathbf{x})^\top \\ 1, (\mathbf{X}_2 - \mathbf{x})^\top \\ \vdots \\ 1, (\mathbf{X}_n - \mathbf{x})^\top \end{pmatrix},$$

$$\mathcal{W} := \begin{pmatrix} W_b(\mathbf{X}_1 - \mathbf{x}) & & 0 \\ & \ddots & \\ 0 & & W_b(\mathbf{X}_n - \mathbf{x}) \end{pmatrix}.$$

As in Section 4.2 we use the notation $W_b(\cdot)$ for a p-dimensional product kernel with bandwidth b and $K_H(\cdot)$ for a d-dimensional general kernel with bandwidth matrix H.

Rewriting (5.1) leads to

$$\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) = \hat{\beta}_0 = e_1^\top (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathcal{Y} = e_1^\top \left(\frac{1}{n} \mathcal{X}^\top \mathcal{W} \mathcal{X} \right)^{-1} \frac{1}{n} \mathcal{X}^\top \mathcal{W} \mathcal{Y}, \tag{5.2}$$

where $e_1 = (1, 0, \dots, 0)^\top \in \mathbb{R}^{p+1}$.

Define now $\mathcal{G}_p$ as

$$\mathcal{G}_p := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & b^{-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^{-2} \end{pmatrix} \in \mathbb{R}^{(p+1,p+1)}.$$

The next step is to insert the identity matrix $I_{p+1} = (\mathcal{G}_p^{-1} \mathcal{G}_p)$ in the equation (5.2). By using that $B^{-1} A^{-1} = (AB)^{-1}$, we get

$$\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) = e_1^\top \left(\frac{1}{n} \mathcal{X}^\top \mathcal{W} \mathcal{X} \right)^{-1} \frac{1}{n} \mathcal{X}^\top \mathcal{W} \mathcal{Y} = e_1^\top \left(\frac{1}{n} \mathcal{G}_p \mathcal{X}^\top \mathcal{W} \mathcal{X} \right)^{-1} \frac{1}{n} \mathcal{G}_p \mathcal{X}^\top \mathcal{W} \mathcal{Y}.$$

Simplifying the matrix product leads to

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\frac{1}{n}\mathcal{G}_p\mathcal{X}^\top\mathcal{W}\mathcal{X}\right)^{-1}\left(\frac{1}{n}\mathcal{G}_p\mathcal{X}^\top\mathcal{W}\mathcal{Y}\right) \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\mathcal{G}_p\begin{pmatrix}1\\ \mathbf{X}_i-\mathbf{x}\end{pmatrix}\left(1,(\mathbf{X}_i-\mathbf{x})^\top\right)\right]^{-1} \\
&\quad\cdot\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\mathcal{G}_p\begin{pmatrix}1\\ \mathbf{X}_i-\mathbf{x}\end{pmatrix}K_H(\mathbf{Z}_i-\mathbf{z})\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix}1\\ b^{-2}(\mathbf{X}_i-\mathbf{x})\end{pmatrix}\left(1,(\mathbf{X}_i-\mathbf{x})^\top\right)\right]^{-1} \\
&\quad\cdot\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix}1\\ b^{-2}(\mathbf{X}_i-\mathbf{x})\end{pmatrix}K_H(\mathbf{Z}_i-\mathbf{z})\right].
\end{aligned} \tag{5.3}$$

We define now ϵ_i as

$$\epsilon_i := K_H(\mathbf{Z}_i-\mathbf{z}) - g(\mathbf{X}_i), \tag{5.4}$$

and by rewriting (5.4) and applying in (5.3) we get

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix}1\\ b^{-2}(\mathbf{X}_i-\mathbf{x})\end{pmatrix}\left(1,(\mathbf{X}_i-\mathbf{x})^\top\right)\right]^{-1} \\
&\quad\cdot\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix}1\\ b^{-2}(\mathbf{X}_i-\mathbf{x})\end{pmatrix}\left(g(\mathbf{X}_i)+\epsilon_i\right)\right].
\end{aligned} \tag{5.5}$$

We know by Taylor expansion that

$$g(\mathbf{X}_i) = g(\mathbf{x}) + (\mathbf{X}_i-\mathbf{x})^\top\nabla g(\mathbf{x}) + \frac{1}{2}(\mathbf{X}_i-\mathbf{x})^\top\mathcal{H}_g(\mathbf{x})(\mathbf{X}_i-\mathbf{x}) + R(\mathbf{x},\mathbf{X}_i) \text{ a.s.},$$

where $\mathcal{H}_g(\mathbf{x})$ is the Hessian of $g(\mathbf{x})$ and $R(\mathbf{x},\mathbf{X}_i)$ is the remainder term. By applying this in (5.5) we get

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix}1\\ b^{-2}(\mathbf{X}_i-\mathbf{x})\end{pmatrix}\left(1,(\mathbf{X}_i-\mathbf{x})^\top\right)\right]^{-1} \\
&\quad\cdot\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix}1\\ b^{-2}(\mathbf{X}_i-\mathbf{x})\end{pmatrix}\left\{\left(1,(\mathbf{X}_i-\mathbf{x})^\top\right)\begin{pmatrix}g(\mathbf{x})\\ \nabla g(\mathbf{x})\end{pmatrix}\right.\right. \\
&\qquad\qquad\left.\left.+\frac{1}{2}(\mathbf{X}_i-\mathbf{x})^\top\mathcal{H}_g(\mathbf{x})(\mathbf{X}_i-\mathbf{x}) + R(\mathbf{x},\mathbf{X}_i) + \epsilon_i\right\}\right].
\end{aligned}$$

Simplifying this equation leads to

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \begin{pmatrix} g(\mathbf{x}) \\ \nabla g(\mathbf{x}) \end{pmatrix} + \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \left(1, (\mathbf{X}_i - \mathbf{x})^\top\right)\right]^{-1} \\ &\quad \cdot \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) \right.\right. \\ &\qquad\qquad \left.\left. + R(\mathbf{x}, \mathbf{X}_i) + \epsilon_i \right\}\right] \\ &= \begin{pmatrix} g(\mathbf{x}) \\ \nabla g(\mathbf{x}) \end{pmatrix} + \mathcal{M}_1^{-1}\left[\mathcal{M}_2 + \mathcal{M}_3\right] + (s.o.), \end{aligned} \tag{5.6}$$

where

$$\begin{aligned}\mathcal{M}_1 &= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \left(1, (\mathbf{X}_i - \mathbf{x})^\top\right)\right] \in \mathbb{R}^{(p+1)\times(p+1)}, \\ \mathcal{M}_2 &= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\}\right] \in \mathbb{R}^{p+1}, \\ \mathcal{M}_3 &= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \{\epsilon_i\}\right] \in \mathbb{R}^{p+1}, \\ (s.o.) &= \mathcal{M}_1^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} R(\mathbf{x}, X_i)\right] \in \mathbb{R}^{p+1}. \end{aligned}$$

Now define the matrix $\mathcal{D}$ as

$$\mathcal{D} := \begin{pmatrix} \sqrt{n|H|^{\frac{1}{2}} b^p} & 0 & \dots & 0 \\ 0 & \sqrt{n|H|^{\frac{1}{2}} b^p b} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{n|H|^{\frac{1}{2}} b^p b} \end{pmatrix} \in \mathbb{R}^{(p+1)\times(p+1)}.$$

We use in the following derivation Lemma A.1 to Lemma A.4, which can be found (with their proofs) in Appendix A.

Rewriting (5.6) and multiplying both sides by $\mathcal{D}$ leads to:

$$\mathcal{D}\left(\hat{\boldsymbol{\beta}} - \begin{pmatrix} g(\mathbf{x}) \\ \nabla g(\mathbf{x}) \end{pmatrix}\right) = \mathcal{D}\mathcal{M}_1^{-1}\left[\mathcal{M}_2 + \mathcal{M}_3\right] + \mathcal{D}(s.o.). \tag{5.7}$$

Lemma A.1 states

$$e_1^\top \mathcal{D}\mathcal{M}_1^{-1}\left[\mathcal{M}_2 + \mathcal{M}_3\right] + e_1^\top \mathcal{D}(s.o.) = f_{\mathbf{X}}(\mathbf{x})^{-1} e_1^\top \mathcal{D}\left[\mathcal{M}_2 + \mathcal{M}_3\right] + o_p(1).$$

So after multiplying equation (5.7) by $e_1^\top$ and then applying Lemma A.1 we get

$$e_1^\top \mathcal{D}\left(\hat{\boldsymbol{\beta}} - \begin{pmatrix} g(\mathbf{x}) \\ \nabla g(\mathbf{x}) \end{pmatrix}\right) = f_{\mathbf{X}}(\mathbf{x})^{-1} e_1^\top \mathcal{D}\left[\mathcal{M}_2 + \mathcal{M}_3\right] + o_p(1).$$

Simplifying leads to

$$\sqrt{n|H|^{\frac{1}{2}} b^p}\left(\hat{\beta}_0 - g(\mathbf{x})\right) = f_{\mathbf{X}}(\mathbf{x})^{-1} e_1^\top \mathcal{D}\mathcal{M}_2 + f_{\mathbf{X}}(\mathbf{x})^{-1} e_1^\top \mathcal{D}\mathcal{M}_3 + o_p(1). \quad (5.8)$$

Now a useful result from Lemma A.2 will be

$$e_1^\top \mathcal{D}\mathcal{M}_2 = f_{\mathbf{X}}(\mathbf{x})\sqrt{n|H|^{\frac{1}{2}} b^p}\mu_K \sum_{j=1}^{p} \frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + o_p(1).$$

Using this result, we can rewrite (5.8) as

$$\begin{aligned}\sqrt{n|H|^{\frac{1}{2}} b^p}\left(\hat{\beta}_0 - g(\mathbf{x})\right) &= f_{\mathbf{X}}(\mathbf{x})^{-1} f_{\mathbf{X}}(\mathbf{x})\sqrt{n|H|^{\frac{1}{2}} b^p}\mu_K \sum_{j=1}^{p} \frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} \\ &\quad + f_{\mathbf{X}}(\mathbf{x})^{-1} e_1^\top \mathcal{D}\mathcal{M}_3 + o_p(1),\end{aligned}$$

and reordering leads to

$$\begin{aligned}\sqrt{n|H|^{\frac{1}{2}} b^p}\left(\hat{\beta}_0 - g(\mathbf{x})\right) - \sqrt{n|H|^{\frac{1}{2}} b^p}\mu_K \sum_{j=1}^{p} \frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} \\ = f_{\mathbf{X}}(\mathbf{x})^{-1} e_1^\top \mathcal{D}\mathcal{M}_3 + o_p(1).\end{aligned} \quad (5.9)$$

At this point we can use the established result from Lemma A.3:

$$e_1^\top \mathcal{D}\mathcal{M}_3 \xrightarrow{d} N\left(0, R_{\mathcal{K}}\nu_K^p f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\right).$$

Together with common results from asymptotic theory we get that

$$f_{\mathbf{X}}(\mathbf{x})^{-1} e_1^\top \mathcal{D}\mathcal{M}_3 + o_p(1) \xrightarrow{d} N\left(0, R_{\mathcal{K}}\nu_K^p \frac{f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}\right).$$

Therefore one gets for the asymptotic distribution of the left hand side in (5.9):

$$\sqrt{n|H|^{\frac{1}{2}} b^p}\left(\hat{\beta}_0 - g(\mathbf{x}) - \mu_K \sum_{j=1}^{p} \frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2}\right) \xrightarrow{d} N\left(0, R_{\mathcal{K}}\nu_K^p \frac{f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}\right). \quad (5.10)$$

Lemma A.4 states

$$g(\mathbf{x}) = f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\mathbf{z}}) + O\left(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))\right),$$

so one can finally also conclude that

$$\sqrt{n|H|^{\frac{1}{2}}b^p}\left(\hat{\beta}_0 - g(\mathbf{x}) - \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2}\right)$$
$$= \sqrt{n|H|^{\frac{1}{2}}b^p} \cdot \left(\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) - f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) - \frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\mathbf{z}}) - \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + O\left(\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))\right)\right).$$

Since we know that $\sqrt{n|H|^{\frac{1}{2}}b^p}\mathrm{tr}(\mathbf{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}})) = o(1)$, we can combine this result with (5.10), which yields finally to Theorem 5.1. □

Remark 5.2. *Recall here that - as for the standard multivariate local least squares regression mentioned in Ruppert and Wand (1994) - the unconditional expectation and therefore the unconditional bias and the variance might not exist. One can see this if one considers that in (4.5) the matrix $\mathcal{W}$ can equal the zero matrix with a positive probability. But by the following derivation one can determine the non-stochastic leading terms of the asymptotic bias and variance.*

By analyzing in detail the proofs of Lemma A.1 and Lemma A.2, we get for the term $o_p(1)$ in (5.9) (see proof of Theorem 5.1) that it is $O_p(b) + \sqrt{n|H|^{\frac{1}{2}}b^p}O_p(b^3)$ and therefore one can rewrite (5.9) as:

$$\sqrt{n|H|^{\frac{1}{2}}b^p}\left\{\left(\hat{\beta}_0 - g(\boldsymbol{x})\right) - \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial x_j^2} + O_p\left(\frac{b}{\sqrt{n|H|^{\frac{1}{2}}b^p}}\right) + O_p(b^3)\right\}$$
$$= f_{\boldsymbol{X}}(\boldsymbol{x})^{-1} e_1^\top \mathcal{D}\mathcal{M}_3.$$

Notice here that since $\frac{1}{\sqrt{n|H|^{\frac{1}{2}}b^{p+2}}} = o(1)$ we get $\frac{b}{\sqrt{n|H|^{\frac{1}{2}}b^p}} = o(b^2)$.
Therefore dividing by $\sqrt{n|H|^{\frac{1}{2}}b^p}$ and applying Lemma A.4 leads to

$$\hat{f}_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) - f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) - \frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_z) - \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial x_j^2}$$
$$+ o_p\left(b^2\right) + O_p(b^3) + O\left(\mathrm{tr}(\boldsymbol{1}_{d^3\times d^3}(H^{\frac{1}{2}} \otimes H^{\frac{1}{2}} \otimes H^{\frac{1}{2}}))\right)$$
$$= f_{\boldsymbol{X}}(\boldsymbol{x})^{-1} e_1^\top \mathcal{M}_3. \tag{5.11}$$

Also by applying Lemma A.6,

$$\begin{aligned} \mathrm{E}\left[f_{\boldsymbol{X}}(\boldsymbol{x})^{-1}e_1^\top \mathcal{M}_3\right] &= \boldsymbol{0}_{p+1}, \\ \mathrm{Var}\left[f_{\boldsymbol{X}}(\boldsymbol{x})^{-1}e_1^\top \mathcal{M}_3\right] &= \frac{1}{n|H|^{\frac{1}{2}}b^p}\nu_K^p R_{\mathcal{K}} \frac{f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} + o\left(\frac{1}{n|H|^{\frac{1}{2}}b^p}\right). \end{aligned} \tag{5.12}$$

Therefore the non-stochastic leading terms of the asymptotic Bias *and* Var *will be those of Corollary 5.2.*

6. Bandwidth selection for conditional density estimation

Our approach for bandwidth selection is based on the MISE (see Section 2.6) of the estimator $\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$. In the following $\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ will be used as a random variable, which depends on the random vectors $\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{i,d})^\top$ with $i \in \{1, \ldots, n\}$ and $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,p})^\top$ with $i \in \{1, \ldots, n\}$. The mean integrated squared error for the estimator $\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ of $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ is

$$\mathrm{MISE}\left[\hat{f}_{\mathbf{Z}|\mathbf{X}}, \mathbf{x}, H, b\right] = \int_{\mathbb{R}^d} \left[\mathrm{Bias}\left[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\right]^2 + \mathrm{Var}\left[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\right]\right] d\mathbf{z},$$

which depends on the response bandwidth matrix H, the conditional bandwidth b, and the conditioning argument $\mathbf{x}$.

6.1. Bandwidth selection

There are many possible ways to select bandwidths in the field of nonparametric statistics, but nearly all of them are based on the following two main approaches. In the first category, which is called the **Cross-Validation methods** one tries to minimize the MISE. Whereas in the second one, the so called **plug-in methods**, one tries to minimize the asymptotic MISE with respect to the bandwidths. In this thesis we will present a special form of the plug-in method. It suits our problem of conditional bandwidth selection, is easy to interpret and fast to compute.

Since the exact MISE is usually hard to calculate exactly, a common workaround is the usage of the asymptotic MISE, the so called AMISE. By assuming that the integral is finite one can determine the AMISE as follows:

$$\mathrm{AMISE}\left[\hat{f}_{\mathbf{Z}|\mathbf{X}}, \mathbf{x}, H, b\right] = \int_{\mathbb{R}^d} \left[\mathrm{ABias}\big[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\big]^2 + \mathrm{AVar}\big[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\big]\right] d\mathbf{z}, \quad (6.1)$$

where the $\mathrm{ABias}(\hat{\theta})$ and $\mathrm{AVar}(\hat{\theta})$ are representing the asymptotic bias and variance of an estimator $\hat{\theta}$.

To simplify the calculation we assume in the following that the bandwidth b is known and fixed. We then estimate the main bandwidth matrix H and choose in an following step b. Note here that the so calculated minimum of the AMISE depending on (b, H) has not to be the global minimum.

Our aim is now to determine the matrix H_{AMISE} which is determined as follows:

$$H_{\text{AMISE}} = \underset{H \in \mathcal{G}}{\operatorname{argmin}} \text{ AMISE} \left[\hat{f}_{\mathbf{Z}|\mathbf{X}}, \mathbf{x}, H, b \right], \tag{6.2}$$

where $\mathcal{G}$ is the space of symmetric positive definite matrices.

A heuristically valid estimator for the optimal bandwidth matrix H_{AMISE} is the following bandwidth matrix:

$$\hat{H}_{RT} = \left(\frac{4}{d+2} \right)^{\frac{2}{d+4}} n^{\frac{-2}{d+4}} \hat{\Sigma}(\mathbf{x}), \tag{6.3}$$

where $\hat{\Sigma}(\mathbf{x})$ is the estimated covariance matrix of $\mathbf{Z}$. This bandwidth matrix is mentioned by Wand (1992) as the resulting optimal bandwidth matrix H_{AMISE} for the standard multivariate kernel density estimator, when estimating a multivariate Gaussian density and using Gaussian kernels. We will in the following refer to this bandwidth selection as the rule of thumb bandwidth matrix. Note here that we will use an estimator for the conditional covariance matrix instead of the usual covariance matrix of $\mathbf{Z}$ (see Section 6.2).

Going further with the calculation of (6.2) we will set $\mathcal{G}$ to be the set of all symmetric matrices, assuming that the positive definite property of H does not influence the result severely. To find the minimum of the AMISE in terms of the bandwidth matrix H we need to have the derivative with respect to the components of H. We introduce the following notation. A derivative $\frac{\partial}{\partial M}$ of a function $f : \mathbb{R}^k \to \mathbb{R}$ with respect to a matrix $M \in \mathbb{R}^{d \times d}$ is defined as the matrix, where every entry of the matrix coincide with the derivative corresponding to the respective matrix component:

$$\left[\frac{\partial}{\partial M} f(\mathbf{x}) \right]_{i,j} := \frac{\partial}{\partial M_{i,j}} f(\mathbf{x}).$$

We will search for a solution H_{AMISE} of the following problem:

$$\frac{\partial}{\partial H} \text{AMISE} \left[\hat{f}_{\mathbf{Z}|\mathbf{X}}, \mathbf{x}, H_{\text{AMISE}}, b \right] = 0_{d \times d}, \tag{6.4}$$

where $0_{d \times d}$ is the zero-matrix in $\mathbb{R}^{d \times d}$.

The main idea is now to use for $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ a multivariate normal density, where the mean is set to be equal to zero $\boldsymbol{\mu} = \mathbf{0}_d$ and covariance matrix is set to be dependent on the realization of the conditioning vector $\Sigma = \Sigma(\mathbf{x})$. The reason for

this idea lies in the setup of our estimation. To clarify this, we state the following lemma.

Lemma 6.1. *Let* $U_{Y_j|\boldsymbol{X}} := F_{Y_j|\boldsymbol{X}}(Y_j|\boldsymbol{X})$ *for* $j \in \{1, \dots, d\}$ *then it holds that:*

$$P(\Phi^{-1}(U_{Y_j|\boldsymbol{X}}) < t|\boldsymbol{X} = \boldsymbol{x}) = \Phi(t).$$

Proof. (Lemma 6.1)

$$\begin{aligned} P(\Phi^{-1}(U_{Y_j|\mathbf{X}}) < t|\mathbf{X} = \mathbf{x}) &= P(U_{Y_j|\mathbf{X}} < \Phi(t)|\mathbf{X} = \mathbf{x}) \\ &= P(F_{Y_j|\mathbf{X}}(Y_j|\mathbf{X}) < \Phi(t)|\mathbf{X} = \mathbf{x}) \\ &= P(F_{Y_j|\mathbf{X}}(Y_j|\mathbf{x}) < \Phi(t)|\mathbf{X} = \mathbf{x}) \\ &= P(Y_j < F^{-1}_{Y_j|\mathbf{X}}(\Phi(t)|\mathbf{x})|\mathbf{X} = \mathbf{x}) \\ &= F_{Y_j|\mathbf{X}}(F^{-1}_{Y_j|\mathbf{X}}(\Phi(t)|\mathbf{x})|\mathbf{x}) = \Phi(t), \end{aligned}$$

where $F^{-1}_{Y_j|\mathbf{X}}(t|\mathbf{x})$ is the quantile function of Y_j given $\mathbf{X} = \mathbf{x}$ for $j \in \{1, \dots, d\}$. □

Since we transform our (pseudo) conditional copula observations in equation (4.1) with the inverse of the normal cumulative distribution function, we know by Lemma 6.1 that the conditional marginals of $\mathbf{Z}$ given $\mathbf{X} = \mathbf{x}$ are standard normal distributed. Knowing that the marginals are standard normal it is reasonable to approximate the conditional joint density $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ for the bandwidth calculation by a multivariate normal density with mean equal to zero.

To calculate the AMISE as described in (6.1) one needs to calculate the integrated AVar, which we know be Corollary 5.2 can be represented as follows:

$$\int_{\mathbb{R}^d} \mathrm{AVar}\big[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\big] d\mathbf{z} = \int_{\mathbb{R}^d} \frac{1}{n|H|^{\frac{1}{2}} b^p} R_{\mathcal{K}} \nu_K^p \frac{f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} d\mathbf{z}.$$

Since a conditional density integrates to one we get

$$\int_{\mathbb{R}^d} \mathrm{AVar}\big[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\big] d\mathbf{z} = \frac{R_{\mathcal{K}} \nu_K^p}{n b^p f_{\mathbf{X}}(\mathbf{x})} \cdot |H|^{-\frac{1}{2}}. \tag{6.5}$$

By assuming that the fact that H is positive definite does not affect the derivative one easily can see that

$$\frac{\partial}{\partial H} \int_{\mathbb{R}^d} \mathrm{AVar}\big[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\big] d\mathbf{z} = -\frac{1}{2} \frac{R_{\mathcal{K}} \nu_K^p}{n b^p f_{\mathbf{X}}(\mathbf{x})} \cdot |H|^{-\frac{1}{2}} (2H^{-1} - H^{-1} \circ I_d). \tag{6.6}$$

Note that we used in (6.6) that the derivative of the determinant of a symmetric matrix $M \in \mathbb{R}^{d \times d}$ with respect to itself is $\frac{\partial}{\partial M}|M| = |M|(2M^{-1} - M^{-1} \circ I_d)$, where $\circ$ is representing the componentwise product of two matrices and I_d is the identity matrix in $\mathbb{R}^{d \times d}$.

In the next step we search for a traceable representation of the integrated ABias, to calculate the AMISE as described in (6.1). For this reason we also use Corollary 5.2 for the ABias expression.

$$\begin{aligned}\int_{\mathbb{R}^d} \mathrm{ABias}\big[\hat{f}_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\big]^2 d\mathbf{z} &= \int_{\mathbb{R}^d} \left[\mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + \frac{1}{2}\mu_{\mathcal{K}} \mathrm{tr}(H\mathcal{H}_{\mathbf{z}})\right]^2 d\mathbf{z} \\ &= \int_{\mathbb{R}^d} \left[\frac{1}{2}\mu_{\mathcal{K}} \mathrm{tr}(H\mathcal{H}_{\mathbf{z}})\right]^2 d\mathbf{z} \\ &\quad + 2\int_{\mathbb{R}^d} \mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} \frac{1}{2}\mu_{\mathcal{K}} \mathrm{tr}(H\mathcal{H}_{\mathbf{z}}) d\mathbf{z} \qquad (6.7) \\ &\quad + \int_{\mathbb{R}^d} \left[\mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2}\right]^2 d\mathbf{z}\end{aligned}$$

Since the last summand is independent of H it will be eliminated at future steps, where we take the derivative with respect to the components of H to calculate the desired minimum. For this reason we will only analyze the other components of the sum in (6.7). We need to state one further result (without proof) that will facilitate the calculation. Let $\phi_{\boldsymbol{\mu},\Sigma}(\mathbf{z})$ denote the joint density of a $\mathbf{Z} \sim N(\boldsymbol{\mu}, \Sigma)$ random vector.

Lemma 6.2. *Let $\boldsymbol{0}_d$ be the zero vector in $\mathbb{R}^d$, then the Hessian $\mathcal{H}_{\boldsymbol{z}}$ of $\phi_{\boldsymbol{0}_d,\Sigma(\boldsymbol{x})}(\boldsymbol{z})$ with respect to z is:*

$$\mathcal{H}_{\boldsymbol{z}} = \phi_{\boldsymbol{0}_d,\Sigma(\boldsymbol{x})}(\boldsymbol{z}).\left[\Sigma(\boldsymbol{x})^{-1}\boldsymbol{z}\boldsymbol{z}^\top\Sigma(\boldsymbol{x})^{-1} - \Sigma(\boldsymbol{x})^{-1}\right]. \qquad (6.8)$$

$$\phi_{\boldsymbol{0}_d,\Sigma}(\boldsymbol{z})\phi_{\boldsymbol{0}_d,\tilde{\Sigma}}(\boldsymbol{z}) = \phi_{\boldsymbol{0}_d,\Sigma+\tilde{\Sigma}}(\boldsymbol{0}_d)\phi_{\boldsymbol{0}_d,\Sigma(\Sigma+\tilde{\Sigma})^{-1}\tilde{\Sigma}}(\boldsymbol{z}). \qquad (6.9)$$

Let $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, \Sigma)$, A and B be constant $d \times d$ symmetric matrices and $\mathrm{tr}(\cdot)$ denote the trace function, then

$$\mathrm{Cov}(\boldsymbol{Z}^\top A\boldsymbol{Z}, \boldsymbol{Z}^\top B\boldsymbol{Z}) = 2\mathrm{tr}(A\Sigma B\Sigma). \qquad (6.10)$$

We begin with the first summand of (6.7).

Lemma 6.3.

$$\begin{aligned}&\int_{\mathbb{R}^d} \left[\frac{1}{2}\mu_{\mathcal{K}} \mathrm{tr}(H\mathcal{H}_{\boldsymbol{z}})\right]^2 d\boldsymbol{z} \\ &\qquad = \frac{1}{16}\mu_{\mathcal{K}}^2 (4\pi)^{-\frac{d}{2}} |\Sigma(\boldsymbol{x})|^{-\frac{1}{2}} \left[2\mathrm{tr}(H\Sigma(\boldsymbol{x})^{-1}H\Sigma(\boldsymbol{x})^{-1}) + \mathrm{tr}(H\Sigma(\boldsymbol{x})^{-1})^2\right].\end{aligned}$$

Remark 6.1. *Lemma 6.2 and some main calculation steps in the proof of Lemma 6.3 are similar to those that Wand (1992) used in the proof of Theorem 1 (see appendix), where they proof for the standard multivariate kernel density estimator an* AMISE *statement, while assuming the real density to be a Gaussian mixture.*

Proof. (Lemma 6.3)

$$\int_{\mathbb{R}^d} \left[\frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\mathbf{z}})\right]^2 d\mathbf{z} = \frac{1}{4}\mu_{\mathcal{K}}^2 \int_{\mathbb{R}^d} \mathrm{tr}(H\mathcal{H}_{\mathbf{z}})^2 d\mathbf{z}.$$

Applying (6.8) in Lemma 6.2 leads to

$$= \frac{1}{4}\mu_{\mathcal{K}}^2 \int_{\mathbb{R}^d} \phi_{\mathbf{0}_d,\Sigma(\mathbf{x})}(\mathbf{z})^2 \mathrm{tr}\left(H\left[\Sigma(\mathbf{x})^{-1}\mathbf{z}\mathbf{z}^\top\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}\right]\right)^2 d\mathbf{z}.$$

Applying (6.9) in Lemma 6.2 leads to

$$= \frac{1}{4}\mu_{\mathcal{K}}^2 \phi_{\mathbf{0}_d,2\Sigma(\mathbf{x})}(\mathbf{0}_d) \int_{\mathbb{R}^d} \phi_{\mathbf{0}_d,\frac{1}{2}\Sigma(\mathbf{x})}(\mathbf{z}) \mathrm{tr}\left(H\left[\Sigma(\mathbf{x})^{-1}\mathbf{z}\mathbf{z}^\top\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}\right]\right)^2 d\mathbf{z}.$$

Let $\tilde{\mathbf{Z}} \sim N(\mathbf{0}_d, \frac{1}{2}\Sigma(\mathbf{x}))$ then

$$= \frac{1}{4}\mu_{\mathcal{K}}^2 \phi_{\mathbf{0}_d,2\Sigma(\mathbf{x})}(\mathbf{0}_d) \mathrm{E}\left[\mathrm{tr}\left(H\left[\Sigma(\mathbf{x})^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}\right]\right)^2\right].$$

Let for notational reasons $\mathrm{tr}(\ldots) := \mathrm{tr}\left(H\left[\Sigma(\mathbf{x})^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}\right]\right)$. Then

$$= \frac{1}{4}\mu_{\mathcal{K}}^2 \phi_{\mathbf{0}_d,2\Sigma(\mathbf{x})}(\mathbf{0}_d) \left\{\mathrm{Cov}(\mathrm{tr}(\ldots), \mathrm{tr}(\ldots)) + \mathrm{E}\left[\mathrm{tr}(\ldots)\right]^2\right\}. \tag{6.11}$$

To show the desired result of Lemma 6.3 we will calculate the terms in (6.11) separately. Since the matrix products are linear functions one gets that

$$\mathrm{E}\left[\mathrm{tr}(\ldots)\right] = \mathrm{tr}\left(H\left[\Sigma(\mathbf{x})^{-1}\mathrm{E}\left[\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top\right]\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}\right]\right),$$

and since $\tilde{\mathbf{Z}} \sim N(\mathbf{0}_d, \frac{1}{2}\Sigma(\mathbf{x}))$ we know that $\mathrm{E}\left[\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top\right] = \frac{1}{2}\Sigma(\mathbf{x})$. Therefore,

$$\mathrm{E}\left[\mathrm{tr}(\ldots)\right] = -\frac{1}{2}\mathrm{tr}\left(H\Sigma(\mathbf{x})^{-1}\right). \tag{6.12}$$

To calculate the Cov term we need to rewrite the $\mathrm{tr}(\ldots)$ term as follows:

$$\mathrm{tr}\left(H\left[\Sigma(\mathbf{x})^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}\right]\right) = \tilde{\mathbf{Z}}^\top\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}\tilde{\mathbf{Z}} - \mathrm{tr}(H\Sigma(\mathbf{x})^{-1}).$$

This can be achieved by multiplying the matrix product out, applying that $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ as well as $\mathrm{tr}(A \cdot B) = \mathrm{tr}(B \cdot A)$ for $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{m\times n}$ and using that $\mathrm{tr}(c) = c$ for $c \in \mathbb{R}$. Therefore we get

$$\mathrm{Cov}(\mathrm{tr}(\ldots), \mathrm{tr}(\ldots)) = \mathrm{Cov}(\tilde{\mathbf{Z}}^\top\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}\tilde{\mathbf{Z}}, \tilde{\mathbf{Z}}^\top\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}\tilde{\mathbf{Z}}).$$

Using then (6.10) from Lemma 6.2 (with $A = B = \Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}$), we get

$$\begin{aligned}\mathrm{Cov}(\mathrm{tr}(\ldots),\mathrm{tr}(\ldots)) &= 2\mathrm{tr}\left(\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}\left\{\frac{1}{2}\Sigma(\mathbf{x})\right\}\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}\left\{\frac{1}{2}\Sigma(\mathbf{x})\right\}\right)\\ &= \frac{1}{2}\mathrm{tr}\left(\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}H\right). \end{aligned} \tag{6.13}$$

Also recall that

$$\phi_{\mathbf{0}_d,2\Sigma(\mathbf{x})}(\mathbf{0}_d) = (4\pi)^{-\frac{d}{2}}|\Sigma(\mathbf{x})|^{-\frac{1}{2}}. \tag{6.14}$$

Finally the desired result of Lemma 6.3 follows directly by plugging (6.12), (6.13) and (6.14) into (6.11). □

By using the representation in Lemma 6.3 one can also calculate the derivative in terms of H:

$$\begin{aligned}&\frac{\partial}{\partial H}\int_{\mathbb{R}^d}\left[\frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\mathbf{z}})\right]^2 d\mathbf{z} \\ &\quad= \frac{1}{4}\mu_{\mathcal{K}}^2(4\pi)^{-\frac{d}{2}}|\Sigma(\mathbf{x})|^{-\frac{1}{2}}\Bigg[2\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1}\circ I_d \\ &\qquad + \mathrm{tr}(H\Sigma(\mathbf{x})^{-1})\left\{\Sigma(\mathbf{x})^{-1} - \frac{1}{2}\Sigma(\mathbf{x})^{-1}\circ I_d\right\}\Bigg],\end{aligned} \tag{6.15}$$

where we used that the derivative of trace functions with respect to a symmetric matrix $M \in \mathbb{R}^{d\times d}$ can be written as $\frac{\partial}{\partial M}\mathrm{tr}(AM) = 2A - A\circ I_d$ and $\frac{\partial}{\partial M}\mathrm{tr}(AMAM) = 4AMA - 2AMA\circ I_d$, when A is a symmetric $\mathbb{R}^{d\times d}$ matrix.

The last part that is needed to calculate the AMISE is the following term of (6.7):

$$2\int_{\mathbb{R}^d}\mu_K\sum_{j=1}^{p}\frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2}\frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\mathbf{z}})d\mathbf{z}. \tag{6.16}$$

By applying the Leibniz rule one can immediately see that the derivative of (6.16) with respect to H is a constant. But it is also worth noting that the calculation of this term appears to be much more challenging then one at first thought.

Remark 6.2. *The interested reader, who wants to use the term (6.16) should consider the following:*

- *The solution of (6.16) and its derivative is quite long. (One can use here Ghazal (1996).)*
- *It requires the ability to estimate first and second derivatives of the covariance matrices $\frac{\partial}{\partial x_j}\Sigma(\boldsymbol{x})$ and $\frac{\partial^2}{\partial^2 x_j}\Sigma(\boldsymbol{x})$.*
- *There will be no obvious solution of (6.4). Therefore we would need to use for example the Newton-Raphson algorithm.*

Since we are looking for a fast calculation of the bandwidth matrix, we will ignore the term (6.16) in our calculation. Confirmation of this procedure is given by the fact that the term is (in simple words) representing the influence of the curvature in terms of $\mathbf{x}$ on our bandwidth matrix. But because we already simplified the AMISE by assuming the joint normality of $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$, one would take a sledgehammer to crack a nut.

By applying the derivative of the integrated AVar (6.6) and our approximation of the squared integrated ABias (6.15) we get

$$\begin{aligned}\frac{\partial}{\partial H}\text{AMISE}\left[\hat{f}_{\mathbf{Z}|\mathbf{X}}, \mathbf{x}, H, b\right] \approx \\ \left(-\frac{1}{2}\right)\frac{R_{\mathcal{K}}\nu_K^p}{nb^p f_{\mathbf{X}}(\mathbf{x})} \cdot |H|^{-\frac{1}{2}}(2H^{-1} - H^{-1} \circ I_d) \\ + \frac{1}{4}\mu_{\mathcal{K}}^2(4\pi)^{-\frac{d}{2}}|\Sigma(\mathbf{x})|^{-\frac{1}{2}}\Big[2\Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1}H\Sigma(\mathbf{x})^{-1} \circ I_d \\ + \operatorname{tr}(H\Sigma(\mathbf{x})^{-1})\left\{\Sigma(\mathbf{x})^{-1} - \frac{1}{2}\Sigma(\mathbf{x})^{-1} \circ I_d\right\}\Big].\end{aligned} \tag{6.17}$$

Our goal is to find a Matrix H_{AMISE} such that the (6.17) is zero. A simple way to do this is by noticing the special structure of the equation in terms of the covariance matrix $\Sigma(\mathbf{x})$. Set

$$H = \Upsilon\Sigma(\mathbf{x}), \tag{6.18}$$

where Υ is representing a term in $\mathbb{R}$. Plugging (6.18) into (6.17) leads to:

$$\begin{aligned}\frac{\partial}{\partial H}\text{AMISE}\left[\hat{f}_{\mathbf{Z}|\mathbf{X}}, \mathbf{x}, \Upsilon\Sigma(\mathbf{x}), b\right] \approx \\ -\frac{1}{2}\frac{R_{\mathcal{K}}\nu_K^p}{nb^p f_{\mathbf{X}}(\mathbf{x})} \cdot \Upsilon^{-\frac{d}{2}}|\Sigma(\mathbf{x})|^{-\frac{1}{2}}\Upsilon^{-1}(2\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1} \circ I_d) \\ + \frac{1}{4}\mu_{\mathcal{K}}^2(4\pi)^{-\frac{d}{2}}|\Sigma(\mathbf{x})|^{-\frac{1}{2}}\Upsilon\Big[2\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1} \circ I_d \\ + \frac{d}{2}\left\{2\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1} \circ I_d\right\}\Big].\end{aligned}$$

Simplifying this leads to

$$\begin{aligned}= \left\{-\frac{1}{2}\frac{R_{\mathcal{K}}\nu_K^p}{nb^p f_{\mathbf{X}}(\mathbf{x})} \cdot |\Sigma(\mathbf{x})|^{-\frac{1}{2}}\Upsilon^{-\frac{d+2}{2}} + \frac{1}{4}\mu_{\mathcal{K}}^2(4\pi)^{-\frac{d}{2}}|\Sigma(\mathbf{x})|^{-\frac{1}{2}}\frac{d+2}{2}\Upsilon\right\} \\ \cdot \left[2\Sigma(\mathbf{x})^{-1} - \Sigma(\mathbf{x})^{-1} \circ I_d\right].\end{aligned} \tag{6.19}$$

We can now see that setting (6.19) to be the zero matrix is equivalent to setting the scalar in front of the matrix expression equal to zero. Let us use the following

notations:

$$\tilde{a} := -\frac{1}{2}\frac{R_{\mathcal{K}}\nu_K^p}{nb^p f_{\mathbf{X}}(\mathbf{x})} \cdot |\Sigma(\mathbf{x})|^{-\frac{1}{2}},$$
$$\tilde{b} := \frac{1}{4}\mu_{\mathcal{K}}^2(4\pi)^{-\frac{d}{2}}|\Sigma(\mathbf{x})|^{-\frac{1}{2}}\frac{d+2}{2}.$$

Therefore setting (6.19) equal to zero is equivalent to:

$$\tilde{a}\Upsilon^{-\frac{d+2}{2}} + \tilde{b}\Upsilon = 0$$
$$\Rightarrow \Upsilon = \left(-\frac{\tilde{b}}{\tilde{a}}\right)^{-\frac{2}{4+d}}. \tag{6.20}$$

Using the definitions of $\tilde{a}$ and $\tilde{b}$ one finally gets the approximated solution of (6.4) by:

$$\hat{H}_{\text{AMISE}} = \left(\frac{(d+2)\mu_{\mathcal{K}}^2(4\pi)^{-\frac{d}{2}}nb^p f_{\mathbf{X}}(\mathbf{x})}{4R_{\mathcal{K}}\nu_K^p}\right)^{-\frac{2}{4+d}}\hat{\Sigma}(\mathbf{x}). \tag{6.21}$$

It is worth noting that this bandwidth is requiring knowledge of the conditional covariance matrix $\Sigma(\mathbf{x})$. One possible way to estimate this conditional covariance matrix is by using the nonparametric procedure described in Section 6.2. For the estimation of $f_{\mathbf{X}}(\mathbf{x})$ one can use an usual multivariate kernel density estimator (see for example Chapter 4 of Wand and Jones (1994)).

It is important to notice that also the bandwidth b is required for the calculation of $\hat{H}_{\text{AMISE}}$. But as Fan et al. (1996) mentioned for given response bandwidth (in our case the H matrix) the rest is just a standard nonparametric problem regressing $\mathcal{K}_H(\mathbf{Z}_i - \mathbf{z})$ on $\mathbf{X}_i$, where several possible methods for estimating the bandwidth b can be found in the literature. In our case we will choose b with a plug-in method derived in Ruppert et al. (1995).

Our approach will therefore be the following: At first we use the rule of thumb bandwidth matrix $\hat{H}_{RT}$, mentioned in (6.3), to establish a bandwidth b and then use this b to calculate the bandwidth matrix $\hat{H}_{\text{AMISE}}$ established in (6.21).

6.2. Estimation of conditional covariance matrix

The following approach for estimating the conditional covariance matrix was introduced by Yin et al. (2010), where they develop a Nadaraya-Watson kernel estimator for $\Sigma(\mathbf{x})$. First recall that the components of the conditional covariance

matrix $\Sigma(\mathbf{x})$ are defined as follows:

$$\begin{aligned}[\Sigma(\mathbf{x})]_{l,k} &:= \mathrm{Cov}(Z_l, Z_k|\mathbf{X} = \mathbf{x}) \\ &= \mathrm{E}[Z_l Z_k|\mathbf{X} = \mathbf{x}] - \mathrm{E}[Z_l|\mathbf{X} = \mathbf{x}]\mathrm{E}[Z_k|\mathbf{X} = \mathbf{x}],\end{aligned}$$

where $(l, k) \in \{1, \ldots, d\}^2$. In our case we know by Lemma 6.1 that the marginal mean of $\mathbf{Z}$ conditioned on $\mathbf{X} = \mathbf{x}$ is 0 and the marginal variance of $\mathbf{Z}$ conditioned on $\mathbf{X} = \mathbf{x}$ is 1. Therefore in our case we have only to compute

$$[\Sigma(\mathbf{x})]_{l,k} = \mathrm{E}[Z_l Z_k|\mathbf{X} = \mathbf{x}],$$

for $(l, k) \in \{1, \ldots, d\}^2$ with $l \neq k$. To estimate this conditional expectation we use the nonparametric local constant estimator, which is also known as the Nadaraya-Watson kernel estimator. Let as usual $\mathbf{z}_i := (z_{i,1}, \ldots, z_{i,d})$ and $\mathbf{x}_i := (x_{i,1}, \ldots, x_{i,p})$ with $i \in \{1, \ldots, n\}$ be the observations to the corresponding response random vector $\mathbf{Z}_i := (Z_{i,1}, \ldots, Z_{i,d})$ and the conditioning random vector $\mathbf{X}_i := (X_{i,1}, \ldots, X_{i,p})$, then the estimator is defined as

$$[\hat{\Sigma}(\mathbf{x})]_{l,k} := \frac{\sum_{i=1}^n K_h(\mathbf{x}_i - \mathbf{x}) z_{i,l} z_{i,k}}{\sum_{i=1}^n K_h(\mathbf{x}_i - \mathbf{x})}, \tag{6.22}$$

where $K_h(\cdot)$ is a product kernel as defined in Section 2.4. Considering the bandwidth h, we will consider (as in Section 2.3 of Yin et al. (2010)) a log likelihood type leave-out-one criterion

$$CV_\Sigma(h) = \frac{1}{n} \sum_{i=1}^n \left[\mathbf{z}_i^\top \hat{\Sigma}_{(-i)}^{-1}(\mathbf{x}_i) \mathbf{z}_i + \log(|\hat{\Sigma}_{(-i)}(\mathbf{x}_i)|) \right],$$

where $\hat{\Sigma}_{(-i)}(\cdot)$ is the estimator of the conditional covariance matrix (6.22) without taking the observation $(\mathbf{z}_i, \mathbf{x}_i)$ into account. The bandwidth h_{CV} for the estimator (6.22) will then be chosen as

$$h_{CV} := \operatorname*{argmin}_{h>0} CV_\Sigma(h).$$

7. Simulation study

The main course of this chapter is to analyze the performance of our conditional copula density estimator under several simulation scenarios. We begin by introducing the simulation setup and the error measure. In the following section we will give an overview of the considered scenarios. Finally we will present the results of the simulation study and provide an analysis and visualization.

7.1. Setup of the simulation study

The entire study was implemented in `R` (R Core Team, 2016). For reasons of illustration we only are considering conditional bivariate copulas ($d = 2$) in this section. To analyze the behavior of our estimator we will generate simulated conditional data, on which our estimator will be applied. This will be done by the following procedure:

- We generate $R = 100$ times
- a sample $\mathbf{x}_i$, $i = 1, \ldots, n$, with sample size $n = 200$ or $n = 1000$
- of a random vector $\mathbf{X} \in \mathbb{R}^p$ with distribution $f_{\mathbf{X}}(\mathbf{x})$,
- then we generated corresponding to every sample $\mathbf{x}_i$ of $\mathbf{X}$
- one sample $\mathbf{u}_i \in \mathbb{R}^2$ of a bivariate conditional copula density $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x}_i)$.

We denote in the following the estimator of the conditional copula density function based on the m'th of the R sample sets (of size n) by $\hat{c}_m(\mathbf{u}|\mathbf{x})$. As described in Chapter 4, this estimator will be derived by (4.2), where the conditional density function in the numerator is calculated by (4.5).

Considering the Kernels we used for the (response) general kernel $\mathcal{K}(\mathbf{z})$ the multivariate standard Gaussian density and for the (condition) product kernel $W(\mathbf{x})$ we used the product of univariate standard Gaussian densities.

As described in Chapter 6, we used for the bandwidth matrix H the rule of thumb bandwidth matrix $\hat{H}_{RT}$ (see (6.3)) as well as $\hat{H}_{\text{AMISE}}$ (see (6.21)). The implementation of the Nadaraya-Watson kernel estimator for the used covariance matrix $\hat{\Sigma}(\mathbf{x})$ (see (6.22) in Section 6.2) employs the `np R-package` (Hayfield and Racine, 2008). For the nonparametric estimation of the joint density function $f_{\mathbf{X}}(\mathbf{x})$ of X in (6.21) the `ks R-package` (Duong, 2016) was used. For the selection

of the bandwidth b we used the `KernSmooth R-package` (Wand, 2015).

Given a condition $\mathbf{x}$ our main error measures will be the integrated absolute error (IAE) and the mean integrated absolute error (MIAE):

$$\begin{aligned}
\mathrm{IAE}\left[\hat{c}_m, \mathbf{x}\right] &:= \int \left|\hat{c}_m(\mathbf{u}|\mathbf{x}) - c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})\right| d\mathbf{u} \qquad \text{for } m \in \{1, \ldots, R\}, \\
\mathrm{MIAE}\left[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}\right] &:= \mathrm{E}\left[\int \left|\hat{c}_m(\mathbf{u}|\mathbf{x}) - c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})\right| d\mathbf{u}\right] \text{ for } m \in \{1, \ldots, R\}.
\end{aligned}$$

Since the R sample sets are created by the same procedure, the theoretical $\mathrm{MIAE}\left[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}\right]$ stays the same for every m.

We will estimate these measures by evaluation on the following grid of values for $\mathbf{u}$: $(u_1, u_2) \in \left\{\left(\frac{l}{11}, \frac{k}{11}\right) : l, k \in \{1, \ldots, 10\}\right\}$. To be more precise we estimate the $\mathrm{IAE}\left[\hat{c}_m, \mathbf{x}\right]$, the $\mathrm{MIAE}\left[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}\right]$ as well as $\sigma_{\widehat{\mathrm{IAE}}[\hat{c}_m, \mathbf{x}]}$ and $\sigma_{\widehat{\mathrm{MIAE}}[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}]}$ (standard deviation of $\widehat{\mathrm{IAE}}\left[\hat{c}_m, \mathbf{x}\right]$ and $\widehat{\mathrm{MIAE}}\left[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}\right]$) empirically by:

$$\begin{aligned}
\widehat{\mathrm{IAE}}\left[\hat{c}_m, \mathbf{x}\right] &:= \frac{1}{100} \sum_{l=1}^{10} \sum_{k=1}^{10} \left|\hat{c}_m\left(\frac{l}{11}, \frac{k}{11} \middle| \mathbf{x}\right) - c_{\mathbf{U}|\mathbf{X}}\left(\frac{l}{11}, \frac{k}{11} \middle| \mathbf{x}\right)\right|, \\
\widehat{\mathrm{MIAE}}\left[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}\right] &:= \frac{1}{R} \sum_{m=1}^{R} \widehat{\mathrm{IAE}}\left[\hat{c}_m, \mathbf{x}\right], \qquad (7.1) \\
\hat{\sigma}_{\widehat{\mathrm{IAE}}[\hat{c}_m, \mathbf{x}]} &:= \sqrt{\frac{1}{R-1} \sum_{m=1}^{R} \left(\widehat{\mathrm{IAE}}\left[\hat{c}_m, \mathbf{x}\right] - \widehat{\mathrm{MIAE}}\left[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}\right]\right)^2}, \\
\hat{\sigma}_{\widehat{\mathrm{MIAE}}[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}]} &:= \frac{\hat{\sigma}_{\widehat{\mathrm{IAE}}[\hat{c}_m, \mathbf{x}]}}{\sqrt{R}}, \qquad (7.2)
\end{aligned}$$

where $\hat{c}_m(\mathbf{u}|\mathbf{x})$ is the estimator based an the m'th of the R sample sets. Recall that $\hat{\sigma}_{\widehat{\mathrm{MIAE}}}$ is a valid estimator because

$$\begin{aligned}
\mathrm{Var}(\widehat{\mathrm{MIAE}}\left[\hat{c}_{\mathbf{U}|\mathbf{X}}, \mathbf{x}\right]) = \mathrm{Var}\left(\frac{1}{R} \sum_{m=1}^{R} \widehat{\mathrm{IAE}}\left[\hat{c}_m, \mathbf{x}\right]\right) &= \frac{1}{R} \mathrm{Var}\left(\widehat{\mathrm{IAE}}\left[\hat{c}_1, \mathbf{x}\right]\right) \\
&\approx \frac{1}{R} \hat{\sigma}^2_{\widehat{\mathrm{IAE}}[\hat{c}_m, \mathbf{x}]},
\end{aligned}$$

where we used that the random variables $\widehat{\mathrm{IAE}}\left[\hat{c}_m, \mathbf{x}\right]$ with $m \in \{1, \ldots, R\}$ are *i.i.d.*, because all of the R sample sets (of size n) are created independently by the same procedure.

A description of the considered sampling scenarios can be found in the next section.

7.2. Simulation scenarios

A summary of all scenarios can be found in Table 7.1. The remainder part of this section will discuss the details of these scenarios as well as provide the reasons for our choices.

	Scenario	$f_{\mathbf{X}}(\mathbf{x})$	$\tau(\mathbf{x})$	$\theta_1(\mathbf{x})$	θ_2	*copula*
$x \in \mathbb{R}$	1	$X \sim N(0,1)$	$\tanh(\frac{x}{2})$	$\sin(\frac{\pi}{2}\tau(x))$	-	Gaussian
	2	$X \sim N(0,1)$	$\tanh(\frac{x}{2})$	$\sin(\frac{\pi}{2}\tau(x))$	3	t-copula
	3	$X \sim N(0,1)$	$\lvert\tanh(\frac{x}{2})\rvert$	$\frac{2\tau(x)}{1-\tau(x)}$	-	Clayton
	4	$X \sim U(-2,2)$	$\tanh(\frac{x}{2})$	$\sin(\frac{\pi}{2}\tau(x))$	-	Gaussian
	5	$X \sim U(-2,2)$	$\tanh(\frac{x}{2})$	$\sin(\frac{\pi}{2}\tau(x))$	3	t-copula
	6	$X \sim U(-2,2)$	$\lvert\tanh(\frac{x}{2})\rvert$	$\frac{2\tau(x)}{1-\tau(x)}$	-	Clayton
$\mathbf{x} \in \mathbb{R}^2$	1	$\mathbf{X} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$	$\tanh(\frac{x_1+x_2}{4})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	-	Gaussian
	2	$\mathbf{X} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$	$\tanh(\frac{x_1+x_2}{4})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	3	t-copula
	3	$\mathbf{X} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$	$\lvert\tanh(\frac{x_1+x_2}{4})\rvert$	$\frac{2\tau(\mathbf{x})}{1-\tau(\mathbf{x})}$	-	Clayton
	4	$\begin{matrix}X_1 \sim U(-2,2)\\X_2 \sim U(-2,2)\end{matrix}, X_1 \perp X_2$	$\tanh(\frac{x_1+x_2}{4})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	-	Gaussian
	5	$\begin{matrix}X_1 \sim U(-2,2)\\X_2 \sim U(-2,2)\end{matrix}, X_1 \perp X_2$	$\tanh(\frac{x_1+x_2}{4})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	3	t-copula
	6	$\begin{matrix}X_1 \sim U(-2,2)\\x_2 \sim U(-2,2)\end{matrix}, X_1 \perp X_2$	$\lvert\tanh(\frac{x_1+x_2}{4})\rvert$	$\frac{2\tau(\mathbf{x})}{1-\tau(\mathbf{x})}$	-	Clayton
	7	$\mathbf{X} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}\right)$	$\tanh(\frac{x_1+x_2}{4})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	-	Gaussian
	8	$\mathbf{X} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}\right)$	$\tanh(\frac{x_1+x_2}{4})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	3	t-copula
	9	$\mathbf{X} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.6\\0.6 & 1\end{pmatrix}\right)$	$\lvert\tanh(\frac{x_1-2x_2}{4})\rvert$	$\frac{2\tau(\mathbf{x})}{1-\tau(\mathbf{x})}$	-	Clayton
$\mathbf{x} \in \mathbb{R}^3$	1	$\begin{matrix}X_1 \sim U(-2,2)\\X_2 \sim U(-2,2)\\X_3 \sim U(-2,2)\end{matrix}, \begin{matrix}X_1 \perp X_2\\X_1 \perp X_3\\X_2 \perp X_3\end{matrix}$	$\tanh(\frac{x_1+x_2+x_3}{6})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	-	Gaussian
	2	$\begin{matrix}X_1 \sim U(-2,2)\\X_2 \sim U(-2,2)\\X_3 \sim U(-2,2)\end{matrix}, \begin{matrix}X_1 \perp X_2\\X_1 \perp X_3\\X_2 \perp X_3\end{matrix}$	$\tanh(\frac{x_1+x_2+x_3}{6})$	$\sin(\frac{\pi}{2}\tau(\mathbf{x}))$	3	t-copula
	3	$\begin{matrix}X_1 \sim U(-2,2)\\X_2 \sim U(-2,2)\\X_3 \sim U(-2,2)\end{matrix}, \begin{matrix}X_1 \perp X_2\\X_1 \perp X_3\\X_2 \perp X_3\end{matrix}$	$\lvert\tanh(\frac{x_1+x_2+x_3}{6})\rvert$	$\frac{2\tau(\mathbf{x})}{1-\tau(\mathbf{x})}$	-	Clayton

Table 7.1.: Scenarios for the simulation study.

The following scenarios can be roughly divided into the situations, where we have $x \in \mathbb{R}$ $(p = 1)$, $\mathbf{x} \in \mathbb{R}^2$ $(p = 2)$ and a small segment with $\mathbf{x} \in \mathbb{R}^3$ $(p = 3)$ (see horizontal areas in Table 7.1).

To analyze the performance in terms of the distribution of $\mathbf{X}$, we will consider cases where $\mathbf{X}$ is uniformly distributed (independent, when $p \geq 2$) and where $\mathbf{X}$ is normally distributed (see column $f_{\mathbf{X}}(\mathbf{x})$ in Table 7.1). The influence of the

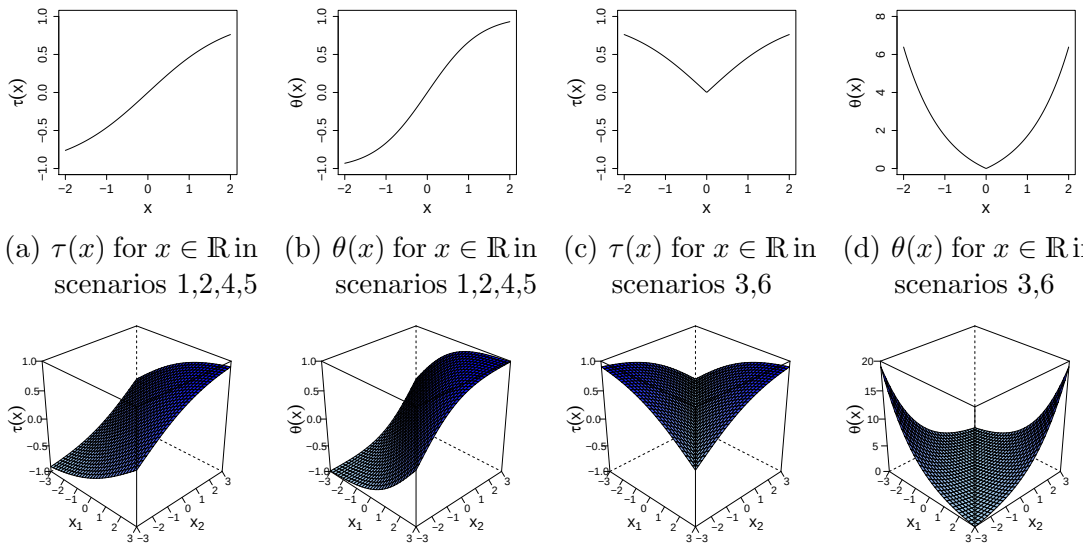
dependence between the components in the $\mathbf{X}$ vector will further be analyzed by the scenarios 7,8 and 9 (for $\mathbf{x} \in \mathbb{R}^2$), where we use multivariate normally distributed $\mathbf{X}$ with correlation $\rho = 0.6$.

We will consider the following 3 different copulas for the upcoming scenarios: The Gaussian copula as an example for a copula with no tail dependence (symmetric), a t-copula as an example for upper and lower tail dependence (symmetric) and a Clayton copula as an example for lower tail dependence (asymmetric). The copula associated with the respective scenarios can be found in the column *copula* of Table 7.1. When we use a t-copula, the degrees of freedom will be $\nu = 3$.

The conditional copula density depends on the realizations of $\mathbf{X}$, only through the Kendall's tau $\tau(\mathbf{x})$ values (see column $\tau(\mathbf{x})$ in Table 7.1).

For the Gaussian and t-copula the Kendall's tau will depend on the realization of $\mathbf{X}$ such that $\tau(\mathbf{x}) = \tanh(\frac{\mathbf{1}^\top \mathbf{x}}{2p})$, where $\mathbf{1}$ is the vector in $\mathbb{R}^p$, where every entry equals 1. Recall here that $\tanh(\cdot)$ is also the inverse of the Fisher Z transformation. In the case of the Clayton copula we will use $\tau(\mathbf{x}) = |\tanh(\frac{\mathbf{1}^\top \mathbf{x}}{2p})|$, because the Kendall's tau values must be positive.

Recall that for the first parameter of the t-copula and the specific parameters of the Gaussian and Clayton copula, there is a direct link between the copula parameter θ and it's Kendall's tau value. These links can be found in the column $\theta_1(\mathbf{x})$ of Table 7.1. Notice here that for the bivariate Gaussian and t-copula the specific parameter $\theta(\mathbf{x})$ is the correlation parameter. The dependence of the Kendall's tau values $\tau(\mathbf{x})$ and the specific copula parameters $\theta_1(\mathbf{x})$ on the conditional vector $\mathbf{x}$ for the scenarios (where $p = 1$ and $p = 2$) is visualized in Figure 7.1.



(a) $\tau(x)$ for $x \in \mathbb{R}$ in scenarios 1,2,4,5 (b) $\theta(x)$ for $x \in \mathbb{R}$ in scenarios 1,2,4,5 (c) $\tau(x)$ for $x \in \mathbb{R}$ in scenarios 3,6 (d) $\theta(x)$ for $x \in \mathbb{R}$ in scenarios 3,6

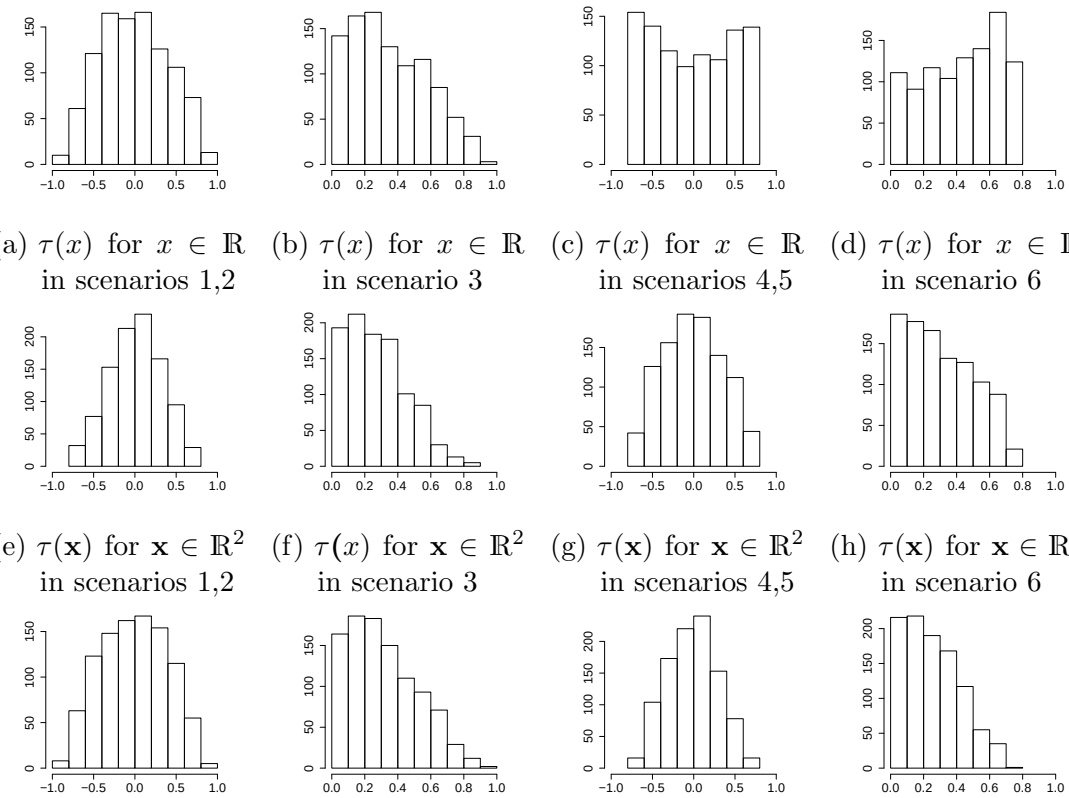
(e) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenarios 1,2,4,5,7,8 (f) $\theta(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenarios 1,2,4,5,7,8 (g) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenarios 3,6,9 (h) $\theta(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenarios 3,6,9

Figure 7.1.: $\tau(\mathbf{x})$ and $\theta(\mathbf{x})$ as a function of $\mathbf{x}$ for scenarios where $x \in \mathbb{R}$ (a)-(d) and $\mathbf{x} \in \mathbb{R}^2$ (e)-(h) (see Table 7.1).

In Figure 7.1 the graphs (a) and (b) represent $\tau(x)$ and $\theta(x)$ as a function

of x in the scenarios 1,2,4,5 where $p = 1$ and the graphs (c) and (d) represent these functions for the Clayton copula scenarios 3 and 6. The graphs (e) and (f) represent $\tau(\mathbf{x})$ and $\theta(\mathbf{x})$ as a function of $\mathbf{x} := (x_1, x_2)$ in the scenarios 1,2,4,5,7,8 where $p = 2$ and the graphs (c) and (d) represent these functions for the Clayton copula scenarios 3,6 and 9 (see Table 7.1 for details).

To get a feeling of how the simulated $\tau(\mathbf{x})$ values are distributed, one can find histograms for the corresponding scenarios in Figure 7.2.



(a) $\tau(x)$ for $x \in \mathbb{R}$ in scenarios 1,2 (b) $\tau(x)$ for $x \in \mathbb{R}$ in scenario 3 (c) $\tau(x)$ for $x \in \mathbb{R}$ in scenarios 4,5 (d) $\tau(x)$ for $x \in \mathbb{R}$ in scenario 6

(e) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenarios 1,2 (f) $\tau(x)$ for $\mathbf{x} \in \mathbb{R}^2$ in scenario 3 (g) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenarios 4,5 (h) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenario 6

(i) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenarios 7,8 (j) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ in scenario 9 (k) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^3$ in scenarios 1,2 (l) $\tau(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^3$ in scenario 3

Figure 7.2.: Histograms of simulated $\tau(\mathbf{x})$ (with $n = 1000$) for the scenarios, where $x \in \mathbb{R}$ (a)-(d), $\mathbf{x} \in \mathbb{R}^2$ (e)-(j) and $\mathbf{x} \in \mathbb{R}^3$ (k)-(l) (see Table 7.1).

At first, one can notice in Figure 7.2 that the histograms for the Clayton copula scenarios (2nd and 4th column) are resulting from their corresponding Gaussian and t-copula scenarios (1st and 3rd column) by adding the negative side mirror image to the positive side. This is unsurprising since we are using for the Clayton copula scenarios the absolute value in $\tau(\mathbf{x})$. The plots (a),(b) are providing the histograms for the scenarios, where X is univariate standard normally distributed, whereas (e),(f),(i),(j) provide those, where $\mathbf{X}$ follows a multivariate normal distribution. Here one can clearly see that due to the positive correlation in the scenarios 7,8,9 ($\mathbf{x} \in \mathbb{R}^2$) the histograms (i),(j) are wider than those of (e),(f).

This was to be expected, because we are using the sum of the components of $\mathbf{x}$ to compute the $\tau(\mathbf{x})$ values and in positive correlation scenarios extremer sums are occurring. The situations were we use uniform distributed $\mathbf{X}$ are represented by the histograms (c),(d),(g),(h),(k) and (l). Recall here for the multivariate cases (g),(h),(k),(l), that we are using the sums to compute the $\tau(\mathbf{x})$'s. These sums however are not uniform distributed and instead are favoring values in the center, which leads to the shapes we are seeing in the histograms (g),(h),(k),(l).

In the next section, we will provide the corresponding results of the conditional copula density estimation.

7.3. Result of simulation study

As mentioned in Section 7.1 we will calculate the estimator of the conditional copula $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ by using the rule of thumb bandwidth matrix $\hat{H}_{RT}$ (see (6.3)) as well as $\hat{H}_{\text{AMISE}}$ (see (6.21)). These estimators will be denoted by $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$ and $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$. Recall also that our error measures of choice are $\widehat{\text{MIAE}}$ (see (7.1)) and $\hat{\sigma}_{\widehat{\text{MIAE}}}$ (see (7.2)).

In order to represent a broad range of different $\tau(\mathbf{x})$ situations, we have chosen for every scenario 9 $\mathbf{x}$-values, on which we will evaluate our estimated conditional copula density (see column $\mathbf{x}$ in Tables 7.2, 7.5, 7.6 and 7.7). Further to guarantee the comparability of the results for one scenario to other scenarios, we have calculated the $\widehat{\text{MIAE}}$ and $\hat{\sigma}_{\widehat{\text{MIAE}}}$ values for the same $\mathbf{x}$-values in each scenario. In all Clayton copula scenarios we omitted the $\mathbf{x}$-values, which led to $\tau(\mathbf{x}) = 0$.

As an indicator for the frequency in which the simulated $\mathbf{x}$-values have occurred near our evaluation $\mathbf{x}$-values, we will use $f_{\mathbf{X}}(\mathbf{x})$.

7.3.1. Result analysis of scenarios where $x \in \mathbb{R}$

We begin by analyzing the results for the scenarios 1-6, where $x \in \mathbb{R}$ ($p = 1$). An overview of these scenarios is given in the first horizontal part of Table 7.1. The calculated $\widehat{\text{MIAE}}$ and $\hat{\sigma}_{\widehat{\text{MIAE}}}$ for these scenarios can be found in Table 7.2.

As expected one can see a significant increase in quality over all scenarios, when the sample size increases from $n = 200$ to $n = 1000$, which is shown both in smaller $\widehat{\text{MIAE}}$ values and in smaller $\hat{\sigma}_{\widehat{\text{MIAE}}}$ values.

Further one can see that an increase of the absolute Kendall's tau values $|\tau(x)|$ results for each scenario in a decreasing performance. Here for $\tau(x)$ between -0.06 and 0.06 the $\widehat{\text{MIAE}}$ are between 0.18 and 0.28 for $n = 200$ (0.11-0.18 for $n = 1000$) whereas for $|\tau(x)|$ values between 0.33 and 0.75 the $\widehat{\text{MIAE}}$ are between 0.21 and 0.76 (0.14-0.60 for $n = 1000$).

In addition we can observe that for small absolute $\tau(x)$ values the scenarios

where X is normally distributed (scenarios 1,2,3) perform better than those were X is uniformly distributed (scenarios 4,5,6). The opposite is the case for big $|\tau(x)|$ values. The reason for this is the distribution of X and the associated frequency in which the simulated x-values occur.A higher occurrence rate (as indicated by $f_X(x)$) leads to a better estimation. For moderate $\tau(x)$ this rate is higher for the scenarios 1,2,3 ($f_X(x) \approx 0.4$) than for the scenarios 4,5,6 ($f_X(x) = 0.25$), for more extreme $\tau(x)$ values the opposite is the case.

Also apparent is, that the conditional copula density estimator has more trouble in the case of the Clayton copula (scenarios 3,6), where we have, compared with the Gaussian and t-copula scenarios (scenarios 1,2,4,5), for almost every $\tau(x)$ situation bigger $\widehat{\text{MIAE}}$ values. This can be explained by the fact that the overall shape of the Clayton copula is harder to estimate, which is also indicated by the asymptotic properties established in Chapter 5. In Corollary 5.3 we see that the second derivatives of $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with respect to $\mathbf{u}$ are affecting the bias (through $\mathcal{H}_{\mathbf{z}}$). Since the integral over this second derivatives will be higher for the Clayton copula, this indicates bigger $\widehat{\text{MIAE}}$ values.

Considering the comparison between the estimator using the H_{RT} bandwidth matrix (see column $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$ in Table 7.2) and the estimator using the H_{AMISE} (see column $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$) we highlighted the better estimation in bold. One can see that in all scenarios where the absolute value of $\tau(x)$ is small, the $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ outperforms the rule of thumb bandwidth approach.

Whereas this behavior is only partly true for bigger absolute values of $\tau(x)$. We notice here that for $|\tau(x)| \geq 0.33$ in scenarios where X is normally distributed (scenarios 1,2,3) the $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$ estimator performs better. In contrary to this, we have for $|\tau(x)| \geq 0.33$ in the scenarios where X is uniformly distributed (scenarios 4,5,6) that $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ is preferable. (Note that in the situations, where the last statements do not hold, the improvement in $\widehat{\text{MIAE}}$ is substantial.) This behavior can be explained by the construction of H_{AMISE} (see (6.21)). Here one can see that the evaluated density $f_{\mathbf{X}}(\mathbf{x})$ occurs in the enumerator. For small $f_{\mathbf{X}}(\mathbf{x})$ values (such in scenarios 1,2,3 for $|\tau(x)| \geq 0.33$) this leads to a big bandwidth and therefore an over-smoothing effect occurs.

					$n = 200$				$n = 1000$			
					$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$	
	x	$f_X(x)$	$\theta(x)$	$\tau(x)$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$
Scenario 1	-1.96	0.058	-0.93	-0.75	0.59	0.018	**0.58**	0.016	**0.36**	0.008	0.41	0.008
	-1.64	0.103	-0.87	-0.68	0.44	0.014	**0.43**	0.012	**0.27**	0.006	0.29	0.006
	-0.67	0.318	-0.49	-0.33	0.24	0.008	**0.22**	0.007	0.15	0.004	**0.14**	0.003
	-0.13	0.396	-0.10	-0.06	0.20	0.005	**0.18**	0.004	0.12	0.004	**0.11**	0.003
	0.00	0.399	0.00	0.00	0.20	0.005	**0.18**	0.004	0.12	0.003	**0.11**	0.003
	0.13	0.396	0.10	0.06	0.20	0.005	**0.18**	0.005	0.12	0.003	**0.11**	0.002
	0.67	0.318	0.49	0.33	0.24	0.006	**0.22**	0.006	0.15	0.004	**0.14**	0.003
	1.64	0.103	0.87	0.68	0.46	0.013	**0.44**	0.011	**0.28**	0.006	0.30	0.007
	1.96	0.058	0.93	0.75	0.60	0.019	**0.58**	0.014	**0.35**	0.007	0.41	0.007
Scenario 2	-1.96	0.058	-0.93	-0.75	0.69	0.077	**0.66**	0.023	**0.40**	0.008	0.48	0.009
	-1.64	0.103	-0.87	-0.68	0.50	0.021	**0.49**	0.012	**0.30**	0.006	0.35	0.007
	-0.67	0.318	-0.49	-0.33	0.25	0.006	**0.24**	0.006	0.16	0.004	**0.15**	0.003
	-0.13	0.396	-0.10	-0.06	0.22	0.006	**0.21**	0.005	0.13	0.003	**0.12**	0.003
	0.00	0.399	0.00	0.00	0.22	0.007	**0.21**	0.005	0.13	0.003	**0.12**	0.002
	0.13	0.396	0.10	0.06	0.22	0.007	**0.21**	0.006	0.13	0.003	**0.12**	0.002
	0.67	0.318	0.49	0.33	0.25	0.007	**0.24**	0.006	0.15	0.004	0.15	0.003
	1.64	0.103	0.87	0.68	**0.45**	0.010	0.48	0.010	**0.30**	0.006	0.34	0.007
	1.96	0.058	0.93	0.75	**0.62**	0.021	0.65	0.013	**0.40**	0.009	0.47	0.010
Scenario 3	-1.96	0.058	6.10	0.75	**0.76**	0.018	0.78	0.013	**0.49**	0.007	0.59	0.008
	-1.64	0.103	4.18	0.68	**0.54**	0.011	0.56	0.009	**0.35**	0.005	0.41	0.005
	-0.67	0.318	0.96	0.33	0.23	0.006	**0.22**	0.005	0.15	0.004	**0.14**	0.003
	-0.13	0.396	0.13	0.06	0.22	0.005	**0.20**	0.005	0.14	0.003	**0.13**	0.003
	0.13	0.396	0.13	0.06	0.22	0.005	**0.20**	0.005	0.14	0.003	**0.13**	0.003
	0.67	0.318	0.96	0.33	0.24	0.006	**0.22**	0.005	0.15	0.003	**0.14**	0.003
	1.64	0.103	4.18	0.68	**0.51**	0.009	0.54	0.008	**0.35**	0.005	0.41	0.006
	1.96	0.058	6.10	0.75	**0.71**	0.029	0.75	0.012	**0.50**	0.008	0.60	0.008
Scenario 4	-1.96	0.25	-0.93	-0.75	0.61	0.020	**0.52**	0.014	0.38	0.010	**0.32**	0.008
	-1.64	0.25	-0.87	-0.68	0.35	0.009	**0.32**	0.008	0.21	0.004	**0.19**	0.004
	-0.67	0.25	-0.49	-0.33	0.25	0.007	**0.22**	0.006	0.16	0.003	**0.14**	0.003
	-0.13	0.25	-0.10	-0.06	0.24	0.007	**0.21**	0.005	0.15	0.004	**0.13**	0.003
	0.00	0.25	0.00	0.00	0.23	0.007	**0.21**	0.006	0.15	0.004	**0.13**	0.003
	0.13	0.25	0.10	0.06	0.24	0.007	**0.21**	0.006	0.15	0.004	**0.13**	0.003
	0.67	0.25	0.49	0.33	0.24	0.007	**0.21**	0.006	0.16	0.004	**0.14**	0.003
	1.64	0.25	0.87	0.68	0.32	0.008	**0.30**	0.007	0.22	0.004	**0.20**	0.004
	1.96	0.25	0.93	0.75	0.57	0.018	**0.48**	0.013	0.40	0.010	**0.33**	0.007
Scenario 5	-1.96	0.25	-0.93	-0.75	0.63	0.017	**0.58**	0.012	0.40	0.008	**0.37**	0.007
	-1.64	0.25	-0.87	-0.68	**0.37**	0.007	0.39	0.007	0.22	0.004	0.22	0.005
	-0.67	0.25	-0.49	-0.33	0.27	0.008	**0.25**	0.007	0.16	0.004	**0.15**	0.003
	-0.13	0.25	-0.10	-0.06	0.25	0.007	**0.22**	0.005	0.15	0.004	**0.14**	0.004
	0.00	0.25	0.00	0.00	0.25	0.007	**0.22**	0.006	0.16	0.004	**0.14**	0.004
	0.13	0.25	0.10	0.06	0.25	0.007	**0.22**	0.006	0.16	0.004	**0.14**	0.004
	0.67	0.25	0.49	0.33	0.25	0.007	**0.23**	0.006	0.16	0.004	**0.15**	0.004
	1.64	0.25	0.87	0.68	**0.37**	0.008	0.38	0.009	0.21	0.004	0.21	0.005
	1.96	0.25	0.93	0.75	0.63	0.013	**0.59**	0.012	0.39	0.008	**0.36**	0.008
Scenario 6	-1.96	0.25	6.10	0.75	0.71	0.018	**0.68**	0.013	**0.46**	0.008	0.47	0.008
	-1.64	0.25	4.18	0.68	**0.42**	0.007	0.45	0.007	**0.26**	0.005	0.29	0.005
	-0.67	0.25	0.96	0.33	0.27	0.008	**0.24**	0.006	0.16	0.004	**0.15**	0.003
	-0.13	0.25	0.13	0.06	0.27	0.007	**0.23**	0.006	0.17	0.004	**0.15**	0.003
	0.13	0.25	0.13	0.06	0.28	0.008	**0.24**	0.007	0.18	0.004	**0.15**	0.003
	0.67	0.25	0.96	0.33	0.29	0.009	**0.26**	0.007	0.17	0.004	**0.15**	0.003
	1.64	0.25	4.18	0.68	**0.42**	0.009	0.44	0.009	**0.27**	0.004	0.30	0.004
	1.96	0.25	6.10	0.75	0.70	0.020	**0.68**	0.016	**0.46**	0.008	0.47	0.008

Table 7.2.: Result of scenarios where $x \in \mathbb{R}$ (see Table 7.1).

As a plausibility test we compere the conditional copula estimator $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$ with an ordinary nonparametric copula estimator $\hat{c}(\mathbf{u})$, which does not use the condition x. We have chosen for $\hat{c}(\mathbf{u})$ the transformation copula density estimator as it is described in Section 3.4 of Nagler (2014). It was found that the results lead to the same conclusions, across all scenarios, therefore we will only provide those for scenario 1. The calculated respective $\widehat{\text{MIAE}}$ and $\hat{\sigma}_{\widehat{\text{MIAE}}}$ for this comparison can be seen in Table 7.3. Here at first one notices immediately that the unconditional estimator for our simulation scenario is not consistent, because it does not perform better when the sample size n increases. The conditional estimator outperforms the unconditional one, in the cases where we have a significant big absolute value for $\tau(x)$. This is unsurprising, because the used model is highly dependent on x and $\hat{c}(\mathbf{u})$ does not take this into account. In the case of moderate $\tau(x)$ values the performance of the two estimators is quite comparable for $n = 1000$ and for $n = 200$ the unconditional estimator is even better than the conditional one. This was to be expected, because the estimator $\hat{c}(\mathbf{u})$ provides an average estimation (over X) of the conditional copula density. Since this average roughly coincides here with the conditional copula density in the case, where $\tau(x)$ is equal 0, we get a good performance of $\hat{c}(\mathbf{u})$ there.

					$n = 200$				$n = 1000$			
					$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}(\mathbf{u})$		$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}(\mathbf{u})$	
	x	$f_X(x)$	$\theta(x)$	$\tau(x)$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$
Scenario 1	-1.96	0.058	-0.93	-0.75	0.59	0.018	1.03	0.004	0.36	0.008	1.02	0.002
	-1.64	0.103	-0.87	-0.68	0.44	0.014	0.83	0.004	0.27	0.006	0.82	0.002
	-0.67	0.318	-0.49	-0.33	0.24	0.008	0.26	0.004	0.15	0.004	0.26	0.002
	-0.13	0.396	-0.10	-0.06	0.20	0.005	0.12	0.003	0.12	0.004	0.12	0.002
	0.00	0.399	0.00	0.00	0.20	0.005	0.11	0.002	0.12	0.003	0.11	0.001
	0.13	0.396	0.10	0.06	0.20	0.005	0.12	0.003	0.12	0.003	0.12	0.002
	0.67	0.318	0.49	0.33	0.24	0.006	0.26	0.004	0.15	0.004	0.26	0.002
	1.64	0.103	0.87	0.68	0.46	0.013	0.83	0.004	0.28	0.006	0.82	0.002
	1.96	0.058	0.93	0.75	0.60	0.019	1.03	0.004	0.35	0.007	1.02	0.002

Table 7.3.: Comparison between $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$ and $\hat{c}(\mathbf{u})$ (not using x) for scenario 1 in the case where $x \in \mathbb{R}$ (see Table 7.1).

To provide a feeling for the sensitivity of our estimator on the bandwidth parameter b, we calculated the $\widehat{\text{MIAE}}$ for shifted b (by 0.1). Again we provide only the results for the scenario 1. For the estimator we used $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$. These results can be found in Table 7.4. The columns denoted by b are representing the results of our estimator, when we use the bandwidth b of the proposed method (see end of Section 6.1). The results contained in the columns $b + 0.1$ and $b - 0.1$ are derived by the same way but using for the bandwidth instead of the before mentioned b the respective shifted version. Note that in the case of $R = 100$ the standard deviation of b was 0.12 for $n = 200$ and 0.09 for $n = 1000$. A histogram of the calculated b's in scenario 1 (with $R = 100$ and $n = 200$) can be seen in Figure 7.3.

As known for kernel estimators in general the performance depends highly on the choice of bandwidth. This can be also observed in Table 7.4, especially taking the estimated standard deviation $\hat{\sigma}_{\widehat{\text{MIAE}}}$ into account. The results reveal also that our bandwidth b tends to undersmooth, because the estimator performs better, when we use $b + 0.1$. At the first look this seems counterintuitive since plug-in estimator are more know to oversmooth. But as Loader (1999) reveals, the bandwidth selection as described by Ruppert et al. (1995) (which we use) tends to undersmooth, which explains the results we are seeing in Table 7.4.

					$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$ for $n = 200$				$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$ for $n = 1000$			
					b		$b+0.1$	$b-0.1$	b		$b+0.1$	$b-0.1$
	x	$f_X(x)$	$\theta(x)$	$\tau(x)$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\widehat{\text{MIAE}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\widehat{\text{MIAE}}$
Scenario 1	-1.96	0.058	-0.93	-0.75	0.57	0.017	0.54	0.64	0.36	0.008	0.35	0.40
	-1.64	0.103	-0.87	-0.68	0.43	0.012	0.41	0.48	0.27	0.006	0.26	0.29
	-0.67	0.318	-0.49	-0.33	0.22	0.006	0.21	0.27	0.14	0.003	0.12	0.16
	-0.13	0.396	-0.10	-0.06	0.20	0.006	0.18	0.23	0.12	0.003	0.11	0.15
	0.00	0.399	0.00	0.00	0.19	0.005	0.18	0.23	0.12	0.003	0.11	0.15
	0.13	0.396	0.10	0.06	0.19	0.006	0.18	0.24	0.13	0.003	0.11	0.15
	0.67	0.318	0.49	0.33	0.22	0.006	0.21	0.27	0.14	0.004	0.13	0.17
	1.64	0.103	0.87	0.68	0.42	0.013	0.40	0.46	0.27	0.006	0.25	0.30
	1.96	0.058	0.93	0.75	0.55	0.014	0.53	0.61	0.34	0.007	0.33	0.38

Table 7.4.: Sensitivity of $\widehat{\text{MIAE}}$ using $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$ on b for scenario 1 in the case where $x \in \mathbb{R}$ (see Table 7.1).

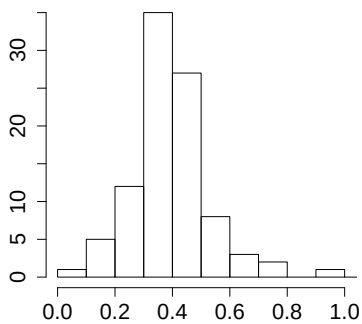
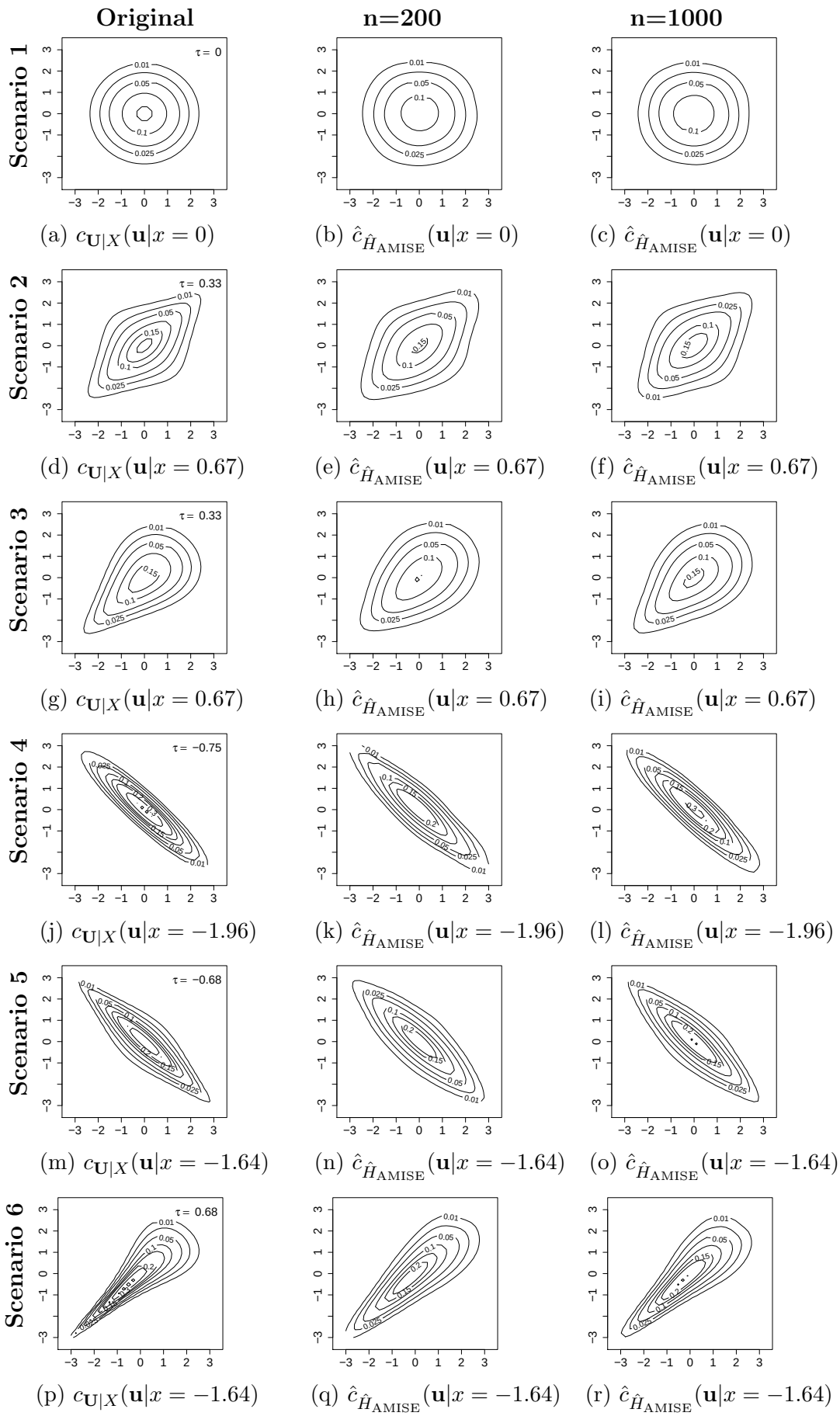


Figure 7.3.: Histogram of calculated bandwidths b for $R = 100$ and $n = 200$ for scenario 1 in the case where $x \in \mathbb{R}$ (see Table 7.1).

To illustrate the performance of our estimator visually, Figure 7.4 and Figure 7.5 are providing the comparison between contour/surface plots of the original copula density $c_{\mathbf{U}|X}(\mathbf{u}|x)$ and the average of the estimation $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x)$. Here the average was calculated over $R = 100$ estimations. The expected value of the estimator is equal to the true conditional copula plus the bias. Since the average is an approximation of the expected value this visualizes the bias of the estimator. Note that for the contour plots we have transformed the densities to the Z-space (Gaussian scale). Due to the immense variety of scenarios and conditions, we have decided to provide only plots for one condition x for every scenario.



(a) $c_{\mathbf{U}|X}(\mathbf{u}|x=0)$ (b) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0)$ (c) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0)$

(d) $c_{\mathbf{U}|X}(\mathbf{u}|x=0.67)$ (e) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$ (f) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$

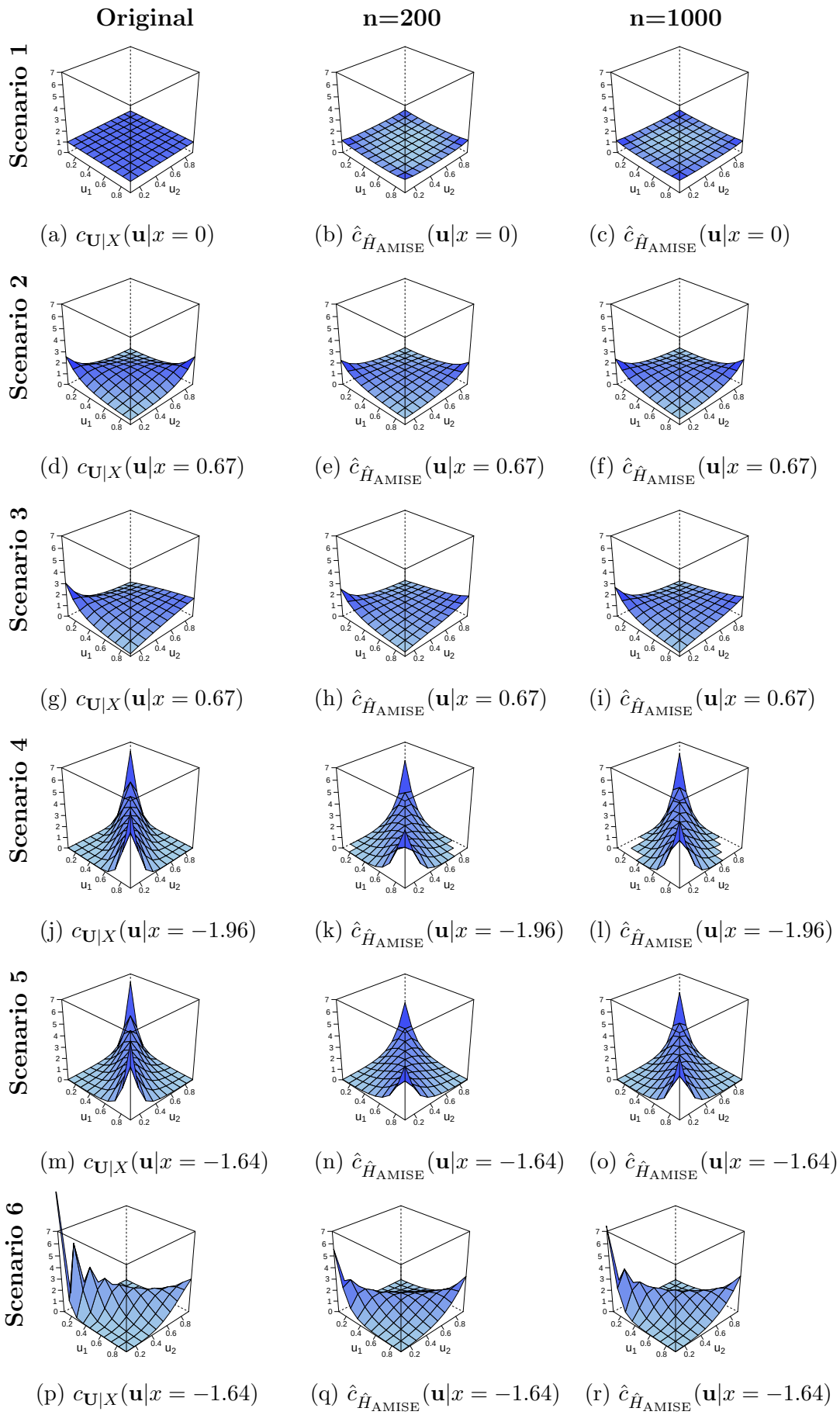
(g) $c_{\mathbf{U}|X}(\mathbf{u}|x=0.67)$ (h) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$ (i) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$

(j) $c_{\mathbf{U}|X}(\mathbf{u}|x=-1.96)$ (k) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.96)$ (l) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.96)$

(m) $c_{\mathbf{U}|X}(\mathbf{u}|x=-1.64)$ (n) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$ (o) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$

(p) $c_{\mathbf{U}|X}(\mathbf{u}|x=-1.64)$ (q) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$ (r) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$

Figure 7.4.: Comparison between contour plots (on Z-space) of original and average conditional copula density estimation for $x \in \mathbb{R}$ scenarios (see Table 7.1).



(a) $c_{\mathbf{U}|X}(\mathbf{u}|x=0)$ (b) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0)$ (c) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0)$

(d) $c_{\mathbf{U}|X}(\mathbf{u}|x=0.67)$ (e) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$ (f) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$

(g) $c_{\mathbf{U}|X}(\mathbf{u}|x=0.67)$ (h) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$ (i) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=0.67)$

(j) $c_{\mathbf{U}|X}(\mathbf{u}|x=-1.96)$ (k) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.96)$ (l) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.96)$

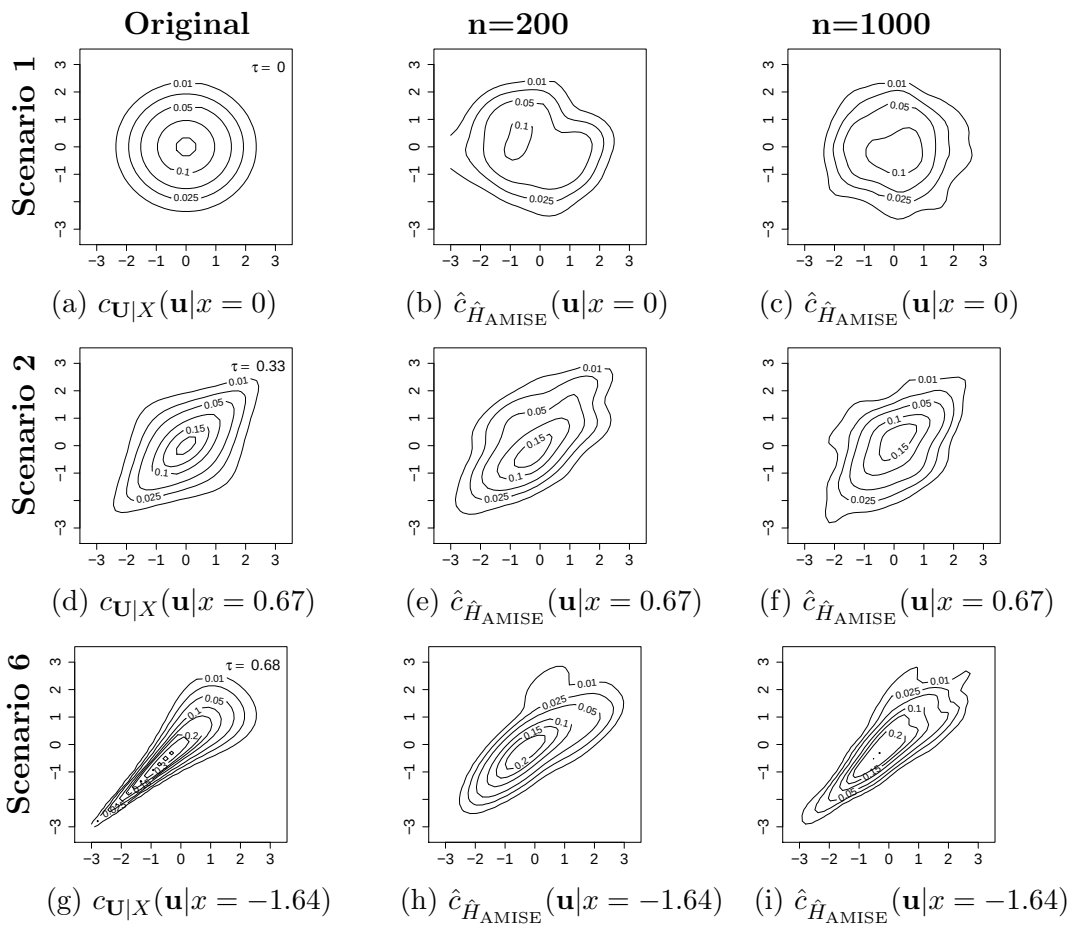
(m) $c_{\mathbf{U}|X}(\mathbf{u}|x=-1.64)$ (n) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$ (o) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$

(p) $c_{\mathbf{U}|X}(\mathbf{u}|x=-1.64)$ (q) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$ (r) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x=-1.64)$

Figure 7.5.: Comparison between surface plots of original and average conditional copula density estimation for $x \in \mathbb{R}$ scenarios (see Table 7.1).

In Figure 7.4 and Figure 7.5 one can find the true conditional copulas, in the first column, whereas the average estimations for $n = 200$ can be found in the second and for $n = 1000$ in the third column. One can clearly see the increase in performance for increasing sample size n. Further one sees that our estimator is interpreting the dependence on x in terms of the correlation parameter right.

Recall that Figure 7.4 and Figure 7.5 are showing average estimations. Therefore the plots appear much smoother than an actual estimation would. To get a feeling for how actual conditional estimations may look like, one can find examples in Figure 7.6, where we compare contour plots (on Z-space) of the original and the example conditional copula density estimator. As one would expect, these contours are not as smooth as those of Figure 7.4. Nevertheless one sees that the estimation is interpreting the overall shape and the dependence on x (i.e. symmetry and correlation) right.



(a) $c_{\mathbf{U}|X}(\mathbf{u}|x = 0)$ (b) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x = 0)$ (c) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x = 0)$

(d) $c_{\mathbf{U}|X}(\mathbf{u}|x = 0.67)$ (e) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x = 0.67)$ (f) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x = 0.67)$

(g) $c_{\mathbf{U}|X}(\mathbf{u}|x = -1.64)$ (h) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x = -1.64)$ (i) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|x = -1.64)$

Figure 7.6.: Comparison between contour plots (on Z-space) of original and examples of conditional copula density estimations for $x \in \mathbb{R}$ scenarios (see Table 7.1).

In the next section we will discuss the performance of the estimator, when using two and three dimensional conditions $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{x} \in \mathbb{R}^3$.

7.3.2. Result analysis of scenarios where $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{x} \in \mathbb{R}^3$

In this section we analyze the results of scenarios 1-9, where $\mathbf{x} \in \mathbb{R}^2$ ($p = 2$) and scenarios 1-3, where $\mathbf{x} \in \mathbb{R}^3$ ($p = 3$). Recall that an overview of these scenarios is given in the second and third horizontal parts of Table 7.1. The calculated $\widehat{\text{MIAE}}$ and $\hat{\sigma}_{\widehat{\text{MIAE}}}$ for these scenarios can be found in Table 7.5 (scenarios 1-6, $\mathbf{x} \in \mathbb{R}^2$), Table 7.6 (scenarios 7-9, $\mathbf{x} \in \mathbb{R}^2$) and Table 7.7 (scenarios 1-3, $\mathbf{x} \in \mathbb{R}^3$).

The results reveal a lot of similarities to those where $x \in \mathbb{R}$. As in the case of $x \in \mathbb{R}$ we also observe that bigger n leads to an significant increase in performance. Further we can see again that in the Clayton copula cases (scenarios 3,6,9 for $\mathbf{x} \in \mathbb{R}^2$ and scenario 3 for $\mathbf{x} \in \mathbb{R}^3$) typically the estimator performs worse than in the Gaussian and t-copula cases (scenarios 1,2,4,5,7,8 for $\mathbf{x} \in \mathbb{R}^2$ and scenarios 1,2 for $\mathbf{x} \in \mathbb{R}^3$).
Contrary to the case of $x \in \mathbb{R}$ we observe here, that smaller $|\tau(\mathbf{x})|$ values do not automatically imply a decrease in $\widehat{\text{MIAE}}$. This is due to the fact, that the probabilities of $\mathbf{x}$ regions are not proportional to $|\tau(\mathbf{x})|$ as it was in case of $x \in \mathbb{R}$. For example one can see in scenario 1 of Table 7.5, that a very moderate $\tau(\mathbf{x}) = 0.09$ performs significantly worse than a $\tau(\mathbf{x}) = 0.45$. Note that the probability of the occurring $\mathbf{x}$-values, causing this results is indicated by the $f_{\mathbf{X}}(\mathbf{x})$ values.
The $f_{\mathbf{X}}(\mathbf{x})$ influence also gets confirmed by comparing the different distributions of $\mathbf{X}$ in $\mathbb{R}^2$. Here for nearly all evaluation $\mathbf{x}$ points the estimator performs better in the scenarios where $\mathbf{X}$ is uniform (scenarios 4,5,6) compared to those where $\mathbf{X}$ is standard normally distributed (scenarios 1,2,3). An exception to this, is naturally the points where $\tau(\mathbf{x}) = 0$, where $f_{\mathbf{X}}(\mathbf{x})$ is bigger for the scenarios 1,2 and 3.
If one compares the standard normally distributed scenarios 1,2,3 (see Table 7.5) with the scenarios 7,8,9 (see Table 7.6), which have positive correlation, one notices that the correlation is influencing the results through a change of the $f_{\mathbf{X}}(\mathbf{x})$ values. Here the results for the evaluation points with $|\tau(\mathbf{x})| \in \{0.45, 0.62\}$ are better in the scenarios 7,8,9, whereas for $|\tau(\mathbf{x})| = 0.09$ the estimations performs significantly worse than in the scenarios 1,2,3.
If the density of $\mathbf{X}$ has bounded support, the estimator will put for $\mathbf{x}$ close to the boundary a considerable amount of probability mass outside the support. This leads to severe bias if the evaluating $\mathbf{x}$-vector is close to the boundary. We can observe this in the scenarios where $\mathbf{X}$ is uniform (scenarios 4-6, $\mathbf{x} \in \mathbb{R}^2$). Here due to this boundary effect the estimator performs significantly worse for $|\tau(\mathbf{x})| = 0.09$ than for $|\tau(\mathbf{x})| = 0.15$.
Considering the comparison between the estimators $\hat{c}_{\hat{H}_{RT}}(\mathbf{u}|\mathbf{x})$ and $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ one notices that the estimator using the $\hat{H}_{\text{AMISE}}$ bandwidth matrix (see (6.21)) gets the lead in nearly all situations. As in the case of $x \in \mathbb{R}$ we highlighted the better estimations in Tables 7.5, 7.6 and 7.7 in bold.
As it is typical in the nonparametric setting the estimator has more trouble as the dimensions of $\mathbf{x}$ increases. This can be noticed by comparing the results of the Tables 7.5 and 7.6, with those of Table 7.7 and taking the respective $\tau(\mathbf{x})$ into account.

						$n = 200$				$n = 1000$			
						$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$	
	x_1	x_2	$f_{\mathbf{X}}(\mathbf{x})$	$\theta(\mathbf{x})$	$\tau(\mathbf{x})$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$
Scenario 1	-1.64	-1.28	0.018	-0.83	-0.62	0.85	0.049	**0.59**	0.026	0.47	0.015	**0.36**	0.010
	-1.28	-0.67	0.056	-0.65	-0.45	0.49	0.027	**0.37**	0.014	0.27	0.008	**0.22**	0.006
	0.67	-1.28	0.056	-0.23	-0.15	0.52	0.038	**0.35**	0.023	0.27	0.008	**0.19**	0.005
	-1.64	1.28	0.018	-0.14	-0.09	0.87	0.045	**0.49**	0.026	0.48	0.016	**0.29**	0.011
	0.00	0.00	0.159	0.00	0.00	0.28	0.015	**0.22**	0.010	0.15	0.005	**0.13**	0.003
	-1.28	1.64	0.018	0.14	0.09	0.84	0.039	**0.46**	0.019	0.47	0.015	**0.28**	0.009
	1.28	-0.67	0.056	0.23	0.15	0.55	0.066	**0.35**	0.015	0.27	0.009	**0.19**	0.006
	0.67	1.28	0.056	0.65	0.45	0.50	0.031	**0.36**	0.016	0.27	0.009	**0.22**	0.005
	1.28	1.64	0.018	0.83	0.62	0.84	0.075	**0.57**	0.024	0.46	0.015	**0.38**	0.010
Scenario 2	-1.64	-1.28	0.018	-0.83	-0.62	0.88	0.059	**0.66**	0.026	0.50	0.021	**0.44**	0.012
	-1.28	-0.67	0.056	-0.65	-0.45	0.47	0.018	**0.39**	0.011	0.30	0.011	**0.26**	0.006
	0.67	-1.28	0.056	-0.23	-0.15	0.49	0.022	**0.35**	0.015	0.29	0.013	**0.22**	0.009
	-1.64	1.28	0.018	-0.14	-0.09	0.82	0.037	**0.47**	0.020	0.53	0.026	**0.32**	0.011
	0.00	0.00	0.159	0.00	0.00	0.29	0.015	**0.24**	0.009	0.15	0.005	**0.14**	0.004
	-1.28	1.64	0.018	0.14	0.09	0.84	0.036	**0.47**	0.019	0.52	0.018	**0.32**	0.011
	1.28	-0.67	0.056	0.23	0.15	0.47	0.019	**0.34**	0.013	0.29	0.012	**0.21**	0.006
	0.67	1.28	0.056	0.65	0.45	0.50	0.027	**0.41**	0.015	0.29	0.010	**0.25**	0.005
	1.28	1.64	0.018	0.83	0.62	0.92	0.070	**0.67**	0.043	0.49	0.016	**0.43**	0.009
Scenario 3	-1.64	-1.28	0.018	3.32	0.62	0.82	0.031	**0.67**	0.017	0.51	0.019	**0.48**	0.011
	-1.28	-0.67	0.056	1.66	0.45	0.47	0.022	**0.39**	0.014	0.30	0.014	**0.25**	0.006
	0.67	-1.28	0.056	0.35	0.15	0.48	0.017	**0.33**	0.012	0.29	0.013	**0.20**	0.007
	-1.64	1.28	0.018	0.20	0.09	0.91	0.071	**0.51**	0.032	0.51	0.025	**0.29**	0.011
	-1.28	1.64	0.018	0.20	0.09	0.79	0.032	**0.45**	0.018	0.49	0.016	**0.29**	0.009
	1.28	-0.67	0.056	0.35	0.15	0.47	0.017	**0.31**	0.010	0.29	0.014	**0.20**	0.008
	0.67	1.28	0.056	1.66	0.45	0.49	0.023	**0.39**	0.012	0.30	0.014	**0.26**	0.008
	1.28	1.64	0.018	3.32	0.62	0.82	0.032	**0.65**	0.015	0.52	0.019	**0.48**	0.008
Scenario 4	-1.64	-1.28	0.062	-0.83	-0.62	0.59	0.025	**0.45**	0.013	0.35	0.011	**0.27**	0.008
	-1.28	-0.67	0.062	-0.65	-0.45	0.54	0.074	**0.38**	0.039	0.27	0.010	**0.21**	0.006
	0.67	-1.28	0.062	-0.23	-0.15	0.48	0.030	**0.34**	0.023	0.27	0.012	**0.19**	0.006
	-1.64	1.28	0.062	-0.14	-0.09	0.58	0.033	**0.37**	0.019	0.33	0.009	**0.23**	0.007
	0.00	0.00	0.062	0.00	0.00	0.43	0.024	**0.30**	0.016	0.27	0.012	**0.19**	0.007
	-1.28	1.64	0.062	0.14	0.09	0.58	0.022	**0.36**	0.012	0.33	0.010	**0.22**	0.008
	1.28	-0.67	0.062	0.23	0.15	0.47	0.022	**0.32**	0.013	0.27	0.012	**0.19**	0.007
	0.67	1.28	0.062	0.65	0.45	0.50	0.025	**0.38**	0.015	0.28	0.012	**0.21**	0.007
	1.28	1.64	0.062	0.83	0.62	0.60	0.026	**0.46**	0.014	0.35	0.011	**0.27**	0.008
Scenario 5	-1.64	-1.28	0.062	-0.83	-0.62	0.62	0.023	**0.50**	0.012	0.36	0.008	**0.30**	0.007
	-1.28	-0.67	0.062	-0.65	-0.45	0.51	0.021	**0.39**	0.012	0.29	0.010	**0.23**	0.006
	0.67	-1.28	0.062	-0.23	-0.15	0.48	0.023	**0.33**	0.014	0.26	0.009	**0.20**	0.006
	-1.64	1.28	0.062	-0.14	-0.09	0.60	0.021	**0.41**	0.014	0.34	0.010	**0.24**	0.007
	0.00	0.00	0.062	0.00	0.00	0.48	0.021	**0.34**	0.011	0.27	0.011	**0.21**	0.007
	-1.28	1.64	0.062	0.14	0.09	0.59	0.019	**0.37**	0.011	0.33	0.009	**0.23**	0.006
	1.28	-0.67	0.062	0.23	0.15	0.53	0.034	**0.36**	0.019	0.28	0.011	**0.21**	0.006
	0.67	1.28	0.062	0.65	0.45	0.50	0.021	**0.39**	0.013	0.28	0.009	**0.23**	0.006
	1.28	1.64	0.062	0.83	0.62	0.65	0.030	**0.52**	0.015	0.36	0.009	**0.31**	0.007
Scenario 6	-1.64	-1.28	0.062	3.32	0.62	0.66	0.034	**0.58**	0.022	0.37	0.009	**0.36**	0.007
	-1.28	-0.67	0.062	1.66	0.45	0.50	0.023	**0.39**	0.014	0.27	0.007	**0.23**	0.005
	0.67	-1.28	0.062	0.35	0.15	0.51	0.023	**0.33**	0.013	0.26	0.009	**0.19**	0.006
	-1.64	1.28	0.062	0.20	0.09	0.65	0.047	**0.39**	0.024	0.33	0.009	**0.22**	0.006
	-1.28	1.64	0.062	0.20	0.09	0.65	0.031	**0.40**	0.017	0.33	0.008	**0.22**	0.006
	1.28	-0.67	0.062	0.35	0.15	0.51	0.026	**0.34**	0.014	0.26	0.009	**0.19**	0.006
	0.67	1.28	0.062	1.66	0.45	0.53	0.027	**0.40**	0.014	0.27	0.008	**0.23**	0.005
	1.28	1.64	0.062	3.32	0.62	0.66	0.033	**0.57**	0.018	0.36	0.007	**0.35**	0.006

Table 7.5.: Result of scenarios 1-6, where $\mathbf{x} \in \mathbb{R}^2$ (see Table 7.1).

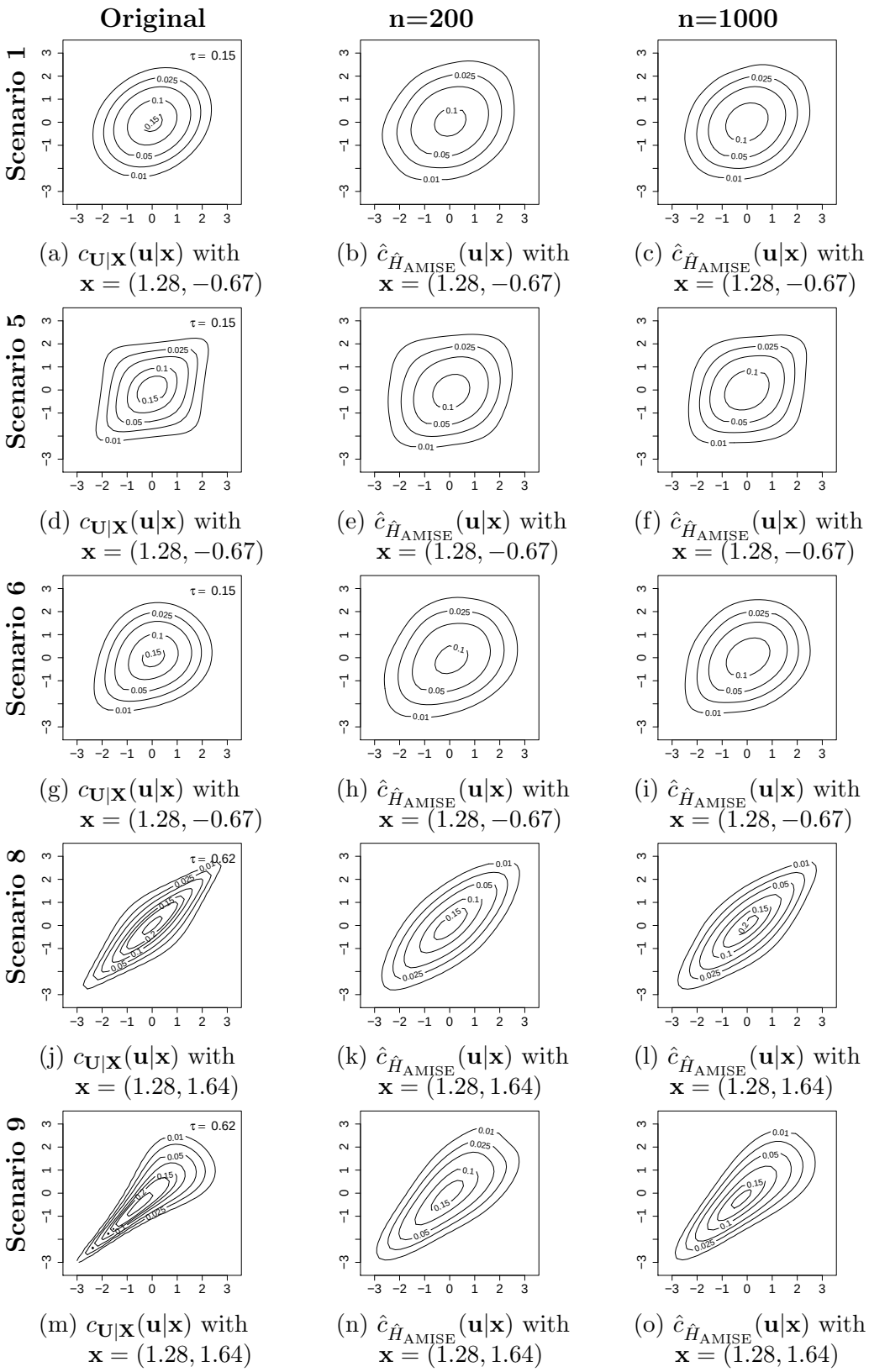
						$n = 200$				$n = 1000$			
						$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$	
	x_1	x_2	$f_X(\mathbf{x})$	$\theta(\mathbf{x})$	$\tau(\mathbf{x})$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$
Scenario 7	-1.64	-1.28	0.048	-0.83	-0.62	0.51	0.017	**0.45**	0.012	0.31	0.013	**0.29**	0.008
	-1.28	-0.67	0.087	-0.65	-0.45	0.36	0.009	**0.29**	0.008	0.24	0.010	**0.20**	0.007
	0.67	-1.28	0.017	-0.23	-0.15	0.78	0.029	**0.43**	0.016	0.51	0.019	**0.29**	0.010
	-1.64	1.28	0.001	-0.14	-0.09	2.30	0.181	**0.65**	0.022	1.56	0.136	**0.49**	0.026
	0.00	0.00	0.199	0.00	0.00	0.24	0.007	**0.20**	0.006	0.16	0.013	**0.14**	0.009
	-1.28	1.64	0.001	0.14	0.09	2.12	0.112	**0.66**	0.025	1.79	0.258	**0.51**	0.027
	1.28	-0.67	0.017	0.23	0.15	0.79	0.030	**0.45**	0.016	0.54	0.030	**0.29**	0.014
	0.67	1.28	0.087	0.65	0.45	0.39	0.012	**0.31**	0.008	0.26	0.014	**0.22**	0.009
	1.28	1.64	0.048	0.83	0.62	0.53	0.017	**0.46**	0.012	0.33	0.015	**0.30**	0.009
Scenario 8	-1.64	-1.28	0.048	-0.83	-0.62	0.64	0.055	**0.54**	0.025	0.33	0.012	0.33	0.009
	-1.28	-0.67	0.087	-0.65	-0.45	0.43	0.018	**0.37**	0.010	0.26	0.011	**0.23**	0.007
	0.67	-1.28	0.017	-0.23	-0.15	0.98	0.059	**0.53**	0.025	0.52	0.026	**0.32**	0.014
	-1.64	1.28	0.001	-0.14	-0.09	2.77	0.293	**0.76**	0.059	1.39	0.081	**0.51**	0.021
	0.00	0.00	0.199	0.00	0.00	0.27	0.012	**0.23**	0.007	0.15	0.007	**0.14**	0.004
	-1.28	1.64	0.001	0.14	0.09	2.60	0.209	**0.72**	0.037	1.63	0.193	**0.58**	0.046
	1.28	-0.67	0.017	0.23	0.15	0.89	0.047	**0.49**	0.023	0.51	0.032	**0.32**	0.020
	0.67	1.28	0.087	0.65	0.45	0.44	0.019	**0.37**	0.012	0.26	0.011	**0.23**	0.007
	1.28	1.64	0.048	0.83	0.62	0.59	0.034	**0.53**	0.028	0.34	0.012	0.34	0.007
Scenario 9	-1.64	-1.28	0.048	3.32	0.62	0.56	0.029	**0.53**	0.013	**0.37**	0.012	0.39	0.007
	-1.28	-0.67	0.087	1.66	0.45	0.40	0.013	**0.35**	0.009	0.26	0.009	**0.23**	0.006
	0.67	-1.28	0.017	0.35	0.15	0.85	0.031	**0.49**	0.021	0.59	0.045	**0.33**	0.020
	-1.64	1.28	0.001	0.20	0.09	2.31	0.166	**0.67**	0.029	1.77	0.156	**0.56**	0.026
	-1.28	1.64	0.001	0.20	0.09	2.65	0.261	**0.69**	0.033	1.77	0.148	**0.55**	0.031
	1.28	-0.67	0.017	0.35	0.15	0.92	0.103	**0.49**	0.041	0.56	0.026	**0.31**	0.014
	0.67	1.28	0.087	1.66	0.45	0.40	0.015	**0.35**	0.009	0.26	0.011	**0.24**	0.007
	1.28	1.64	0.048	3.32	0.62	0.55	0.016	**0.54**	0.011	**0.37**	0.010	0.39	0.006

Table 7.6.: Result of scenarios 7,8 and 9, where $\mathbf{x} \in \mathbb{R}^2$ (see Table 7.1).

							$n = 200$				$n = 1000$			
							$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{RT}}(\mathbf{u}\|\mathbf{x})$		$\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}\|\mathbf{x})$	
	x_1	x_2	x_3	$f_X(\mathbf{x})$	$\theta(\mathbf{x})$	$\tau(\mathbf{x})$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$	$\widehat{\text{MIAE}}$	$\hat{\sigma}_{\widehat{\text{MIAE}}}$
Scenario 1	-1.5	-1.0	-1.0	0.016	-0.73	-0.53	0.78	0.031	**0.49**	0.015	0.67	0.132	**0.39**	0.069
	-1.0	0.5	-1.5	0.016	-0.48	-0.32	0.82	0.034	**0.45**	0.017	0.52	0.023	**0.30**	0.012
	-0.25	-0.5	-0.5	0.016	-0.32	-0.21	0.72	0.041	**0.39**	0.019	0.49	0.040	**0.28**	0.017
	-1.0	-0.5	0.5	0.016	-0.26	-0.17	0.72	0.030	**0.39**	0.016	0.51	0.035	**0.30**	0.018
	0.0	0.0	0.0	0.016	0.00	0.00	0.68	0.033	**0.37**	0.013	0.48	0.029	**0.26**	0.014
	-0.5	0.5	1.0	0.016	0.26	0.17	0.69	0.032	**0.39**	0.014	0.49	0.028	**0.27**	0.014
	0.25	0.5	0.5	0.016	0.32	0.21	0.69	0.027	**0.39**	0.014	0.47	0.025	**0.27**	0.013
	1.5	1.0	-0.5	0.016	0.48	0.32	0.84	0.045	**0.45**	0.022	0.54	0.027	**0.32**	0.013
	1.0	1.0	1.5	0.016	0.73	0.53	0.85	0.044	**0.53**	0.023	0.54	0.021	**0.34**	0.012
Scenario 2	-1.5	-1.0	-1.0	0.016	-0.73	-0.53	0.86	0.046	**0.58**	0.021	0.53	0.023	**0.37**	0.012
	-1.0	0.5	-1.5	0.016	-0.48	-0.32	0.85	0.040	**0.50**	0.018	0.52	0.022	**0.32**	0.011
	-0.25	-0.5	-0.5	0.016	-0.32	-0.21	0.71	0.030	**0.44**	0.016	0.47	0.025	**0.30**	0.012
	-1.0	-0.5	0.5	0.016	-0.26	-0.17	0.77	0.033	**0.42**	0.017	0.47	0.028	**0.29**	0.014
	0.0	0.0	0.0	0.016	0.00	0.00	0.70	0.028	**0.41**	0.015	0.48	0.024	**0.29**	0.012
	-0.5	0.5	1.0	0.016	0.26	0.17	0.71	0.030	**0.41**	0.015	0.50	0.035	**0.30**	0.016
	0.25	0.5	0.5	0.016	0.32	0.21	0.72	0.033	**0.43**	0.014	0.49	0.027	**0.29**	0.012
	1.5	1.0	-0.5	0.016	0.48	0.32	0.85	0.035	**0.49**	0.022	0.51	0.027	**0.33**	0.013
	1.0	1.0	1.5	0.016	0.73	0.53	0.88	0.044	**0.57**	0.018	0.56	0.033	**0.38**	0.015
Scenario 3	-1.5	-1.0	-1.0	0.016	2.21	0.53	0.79	0.041	**0.56**	0.014	0.53	0.023	**0.40**	0.010
	-1.0	0.5	-1.5	0.016	0.95	0.32	0.80	0.040	**0.44**	0.020	0.51	0.020	**0.30**	0.010
	-0.25	-0.5	-0.5	0.016	0.52	0.21	0.72	0.053	**0.40**	0.020	0.46	0.022	**0.27**	0.012
	-1.0	-0.5	0.5	0.016	0.40	0.17	0.73	0.038	**0.40**	0.018	0.44	0.019	**0.25**	0.009
	-0.5	0.5	1.0	0.016	0.40	0.17	0.67	0.026	**0.39**	0.014	0.45	0.019	**0.27**	0.010
	0.25	0.5	0.5	0.016	0.52	0.21	0.74	0.056	**0.42**	0.024	0.46	0.029	**0.27**	0.013
	1.5	1.0	-0.5	0.016	0.95	0.32	0.84	0.042	**0.46**	0.020	0.51	0.021	**0.32**	0.012
	1.0	1.0	1.5	0.016	2.21	0.53	0.81	0.055	**0.57**	0.023	0.54	0.032	**0.41**	0.018

Table 7.7.: Result of scenarios 1,2 and 3, where $\mathbf{x} \in \mathbb{R}^3$ (see Table 7.1).

The following figures are providing selected contour plots of the average (over R=100) estimations for the scenarios $\mathbf{x} \in \mathbb{R}^2$ (Figure 7.7) and $\mathbf{x} \in \mathbb{R}^3$ (Figure 7.8).



(a) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$ (b) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$ (c) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$

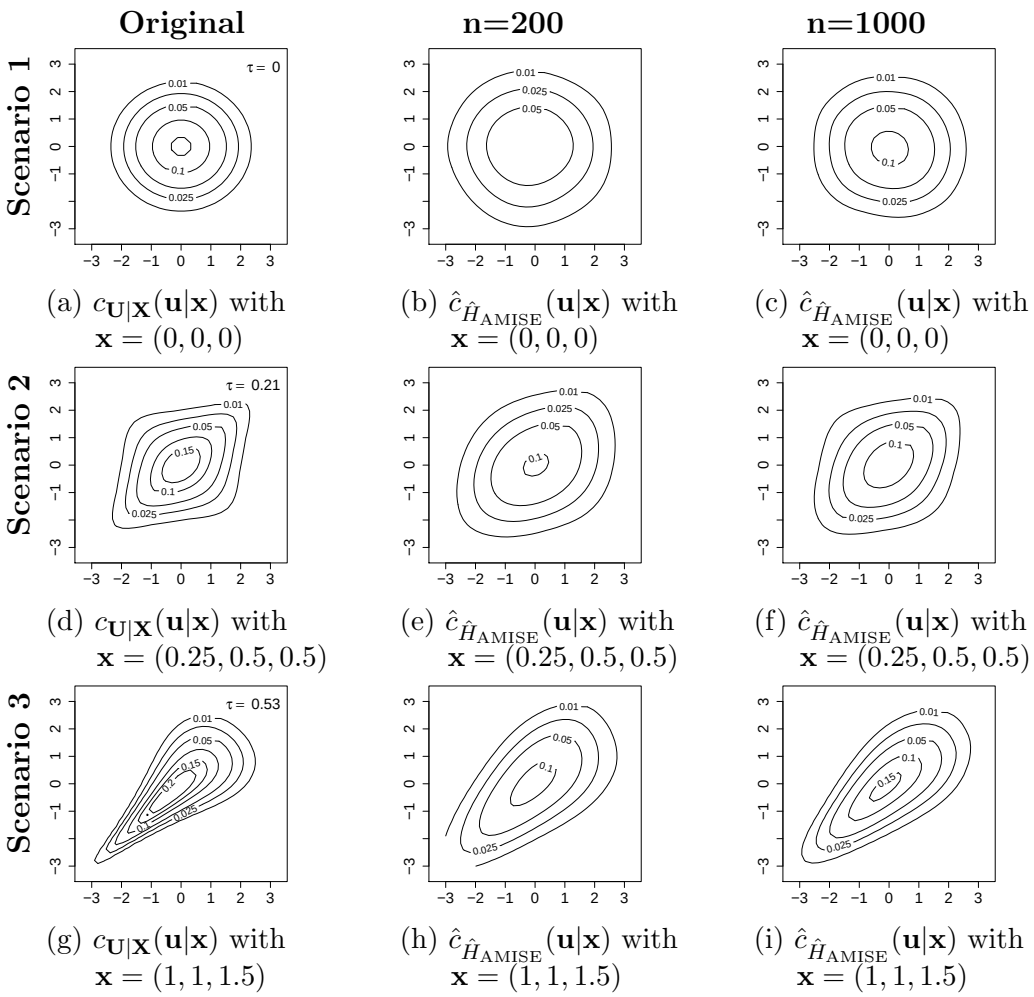
(d) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$ (e) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$ (f) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$

(g) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$ (h) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$ (i) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, -0.67)$

(j) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, 1.64)$ (k) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, 1.64)$ (l) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, 1.64)$

(m) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, 1.64)$ (n) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, 1.64)$ (o) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1.28, 1.64)$

Figure 7.7.: Comparison between contour plots (on Z-space) of original and average conditional copula density estimations for $\mathbf{x} \in \mathbb{R}^2$ scenarios (see Table 7.1).



(a) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (0, 0, 0)$ (b) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (0, 0, 0)$ (c) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (0, 0, 0)$

(d) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (0.25, 0.5, 0.5)$ (e) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (0.25, 0.5, 0.5)$ (f) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (0.25, 0.5, 0.5)$

(g) $c_{\mathbf{U}|\mathbf{X}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1, 1, 1.5)$ (h) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1, 1, 1.5)$ (i) $\hat{c}_{\hat{H}_{\text{AMISE}}}(\mathbf{u}|\mathbf{x})$ with $\mathbf{x} = (1, 1, 1.5)$

Figure 7.8.: Comparison between contour plots (on Z-space) of original and average conditional copula density estimations for $\mathbf{x} \in \mathbb{R}^2$ scenarios (see Table 7.1).

To summarize the results of the simulation study, we have observed, that the main factors influencing the performance of the estimator are: The overall shape of the copula (represented in our study by $\tau(\mathbf{x})$) and the probability of $\mathbf{x}$ occurring (represented in our study by $f_{\mathbf{X}}(\mathbf{x})$). We have seen that for $p > 1$ the bandwidth matrix H_{AMISE} is to prefer, whereas this is not so clear when $p = 1$. Unsurprisingly the estimator has more trouble when p is increasing.

8. Conclusion

This thesis is concerned with kernel estimation of conditional copula densities. We presented a novel approach that is based on transformation kernels combined with local linear regression. We provided the asymptotic properties of this estimator and suggested a bandwidth selection procedure.

We analyzed the performance in a simulation study containing a variety of scenarios, using univariate as well as multivariate covariates. In all scenarios, the method proved to work well. However, in the case of high-dimensional covariates the estimator performed worse, due to the curse of dimensionality.

Possible applications of this estimator can be found in the fields of exploratory analysis of conditional data sets. A further application is the field of vine copulas (see e.g., Aas et al., 2009), where one could use the conditional copula density estimator to estimate non-simplified pair-copula constructions.

A. Additional lemmas and their proofs

The following lemmas are used in the proof of Theorem 5.1. As mentioned in the beginning of Section 5.2, these lemmas and their proofs are based on the proof of Theorem 2.7 in Li and Racine (2006). For reasons of clarity we first state all lemmas and then proof them.

Lemma A.1. *Using the notations of Section 5.2 it holds that*

$$e_1^\top \mathcal{D}\mathcal{M}_1^{-1}\left[\mathcal{M}_2+\mathcal{M}_3\right]+e_1^\top\mathcal{D}(s.o.)=f_{\boldsymbol{X}}(\boldsymbol{x})^{-1}e_1^\top\mathcal{D}\left[\mathcal{M}_2+\mathcal{M}_3\right]+o_p(1).$$

Lemma A.2. *Using the notations of Section 5.2 it holds that*

$$\mathcal{M}_2=\begin{pmatrix} f_{\boldsymbol{X}}(\boldsymbol{x})\mu_K\sum_{j=1}^p\frac{b^2}{2}\frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial x_j^2}+O_p(b^3)\\ O_p(b)\,\boldsymbol{1}_p\end{pmatrix},$$

and

$$e_1^\top\mathcal{D}\mathcal{M}_2=f_{\boldsymbol{X}}(\boldsymbol{x})\sqrt{n|H|^{\frac{1}{2}}b^p}\mu_K\sum_{j=1}^p\frac{b^2}{2}\frac{\partial^2 f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})}{\partial x_j^2}+o_p(1),$$

where $\boldsymbol{1}_p$ is the vector in $\mathbb{R}^p$ where every entry is one.

Lemma A.3. *Using the notations of Section 5.2 it holds that*

$$e_1^\top\mathcal{D}\mathcal{M}_3=\sqrt{n|H|^{\frac{1}{2}}b^p}\cdot e_1^\top\mathcal{M}_3\to N\left(0,R_{\mathcal{K}}\nu_K^p f_{\boldsymbol{X}}(\boldsymbol{x})f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})\right).$$

Lemma A.4. *Using the notations of Section 5.2 it holds that*

$$g(\boldsymbol{x})=f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x})+\frac{1}{2}\mu_{\mathcal{K}}\mathrm{tr}(H\mathcal{H}_{\boldsymbol{z}})+O\Big(\mathrm{tr}(\boldsymbol{1}_{d^3\times d^3}(H^{\frac{1}{2}}\otimes H^{\frac{1}{2}}\otimes H^{\frac{1}{2}}))\Big).$$

To allow for fluent reading of the proofs of Lemma A.1 to Lemma A.4 (see Page 97) we will need to do some preparation work, i.e. we derive basic notations and some minor lemmas (Lemma A.5 to Lemma A.7). For notational convenience define

the function $K^p(\cdot)$, which is equivalent to the product kernel with bandwidth equal to one:

$$K^p(\mathbf{x}) := \left\{ \prod_{j=1}^{p} K(x_j) \right\} \text{ for } \mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p. \tag{A.1}$$

It is now possible to state the following minor lemma.

Lemma A.5. *Using the notations of Section 5.2 it holds that:*

$$E\left[E\left[K^p\left(\frac{\boldsymbol{X}_1 - \boldsymbol{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\boldsymbol{Z}_1 - \boldsymbol{z})\right) \middle| \boldsymbol{X}_1 \right]^2 \right] = O(|H| b^p).$$

Proof. (Lemma A.5)

$$\begin{aligned} &E\left[E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1 \right]^2 \right] \\ &\quad = E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right)^2 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1 \right]^2 \right] \\ &\quad = \int_{\mathbb{R}^p} K^p\left(\frac{\tilde{\mathbf{x}} - \mathbf{x}}{b}\right)^2 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1 = \tilde{\mathbf{x}} \right]^2 f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}. \end{aligned} \tag{A.2}$$

Note that

$$E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1 = \tilde{\mathbf{x}} \right] = |H|^{\frac{1}{2}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}}) + O(|H|^{\frac{1}{2}}),$$

because

$$\begin{aligned} E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1 = \tilde{\mathbf{x}} \right] &= \int_{\mathbb{R}^d} \mathcal{K}\left(H^{-\frac{1}{2}}(\tilde{\mathbf{z}} - \mathbf{z})\right) f_{\mathbf{Z}|\mathbf{X}}(\tilde{\mathbf{z}}|\tilde{\mathbf{x}}) d\tilde{\mathbf{z}} \\ &= \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) f_{\mathbf{Z}|\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z}|\tilde{\mathbf{x}}) |H|^{\frac{1}{2}} d\bar{\mathbf{z}} \\ &= |H|^{\frac{1}{2}} \left\{ \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}}) d\bar{\mathbf{z}} + O(\mathrm{tr}(\mathbf{1}_{d\times d} H^{\frac{1}{2}})) \right\} \\ &= |H|^{\frac{1}{2}} \left\{ f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}}) \underbrace{\int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}}) d\bar{\mathbf{z}}}_{=1} + O(\mathrm{tr}(\mathbf{1}_{d\times d} H^{\frac{1}{2}})) \right\} \\ &= |H|^{\frac{1}{2}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}}) + O(|H|^{\frac{1}{2}} \mathrm{tr}(\mathbf{1}_{d\times d} H^{\frac{1}{2}})) \\ &= |H|^{\frac{1}{2}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}}) + O(|H|^{\frac{1}{2}}). \end{aligned}$$

Where we used in the second step the substitution $\tilde{\mathbf{z}} = (H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z})$ and in the third the linear Taylor expansions (see Theorem 2.1). Note also that we used implicitly that $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ has bounded continuous first order derivatives with respect to $\mathbf{z}$ for all $\mathbf{x}$ (see Remark 3.1) and it holds that $\int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})\, \bar{z}_j d\bar{\mathbf{z}} < \infty$ $j \in \{1, \ldots, d\}$.

Using this in (A.2) we get

$$
\begin{aligned}
&E\left[E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right)\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right)\middle|\mathbf{X}_1\right]^2\right] \\
&\quad = \int_{\mathbb{R}^p} K^p\left(\frac{\tilde{\mathbf{x}} - \mathbf{x}}{b}\right)^2 \left[|H|^{\frac{1}{2}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}}) + O(|H|^{\frac{1}{2}})\right]^2 f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \\
&\quad = \int_{\mathbb{R}^p} K^p(\bar{\mathbf{x}})^2 \left[|H|^{\frac{1}{2}} f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}} + \mathbf{x}) + O(|H|^{\frac{1}{2}})\right]^2 f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x}) b^p d\bar{\mathbf{x}} \\
&\quad = \int_{\mathbb{R}^p} K^p(\bar{\mathbf{x}})^2 \left[|H|^{\frac{1}{2}} \left\{f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + O\left(b\sum_{j=1}^p \bar{x}_j\right)\right\} + O(|H|^{\frac{1}{2}})\right]^2 \\
&\qquad\qquad \cdot \left\{f_{\mathbf{X}}(\mathbf{x}) + O\left(b\sum_{j=1}^p \bar{x}_j\right)\right\} b^p d\bar{\mathbf{x}} \\
&\quad = \left[O(|H|^{\frac{1}{2}})\right]^2 O(b^p) O(1) = O(|H| b^p).
\end{aligned}
$$

We used the substitution $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$ and linear expansions of $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}} + \mathbf{x})$ and $f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x})$ (see Corollary 2.1). Also we used implicitly that we know that $\int_{\mathbb{R}^p} K^p(\bar{\mathbf{x}})^2 \bar{x}_j \bar{x}_l \bar{x}_k d\bar{\mathbf{x}} < \infty$ for $j, l, k \in \{1, \ldots, p\}$ (by applying Fubini on our kernel conditions) and that we assume that $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})$ and $f_{\mathbf{X}}(\mathbf{x})$ have bounded continuous first order derivatives with respect to $\mathbf{x}$ (see Remark 3.1). □

An important minor result for establishing Lemma A.1 and Lemma A.3 will be the following Lemma:

Lemma A.6. *Using the notations of Section 5.2 it holds that:*

$$
\begin{aligned}
\mathrm{E}\left[\mathcal{M}_3\right] &= \boldsymbol{0}_{p+1}, \\
\mathrm{Var}\left[e_1^\top \mathcal{M}_3\right] &= \frac{1}{n|H|^{\frac{1}{2}} b^p} \nu_K^p R_{\mathcal{K}} f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Z}|\boldsymbol{X}}(\boldsymbol{z}|\boldsymbol{x}) + o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p}\right),
\end{aligned}
$$

and for $l \in \{1, \ldots, p\}$, it holds

$$
\mathrm{Var}\left[e_{l+1}^\top \mathcal{M}_3\right] = O\left(\frac{1}{n|H|^{\frac{1}{2}} b^{p+2}}\right),
$$

where e_j for $j \in \{1, \ldots, p+1\}$ is representing the vector in $\mathbb{R}^{p+1}$ s.t. every entry is zero except of the j'th one which is equal to 1.

Proof. (Lemma A.6)
We start by showing that

$$E[\mathcal{M}_3] = \mathbf{0}_{p+1}.$$

Per definition it holds that

$$\begin{aligned} E[\mathcal{M}_3] &= E\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \{\epsilon_i\}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[\begin{pmatrix} W_b(\mathbf{X}_i - \mathbf{x})\epsilon_i \\ b^{-2}W_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})\epsilon_i \end{pmatrix}\right]. \end{aligned}$$

Using that

$$\epsilon_i = K_H(\mathbf{Z}_i - \mathbf{z}) - E\left[K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right],$$

leads to

$$E[\mathcal{M}_3] = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} E\Big[W_b(\mathbf{X}_i - \mathbf{x})\Big\{K_H(\mathbf{Z}_i - \mathbf{z}) - E\left[K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right]\Big\}\Big] \\ E\Big[b^{-2}W_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})\Big\{K_H(\mathbf{Z}_i - \mathbf{z}) - E\left[K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right]\Big\}\Big] \end{pmatrix}.$$

Now because $W_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})$ and $W_b(\mathbf{X}_i - \mathbf{x})$ are $\sigma(\mathbf{X}_i)$ measurable it holds that

$$\begin{aligned} W_b(\mathbf{X}_i - \mathbf{x})E\left[K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right] &= E\left[W_b(\mathbf{X}_i - \mathbf{x})K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right] \\ W_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})E\left[K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right] &= E\left[W_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right]. \end{aligned}$$

Using the tower lemma $E[E[\cdot|\mathbf{X}_i]] = E[\cdot]$ and the linearity of the expected value leads then immediately to the result $E[\mathcal{M}_3] = \mathbf{0}_{p+1}$.

The next step is to analyze the variance. It holds that

$$\mathrm{Var}(e_1^\top \mathcal{M}_3) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})\epsilon_i\right).$$

One can easily see this because the $(\mathbf{X}_i, \mathbf{Z}_i)$ for $i \in \{1, \ldots, n\}$ are *i.i.d.*, thus also $W_b(\mathbf{X}_i - \mathbf{x})\epsilon_i$ for $i \in \{1, \ldots, n\}$ are *i.i.d.*. This leads to

$$\mathrm{Var}(e_1^\top \mathcal{M}_3) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})\epsilon_i\right) = \frac{1}{n}\mathrm{Var}\left(W_b(\mathbf{X}_1 - \mathbf{x})\epsilon_1\right).$$

From the calculation of $\mathrm{E}[\mathcal{M}_3]$ we know that $\mathrm{E}[W_b(\mathbf{X}_1 - \mathbf{x})\epsilon_1] = 0$ and get

$$\mathrm{Var}(e_1^\top \mathcal{M}_3) = \frac{1}{n}E\left[\left(W_b(\mathbf{X}_1 - \mathbf{x})\epsilon_1\right)^2\right].$$

Applying the definition of ϵ_i leads to

$$\begin{aligned}\mathrm{Var}(e_1^\top \mathcal{M}_3) &= \frac{1}{n} E\Big[\left(W_b(\mathbf{X}_1 - \mathbf{x})\{K_H(\mathbf{Z}_1 - \mathbf{z}) - E\left[K_H(\mathbf{Z}_1 - \mathbf{z})|\mathbf{X}_i\right]\}\right)^2 \Big] \\ &= \frac{1}{n} E\Bigg[\Big(W_b(\mathbf{X}_1 - \mathbf{x})K_H(\mathbf{Z}_1 - \mathbf{z}) - E\Big[W_b(\mathbf{X}_1 - \mathbf{x})K_H(\mathbf{Z}_1 - \mathbf{z})|\mathbf{X}_i\Big]\Big)^2 \Bigg].\end{aligned} \tag{A.3}$$

Remember that

$$W_b(\mathbf{X}_i - \mathbf{x})K_H(\mathbf{Z}_i - \mathbf{z}) = \frac{1}{\sqrt{|H|}b^p} \left\{ \prod_{j=1}^{p} K\left(\frac{X_{i,j} - x_j}{b}\right) \right\} \left\{ \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_i - \mathbf{z})\right) \right\}.$$

By rewriting this with the function $K^p(\cdot)$ defined in (A.1) we get

$$W_b(\mathbf{X}_i - \mathbf{x})K_H(\mathbf{Z}_i - \mathbf{z}) = \frac{1}{\sqrt{|H|}b^p} K^p\left(\frac{\mathbf{X}_i - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_i - \mathbf{z})\right).$$

By using this notion in our previous result (A.3) we get

$$\begin{aligned}\mathrm{Var}(e_1^\top \mathcal{M}_3) = \frac{1}{n|H|b^{2p}} E\Bigg[\Bigg(& K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \\ & - E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1\right]\Bigg)^2 \Bigg].\end{aligned}$$

Note that we can see the term above as the expected value of the conditional variance,

$$\mathrm{Var}(e_1^\top \mathcal{M}_3) = \frac{1}{n|H|b^{2p}} \mathrm{E}\left[\mathrm{Var}\left(K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1\right)\right],$$

and by the law of total variance we get

$$\begin{aligned}\mathrm{Var}(e_1^\top \mathcal{M}_3) &= \frac{1}{n|H|b^{2p}} \Bigg\{ \mathrm{Var}\left(K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right)\right) \\ &\qquad - \mathrm{Var}\left(\mathrm{E}\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1\right]\right)\Bigg\} \\ &= \frac{1}{n|H|b^{2p}} \Bigg\{ \mathrm{E}\left(K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right)^2 \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right)^2\right) \\ &\qquad - \mathrm{E}\left(K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right)\right)^2 \\ &\qquad - \mathrm{E}\left(\mathrm{E}\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1\right]^2\right) \\ &\qquad + \mathrm{E}\left(\mathrm{E}\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right) \middle| \mathbf{X}_1\right]\right)^2 \Bigg\}.\end{aligned}$$

By the tower lemma $E[E[\cdot|\mathbf{X}_i]] = E[\cdot]$ we get

$$\mathrm{Var}(e_1^\top \mathcal{M}_3) = V_1 - V_2, \tag{A.4}$$

where

$$V_1 = \frac{1}{n|H|b^{2p}} E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right)^2 \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right)^2\right],$$
$$V_2 = \frac{1}{n|H|b^{2p}} E\left[E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right)\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right)\middle|\mathbf{X}_1\right]^2\right].$$

We will now show that

$$\begin{aligned} V_1 &= \frac{1}{nb^p|H|^{\frac{1}{2}}} R_{\mathcal{K}} \nu_K^p f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + o\left(\frac{1}{nb^p|H|^{\frac{1}{2}}}\right), \\ V_2 &= o\left(\frac{1}{nb^p|H|^{\frac{1}{2}}}\right). \end{aligned} \tag{A.5}$$

Using the definition of expected value yields

$$\begin{aligned} V_1 &= \frac{1}{n|H|b^{2p}} E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b}\right)^2 \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z})\right)^2\right] \\ &= \frac{1}{n|H|b^{2p}} \int_{\mathbb{R}^{p+d}} K^p\left(\frac{\tilde{\mathbf{x}} - \mathbf{x}}{b}\right)^2 \mathcal{K}\left(H^{-\frac{1}{2}}(\tilde{\mathbf{z}} - \mathbf{z})\right)^2 \cdot f_{\mathbf{Z},\mathbf{X}}(\tilde{\mathbf{z}}, \tilde{\mathbf{x}}) d(\tilde{\mathbf{z}}, \tilde{\mathbf{x}}). \end{aligned}$$

Substitution with $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$ and $\tilde{\mathbf{z}} = (H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z})$ leads to

$$\begin{aligned} V_1 &= \frac{1}{n|H|b^{2p}} \int_{\mathbb{R}^{p+d}} K^p(\bar{\mathbf{x}})^2 \mathcal{K}(\bar{\mathbf{z}})^2 \cdot f_{\mathbf{Z},\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z}, b\bar{\mathbf{x}} + \mathbf{x}) |H|^{\frac{1}{2}} b^p d(\bar{\mathbf{z}}, \bar{\mathbf{x}}) \\ &= \frac{1}{n|H|^{\frac{1}{2}} b^p} \int_{\mathbb{R}^{p+d}} K^p(\bar{\mathbf{x}})^2 \mathcal{K}(\bar{\mathbf{z}})^2 \cdot f_{\mathbf{Z},\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z}, b\bar{\mathbf{x}} + \mathbf{x}) d(\bar{\mathbf{z}}, \bar{\mathbf{x}}). \end{aligned}$$

The next step will be a Taylor expansion

$$f_{\mathbf{Z},\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}} + \mathbf{z}, b\bar{\mathbf{x}} + \mathbf{x}) = f_{\mathbf{Z},\mathbf{X}}(\mathbf{z}, \mathbf{x}) + \sum_{\substack{(k,r) \\ \in\{1,\dots,d\}^2}} O\left([H^{\frac{1}{2}}]_{k,r}\bar{z}_r\right) + \sum_{r=1}^{p} O\left(b\bar{x}_r\right).$$

Using this we get

$$V_1 = \frac{1}{n|H|^{\frac{1}{2}} b^p}\left\{f_{\mathbf{Z},\mathbf{X}}(\mathbf{z}, \mathbf{x}) \int_{\mathbb{R}^{p+d}} K^p(\bar{\mathbf{x}})^2 \mathcal{K}(\bar{\mathbf{z}})^2 d(\bar{\mathbf{z}}, \bar{\mathbf{x}}) + O(\mathrm{tr}(\mathbf{1}_{d\times d} H^{\frac{1}{2}})) + O(b)\right\}.$$

Recall that

$$\int_{\mathbb{R}^{p+d}} K^p(\bar{\mathbf{x}})^2 \mathcal{K}(\bar{\mathbf{z}})^2 d(\bar{\mathbf{z}}, \bar{\mathbf{x}}) = \int_{\mathbb{R}^{p+d}} \left(\left\{ \prod_{j=1}^{p} K(\bar{x}_j) \right\} \mathcal{K}(\bar{\mathbf{z}}) \right)^2 d(\bar{\mathbf{z}}, \bar{\mathbf{x}})$$
$$= \left\{ \prod_{j=1}^{p} \int_{\mathbb{R}} K(\bar{x}_j)^2 d\bar{x}_j \right\} \left\{ \int_{\mathbb{R}^d} \mathcal{K}(\bar{\mathbf{z}})^2 d\bar{\mathbf{z}} \right\} = \nu_K^p R_{\mathcal{K}},$$

and also

$$\frac{1}{n|H|^{\frac{1}{2}} b^p} O(\mathrm{tr}(\mathbf{1}_{d\times d} H^{\frac{1}{2}})) = o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p} \right) \text{ because } O(\mathrm{tr}(\mathbf{1}_{d\times d} H^{\frac{1}{2}})) = o(1),$$
$$\frac{1}{n|H|^{\frac{1}{2}} b^p} O(b) = o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p} \right) \text{ because } O(b) = o(1).$$

So we conclude

$$\Rightarrow V_1 = \frac{1}{n|H|^{\frac{1}{2}} b^p} \nu_K^p R_{\mathcal{K}} f_{\mathbf{Z},\mathbf{X}}(\mathbf{z}, \mathbf{x}) + o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p} \right)$$
$$= \frac{1}{n|H|^{\frac{1}{2}} b^p} \nu_K^p R_{\mathcal{K}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p} \right).$$

The next step is to analyze V_2:

$$V_2 = \frac{1}{n|H|b^{2p}} E\left[E\left[K^p\left(\frac{\mathbf{X}_1 - \mathbf{x}}{b} \right) \mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1 - \mathbf{z}) \right) \middle| \mathbf{X}_1 \right]^2 \right].$$

By using Lemma A.5 we get the needed result for V_2:

$$V_2 = \frac{1}{n|H|b^{2p}} O(|H|b^p) = o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p} \right).$$

Therefore it holds by (A.4) and (A.5) that

$$\mathrm{Var}(e_1^\top \mathcal{M}_3) = \frac{1}{n|H|^{\frac{1}{2}} b^p} \nu_K^p R_{\mathcal{K}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p} \right).$$

Finally we analyze the last expression in Lemma A.6. For $l \in \{1, \ldots, p\}$, it holds

$$\mathrm{Var}\left[e_{l+1}^\top \mathcal{M}_3 \right] = O\left(\frac{1}{n|H|^{\frac{1}{2}} b^{p+2}} \right).$$

By definition of $\mathcal{M}_3$, we can easily see that for $l \in \{1, \ldots, p\}$

$$\mathrm{Var}\left[e_{l+1}^{\top}\mathcal{M}_3\right] = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} b^{-2}W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\epsilon_i\right). \tag{A.6}$$

So we have to prove that for $l \in \{1, \ldots, p\}$

$$V_3 := \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} b^{-2}W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\epsilon_i\right) = O\left(\frac{1}{n|H|^{\frac{1}{2}}b^{p+2}}\right).$$

Since $W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\epsilon_i$ for $i \in \{1, \ldots, n\}$ are *i.i.d.* we get

$$V_3 = \frac{1}{nb^4}\mathrm{Var}\left(W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)\epsilon_1\right).$$

By proving that $\mathrm{E}[\mathcal{M}_3] = \mathbf{0}_{p+1}$ we have seen that $\mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)\epsilon_1\right] = 0$. Therefore

$$\begin{aligned}
V_3 &= \frac{1}{nb^4}\mathrm{E}\left[\left(W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)\epsilon_1\right)^2\right] \\
&= \frac{1}{nb^4}E\left[\{W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z}) \right. \\
&\quad \left. -E\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})|\mathbf{X}_1\right]\}^2\right].
\end{aligned}$$

We can see the term above as the expected value of the conditional variance,

$$V_3 = \frac{1}{nb^4}\mathrm{E}\left[\mathrm{Var}\left(W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})\middle|\mathbf{X}_1\right)\right],$$

and by law of total variance we get

$$\begin{aligned}
V_3 &= \frac{1}{nb^4}\left\{\mathrm{Var}\left(W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})\right)\right. \\
&\quad \left. - \mathrm{Var}\left(\mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})|\mathbf{X}_1\right]\right)\right\} \\
&= \frac{1}{nb^4}\left\{\mathrm{E}\left[\left(W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})\right)^2\right]\right. \\
&\quad - \mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})\right]^2 \\
&\quad - \mathrm{E}\left(\mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})|\mathbf{X}_1\right]^2\right) \\
&\quad \left. + \mathrm{E}\left(\mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)K_H(\mathbf{Z}_1 - \mathbf{z})|\mathbf{X}_1\right]\right)^2\right\}.
\end{aligned}$$

By the tower lemma $E[E[\cdot|\mathbf{X}_i]] = E[\cdot]$ we get:

$$V_3 = \tilde{V}_1 - \tilde{V}_2 \tag{A.7}$$

where

$$\tilde{V}_1 = \frac{1}{nb^4}\left\{\mathrm{E}\left[(W_b(\mathbf{X}_1-\mathbf{x})(X_{1,l}-x_l)K_H(\mathbf{Z}_1-\mathbf{z}))^2\right]\right\},$$
$$\tilde{V}_2 = \frac{1}{nb^4}\left\{\mathrm{E}\left(\mathrm{E}\left[W_b(\mathbf{X}_1-\mathbf{x})(X_{1,l}-x_l)K_H(\mathbf{Z}_1-\mathbf{z})|\mathbf{X}_1\right]^2\right)\right\}.$$

As defined in (A.1) we use the notion of $K^p(\cdot)$. Note that

$$W_b(\mathbf{x}) = \frac{1}{b^p}K^p\left(\frac{1}{b}\mathbf{x}\right),$$
$$K_H(\mathbf{z}) = \frac{1}{|H|^{\frac{1}{2}}}\mathcal{K}\left(H^{-\frac{1}{2}}\mathbf{z}\right),$$

Using the definition of expected value results in

$$\begin{aligned}\tilde{V}_1 &= \frac{1}{nb^4}\left\{\mathrm{E}\left[(W_b(\mathbf{X}_1-\mathbf{x})(X_{1,l}-x_l)K_H(\mathbf{Z}_1-\mathbf{z}))^2\right]\right\}\\ &= \frac{1}{nb^4}\int_{\mathbb{R}^{p+d}} W_b(\tilde{\mathbf{x}}-\mathbf{x})^2 K_H(\tilde{\mathbf{z}}-\mathbf{z})^2(\tilde{x}_l-x_l)^2\cdot f_{\mathbf{Z},\mathbf{X}}(\tilde{\mathbf{z}},\tilde{\mathbf{x}})d(\tilde{\mathbf{z}},\tilde{\mathbf{x}})\\ &= \frac{1}{nb^4}\frac{1}{|H|b^{2p}}\int_{\mathbb{R}^{p+d}} K^p\left(\frac{\tilde{\mathbf{x}}-\mathbf{x}}{b}\right)^2\mathcal{K}\left(H^{-\frac{1}{2}}(\tilde{\mathbf{z}}-\mathbf{z})\right)^2(\tilde{x}_l-x_l)^2 f_{\mathbf{Z},\mathbf{X}}(\tilde{\mathbf{z}},\tilde{\mathbf{x}})d(\tilde{\mathbf{z}},\tilde{\mathbf{x}}).\end{aligned}$$

Substitution with $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}}+\mathbf{x})$ and $\tilde{\mathbf{z}} = (H^{\frac{1}{2}}\bar{\mathbf{z}}+\mathbf{z})$ leads to

$$\begin{aligned}\tilde{V}_1 &= \frac{1}{nb^4}\frac{1}{|H|b^{2p}}\int_{\mathbb{R}^{p+d}} K^p(\bar{\mathbf{x}})^2\mathcal{K}(\bar{\mathbf{z}})^2(b\bar{x}_l)^2\cdot f_{\mathbf{Z},\mathbf{X}}(h\bar{\mathbf{z}}+\mathbf{z},b\bar{\mathbf{x}}+\mathbf{x})|H|^{\frac{1}{2}}b^p d(\bar{\mathbf{z}},\bar{\mathbf{x}})\\ &= \frac{1}{nb^2}\frac{1}{|H|^{\frac{1}{2}}b^p}\int_{\mathbb{R}^{p+d}} K^p(\bar{\mathbf{x}})^2\mathcal{K}(\bar{\mathbf{z}})^2\bar{x}_l^2\cdot f_{\mathbf{Z},\mathbf{X}}(h\bar{\mathbf{z}}+\mathbf{z},b\bar{\mathbf{x}}+\mathbf{x})d(\bar{\mathbf{z}},\bar{\mathbf{x}}).\end{aligned}$$

Now we use a Taylor expansion

$$f_{\mathbf{Z},\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}}+\mathbf{z},b\bar{\mathbf{x}}+\mathbf{x}) = f_{\mathbf{Z},\mathbf{X}}(\mathbf{z},\mathbf{x}) + \sum_{\substack{(k,r)\\ \in\{1,\dots,d\}^2}} O\left([H^{\frac{1}{2}}]_{k,r}\bar{z}_r\right) + \sum_{r=1}^{p} O\left(b\bar{x}_r\right).$$

Using this we get

$$\tilde{V}_1 = O\left(\frac{1}{n|H|^{\frac{1}{2}}b^{p+2}}\right).$$

Now we analyze $\tilde{V}_2$:

$$\begin{aligned}\tilde{V}_2 &= \frac{1}{nb^4}\left\{\mathrm{E}\left(\mathrm{E}\left[W_b(\mathbf{X}_1-\mathbf{x})(X_{1,l}-x_l)K_H(\mathbf{Z}_1-\mathbf{z})|\mathbf{X}_1\right]^2\right)\right\}\\ &= \frac{1}{nb^4}\mathrm{E}\left(W_b(\mathbf{X}_1-\mathbf{x})^2(X_{1,l}-x_l)^2\mathrm{E}\left[K_H(\mathbf{Z}_1-\mathbf{z})|\mathbf{X}_1\right]^2\right)\\ &= \frac{1}{n|H|b^{2p+4}}\int_{R^p} K^p\left(\frac{\tilde{\mathbf{x}}-\mathbf{x}}{b}\right)^2(\tilde{x}_l-x_l)^2 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1=\tilde{\mathbf{x}}\right]^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}.\end{aligned}$$

We already showed in the proof of Lemma A.5 that

$$E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)\middle|\mathbf{X}_1=\tilde{\mathbf{x}}\right]=|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}})+O(|H|^{\frac{1}{2}}).$$

Therefore

$$\tilde{V}_2=\frac{1}{n|H|b^{2p+4}}\int_{R^p}K^p\left(\frac{\tilde{\mathbf{x}}-\mathbf{x}}{b}\right)^2(\tilde{x}_l-x_l)^2\left[|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}})+O(|H|^{\frac{1}{2}})\right]^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}.$$

The next step is the usual substitution with $\tilde{\mathbf{x}}=(b\bar{\mathbf{x}}+\mathbf{x})$:

$$\tilde{V}_2=\frac{1}{n|H|b^{2p+4}}\int_{R^p}K^p\left(\bar{\mathbf{x}}\right)^2b^2\bar{x}_l^2\left[|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}}+\mathbf{x})+O(|H|^{\frac{1}{2}})\right]^2 f_{\mathbf{X}}(b\bar{\mathbf{x}}+\mathbf{x})b^pd\bar{\mathbf{x}}.$$

Then use the linear Taylor expansions

$$f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}}+\mathbf{x})=f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right),$$
$$f_{\mathbf{X}}(b\bar{\mathbf{x}}+\mathbf{x})=f_{\mathbf{X}}(\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right),$$

to get

$$\tilde{V}_2=\frac{1}{n|H|b^{p+2}}\int_{R^p}\bar{x}_l^2K^p\left(\bar{\mathbf{x}}\right)^2\left[|H|^{\frac{1}{2}}\left(f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right)\right)+O(|H|^{\frac{1}{2}})\right]^2$$
$$\left(f_{\mathbf{X}}(\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right)\right)d\bar{\mathbf{x}}.$$

Finally factoring $O(|H|)$ out and noticing that the remaining integral is $O(1)$ leads to

$$\tilde{V}_2=O\left(\frac{1}{nb^{p+2}}\right)=O\left(\frac{1}{n|H|^{\frac{1}{2}}b^{p+2}}\right).$$

Combing now this result with (A.7) completes the proof of the last statement of Lemma A.6. □

The last minor Lemma will help us to analyze the asymptotic behavior of $\mathcal{M}_1$.

Lemma A.7.

$$\mathcal{M}_1=\begin{pmatrix} f_{\boldsymbol{X}}(\boldsymbol{x}) & \boldsymbol{0}_p^\top \\ \mu_K\nabla f_{\boldsymbol{X}}(\boldsymbol{x}) & \mu_K f_{\boldsymbol{X}}(\boldsymbol{x})I_p\end{pmatrix}+\begin{pmatrix} O_p(b^2) & O_p(b^2)\boldsymbol{1}_p^\top \\ o_p(1)\boldsymbol{1}_p & o_p(1)\boldsymbol{1}_{p\times p}\end{pmatrix},$$

where I_p is the p-dimensional identity matrix, $\nabla f_{\boldsymbol{X}}(\boldsymbol{x})$ is the gradient of $f_{\boldsymbol{X}}(\boldsymbol{x})$, $\boldsymbol{0}_p$ is the zero vector in $\mathbb{R}^p$, $\boldsymbol{1}_p$ the vector in $\mathbb{R}^p$ and $\boldsymbol{1}_{p\times p}$ the matrix in $\mathbb{R}^{p\times p}$ where every entry is one.

Proof. (Lemma A.7)
To proof this statement we analyze the matrix $\mathcal{M}_1$ blockwise:

$$\mathcal{M}_1 = \begin{pmatrix} \mathcal{M}_1^{11} & \mathcal{M}_1^{12} \\ \mathcal{M}_1^{21} & \mathcal{M}_1^{22} \end{pmatrix},$$

where $\mathcal{M}_1^{11} \in \mathbb{R}$, $\mathcal{M}_1^{12} \in \mathbb{R}^{1\times p}$,$\mathcal{M}_1^{21} \in \mathbb{R}^{p\times 1}$ and $\mathcal{M}_1^{22} \in \mathbb{R}^{p\times p}$. Note that

$$\mathcal{M}_1^{11} = \frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \in \mathbb{R}.$$

This is the well known standard multivariate kernel density estimator $\hat{f}_{\mathbf{X}}(\mathbf{x})$ of $f_{\mathbf{X}}(\mathbf{x})$ for which we know $\hat{f}_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) + O_p(b^2)$.

Now we analyze $\mathcal{M}_1^{12}$:

$$\begin{aligned}\mathcal{M}_1^{12} &= \frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^\top \in \mathbb{R}^{1\times p} \\ &= \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,1} - x_1), \ldots, \tfrac{1}{n}\textstyle\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,p} - x_p)\right).\end{aligned}$$

We will show that for $l \in \{1, \ldots, p\}$,

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] = O(b^2),$$

and

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] = o(1).$$

Then by using that a vanishing variance implies convergence in L^2 to the expectation, and convergence in L^2 implies convergence in probability, we get that

$$\mathcal{M}_1^{12} = O_p(b^2)\mathbf{1}_p^\top.$$

We now prove that

$$E^{12} := \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] = O(b^2),$$

and

$$V^{12} := \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] = o(1).$$

For $l \in \{1, \ldots, p\}$

$$\begin{aligned}
E^{12} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)] \\
&= \mathrm{E}[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)] \\
&= \int_{R^p} W_b(\tilde{\mathbf{x}} - \mathbf{x})(\tilde{x}_l - x_l) f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \\
&= \int_{R^p} W_b(\tilde{\mathbf{x}} - \mathbf{x})(\tilde{x}_l - x_l) f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \\
&= \int_{R^p} \frac{1}{b^p}\left\{\prod_{j=1}^{p} K\left(\frac{\tilde{x}_j - x_j}{b}\right)\right\}(\tilde{x}_l - x_l) f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}.
\end{aligned}$$

By using change of variable $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$ we get:

$$\begin{aligned}
E^{12} &= \int_{R^p} \left\{\prod_{j=1}^{p} K(\bar{x}_j)\right\}(b\bar{x}_l) f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x}) d\bar{\mathbf{x}} \\
&= \int_{R^p} \left\{\prod_{j=1}^{p} K(\bar{x}_j)\right\} b\bar{x}_l \Bigg\{ f_{\mathbf{X}}(\mathbf{x}) + b\bar{\mathbf{x}}^\top \nabla_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \\
&\qquad\qquad + \sum_{\substack{(r_1,r_2)\\ \in\{1,\ldots,d\}^2}} O\left(b^2 \bar{x}_{r_1}\bar{x}_{r_2}\right)\Bigg\} d\bar{\mathbf{x}} \\
&= b f_{\mathbf{X}}(\mathbf{x}) \left\{ \underbrace{\int_R x_l K(\bar{x}_l)\, d\bar{x}_l}_{=0} \prod_{\substack{j=1\\ j\neq l}}^{p} \int_R K(\bar{x}_j)\, d\bar{x}_j \right\} \\
&\quad + b^2 \sum_{m=1}^{p} \int_{R^p} \left\{\prod_{j=1}^{p} K(\bar{x}_j)\right\} \bar{x}_l \left\{\frac{\partial}{\partial x_m} f_{\mathbf{X}}(\mathbf{x}) \bar{x}_m\right\} d\bar{\mathbf{x}} + O(b^3) \\
&= b^2 \sum_{m=1}^{p} \frac{\partial f}{\partial x_m}(\mathbf{x}) \int_{R^p} \left\{\prod_{j=1}^{p} K(\bar{x}_j)\right\} \bar{x}_l \{\bar{x}_m\}\, d\bar{\mathbf{x}} + O(b^3) \\
&= b^2 \frac{\partial f}{\partial x_l}(\mathbf{x}) \left\{\int_R K(\bar{x}_l)\, \bar{x}_l^2 d\bar{x}_l\right\} \underbrace{\left\{\prod_{\substack{j=1\\ j\neq l}}^{p} \int_R K(\bar{x}_j)\, d\bar{x}_j\right\}}_{=1} + O(b^3) \\
&\quad + b^2 \sum_{\substack{m=1\\ m\neq l}}^{p} \frac{\partial f}{\partial x_m}(\mathbf{x}) \left\{\prod_{\substack{j=1\\ j\neq l\\ j\neq m}}^{p} \int_R \{K(\bar{x}_j)\}\, d\bar{x}_j\right\} \underbrace{\left\{\prod_{j\in\{l,m\}} \int_R K(\bar{x}_j)\, \bar{x}_j d\bar{x}_j\right\}}_{=0}
\end{aligned}$$

$$E^{12} = b^2 \frac{\partial f}{\partial x_l}(\mathbf{x})\mu_K + O(b^3) = O(b^2),$$

because $\int_R K(\bar{x})\, d\bar{x} = 1$ and $\int_R xK(x)\, dx = 0$.

The next step is to analyze the variance. For $l \in \{1, \ldots, p\}$,

$$\begin{aligned} V^{12} &:= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right) \\ &= \frac{1}{n}\mathrm{Var}\left(W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)\right) \\ &= \frac{1}{n}\mathrm{E}\left[(W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l))^2\right] + O(\frac{b^4}{n}) \\ &= \frac{1}{n}\int_{R^p} (W_b(\tilde{\mathbf{x}} - \mathbf{x})(\tilde{x}_l - x_l))^2 f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} + O(\frac{b^4}{n}). \end{aligned}$$

By using the definition of the product kernel $W_b(\cdot)$ we get

$$V^{12} = \frac{1}{n}\int_{R^p} \frac{1}{b^{2p}} \left\{\prod_{j=1}^{p} K\left(\frac{\tilde{x}_j - x_j}{b}\right)^2\right\} (\tilde{x}_l - x_l)^2 f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} + O(\frac{b^4}{n}).$$

Change of variable $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$ leads to

$$V^{12} = \frac{1}{nb^p}\int_{R^p} \left\{\prod_{j=1}^{p} K(\bar{x}_j)^2\right\} (b\bar{x}_l)^2 f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x}) d\bar{\mathbf{x}} + O(\frac{b^4}{n}).$$

Since $\int_R x^k K(x)^2\, dx < \infty$ for $k \in \{0, 2, 3\}$ we get by the usual Taylor expansion

$$V^{12} = O\left(\frac{1}{nb^{p-2}}\right) + O\left(\frac{b^4}{n}\right) = O\left(\frac{1}{nb^{p-2}}\right) = o(1),$$

because $\frac{b^4}{n} = O\left(\frac{1}{nb^{p-2}}\right)$ and $nb^{p+2}|H|^{\frac{1}{2}} \to \infty$.

Now we analyze $\mathcal{M}_1^{21}$

$$\begin{aligned} \mathcal{M}_1^{21} &= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(\mathbf{X}_i - \mathbf{x})\right] \\ &= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(X_{i,1} - x_1) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(X_{i,p} - x_p) \end{pmatrix} \in \mathbb{R}^{p\times 1}. \end{aligned}$$

In the following we will show that for $l \in \{1, \dots, p\}$,

$$E^{21} := \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(X_{i,l} - x_l)\right] = \frac{\partial}{\partial x_l} f_{\mathbf{X}}(\mathbf{x})\mu_K + o(1),$$
$$V^{21} := \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(X_{i,l} - x_l)\right] = o(1).$$

By the same argument as used in the proof of $\mathcal{M}_1^{12}$ before this shows that $\mathcal{M}_1^{21} = \mu_K \nabla f_{\mathbf{X}}(\mathbf{x}) + o_p(1)$.

We have already shown in the proof of $\mathcal{M}_1^{12}$ that

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] = b^2 \frac{\partial f}{\partial x_l}(\mathbf{x})\mu_K + O(b^3),$$

and there fore we get for $l \in \{1, \dots, p\}$,

$$\begin{aligned}\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(X_{i,l} - x_l)\right] &= b^{-2}\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] \\ &= \frac{\partial f}{\partial x_l}(\mathbf{x})\mu_K + o(1).\end{aligned}$$

We also showed before that for $l \in \{1, \dots, p\}$,

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] = O\left(\frac{1}{nb^{p-2}}\right),$$

and therefore

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(X_{i,l} - x_l)\right] &= \frac{1}{b^4}\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)\right] \\ &= O\left(\frac{1}{nb^{p+2}}\right) = o(1).\end{aligned}$$

Note here that we assumed that $n|H|^{\frac{1}{2}} b^{p+2} \to \infty$ and therefore $nb^{p+2} \to \infty$.

The remaining part is to analyze $\mathcal{M}_1^{22} \in \mathbb{R}^{p\times p}$:

$$\begin{aligned}
\mathcal{M}_1^{22} &= \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})b^{-2}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^\top\right] \\
&= b^{-2}\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix} X_{i,1}-x_1 \\ \vdots \\ X_{i,p}-x_p \end{pmatrix}\begin{pmatrix} X_{i,1}-x_1 & \dots & X_{i,p}-x_p \end{pmatrix} \\
&= b^{-2}\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})\begin{pmatrix} (X_{i,1}-x_1)^2 & \dots & (X_{i,1}-x_1)(X_{i,p}-x_p) \\ \vdots & \ddots & \vdots \\ (X_{i,p}-x_p)(X_{i,1}-x_1) & \dots & (X_{i,p}-x_p)^2 \end{pmatrix}.
\end{aligned}$$

We will show that for $l, m \in \{1,\dots,p\}$,

$$\begin{aligned}
E^{22} &:= \mathrm{E}\left[b^{-2}\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})(X_{i,l}-x_l)^2\right] = f_{\mathbf{X}}(\mathbf{x})\mu_K + o(1), \\
V^{22} &:= \mathrm{Var}\left[b^{-2}\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})(X_{i,l}-x_l)(X_{i,m}-x_m)\right] = o(1),
\end{aligned}$$

and for $l, m \in \{1,\dots,p\}$ with $l \neq m$,

$$E_{lm}^{22} := \mathrm{E}\left[b^{-2}\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i-\mathbf{x})(X_{i,l}-x_l)(X_{i,m}-x_m)\right] = 0.$$

This is sufficient to prove

$$\mathcal{M}_1^{22} = \mu_K f_{\mathbf{X}}(\mathbf{x}) I_p + o_p(1)\mathbf{1}_{p\times p}.$$

For $l \in \{1,\dots,p\}$ it holds

$$\begin{aligned}
E^{22} &= b^{-2}\mathrm{E}\left[W_b(\mathbf{X}_1-\mathbf{x})(X_{1,l}-x_l)^2\right] \\
&= b^{-2}\int_{R^p} W_b(\tilde{\mathbf{x}}-\mathbf{x})(\tilde{x}_l-x_l)^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}.
\end{aligned}$$

By using the definition of the product kernel $W_b(\cdot)$ we get

$$E^{22} = \int_{R^p} \frac{1}{b^{p+2}}\left\{\prod_{j=1}^{p} K\left(\frac{\tilde{x}_j - x_j}{b}\right)\right\}(\tilde{x}_l-x_l)^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}.$$

By using this and change of variable $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$,

$$\begin{aligned}
E^{22} &= \int_{R^p} \left\{ \prod_{j=1}^{p} K\left(\bar{x}_j\right) \right\} \bar{x}_l^2 f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x}) d\bar{\mathbf{x}} \\
&= \int_{R^p} \left\{ \prod_{j=1}^{p} K\left(\bar{x}_j\right) \right\} \bar{x}_l^2 \left\{ f_{\mathbf{X}}(\mathbf{x}) + b \sum_{r\in\{1,\ldots,p\}} O\Big(\bar{x}_r\Big) \right\} d\bar{\mathbf{x}} \\
&= f_{\mathbf{X}}(\mathbf{x}) \underbrace{\int_R x_l^2 K\left(\bar{x}_l\right) d\bar{x}_l}_{=\mu_K} \prod_{\substack{j=1\\ j\neq l}}^{p} \underbrace{\int_R K\left(\bar{x}_j\right) d\bar{x}_j}_{=1} + O(b) \\
&= f_{\mathbf{X}}(\mathbf{x})\mu_K + o(1).
\end{aligned}$$

For $l, m \in \{1, \ldots, p\}$ and $l \neq m$,

$$\begin{aligned}
E_{lm}^{22} &= b^{-2}\mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)(X_{1,m} - x_m)\right] \\
&= b^{-2} \int_{R^p} W_b(\tilde{\mathbf{x}} - \mathbf{x})(\tilde{x}_l - x_l)(\tilde{x}_m - x_m) f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \\
&= \int_{R^p} \frac{1}{b^{p+2}} \left\{ \prod_{j=1}^{p} K\left(\frac{\tilde{x}_j - x_j}{b}\right) \right\} (\tilde{x}_l - x_l)(\tilde{x}_m - x_m) f_{\mathbf{X}}(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}}.
\end{aligned}$$

By using this and change of variable $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$,

$$\begin{aligned}
E_{lm}^{22} &= \int_{R^p} \left\{ \prod_{j=1}^{p} K\left(\bar{x}_j\right) \right\} \bar{x}_l \bar{x}_m f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x}) d\bar{\mathbf{x}} \\
&= \int_{R^p} \left\{ \prod_{j=1}^{p} K\left(\bar{x}_j\right) \right\} \bar{x}_l \bar{x}_m \left\{ f_{\mathbf{X}}(\mathbf{x}) + b \sum_{r\in\{1,\ldots,p\}} O\Big(\bar{x}_r\Big) \right\} d\bar{\mathbf{x}} \\
&= f_{\mathbf{X}}(\mathbf{x}) \left\{ \prod_{j\in\{l,m\}} \underbrace{\int_R x_j K\left(\bar{x}_j\right) d\bar{x}_j}_{=0} \right\} \left\{ \prod_{\substack{j=1\\ j\notin\{l,m\}}}^{p} \underbrace{\int_R K\left(\bar{x}_j\right) d\bar{x}_j}_{=1} \right\} \\
&\quad + b \sum_{r=1}^{p} \int_{R^p} \left\{ \prod_{j=1}^{p} K\left(\bar{x}_j\right) \right\} \bar{x}_l \bar{x}_m O\Big(\bar{x}_r\Big) d\bar{\mathbf{x}} \\
&= 0.
\end{aligned}$$

Note that the second summand was also zero by Fubini. We have calculated that for $l, m \in \{1, \ldots, p\}$ that $E^{22} = f_{\mathbf{X}}(\mathbf{x})\mu_K + o(1)$ therefore

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)^2\right] = o(1),$$

and for $l \neq m$ we have $E^{22}_{lm} = 0$ therefore

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})(X_{i,l} - x_l)(X_{i,m} - x_m)\right] = 0.$$

This implies now that for $l, m \in \{1, \ldots, p\}$,

$$\begin{aligned}
V^{22} &= \frac{1}{nb^{-4}}\mathrm{Var}\left[W_b(\mathbf{X}_1 - \mathbf{x})(X_{1,l} - x_l)(X_{1,m} - x_m)\right] \\
&= \frac{1}{nb^{-4}}\mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})^2(X_{1,l} - x_l)^2(X_{1,m} - x_m)^2\right] + o(1) \\
&= \frac{1}{nb^{-4}}\int_{R^p} W_b(\tilde{\mathbf{x}} - \mathbf{x})^2(\tilde{x}_l - x_l)^2(\tilde{x}_m - x_m)^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}} + o(1) \\
&= \frac{1}{nb^{-4}}\int_{R^p} \frac{1}{b^{2p}}\left\{\prod_{j=1}^{p} K\left(\frac{\tilde{x}_j - x_j}{b}\right)^2\right\}(\tilde{x}_l - x_l)^2(\tilde{x}_m - x_m)^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}} + o(1).
\end{aligned}$$

By using this and change of variable $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$

$$\begin{aligned}
V^{22} &= \frac{1}{nb^{-4}}\frac{1}{b^p}\int_{R^p}\left\{\prod_{j=1}^{p} K\left(\bar{x}_j\right)^2\right\}(b\bar{x}_l)^2(b\bar{x}_m)^2 f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x})d\bar{\mathbf{x}} + o(1) \\
&= \frac{1}{nb^p}\int_{R^p}\left\{\prod_{j=1}^{p} K\left(\bar{x}_j\right)^2\right\}\bar{x}_l^2\bar{x}_m^2\left(f_{\mathbf{X}}(\mathbf{x}) + b\sum_{r=1}^{p} O\left(\bar{x}_r\right)\right)d\bar{\mathbf{x}} + o(1) \\
&= O(\frac{1}{nb^p}) + o(1) = o(1),
\end{aligned}$$

where we used in the last step that $\int_{\mathbb{R}} |z^5|K(z)dz < \infty$. This completes the proof of Lemma A.7. □

We begin with the proofs of our main lemmas (Lemma A.1 to Lemma A.4). It will be more convenient for us to prove the stated lemmas in the following (reversed-) order: Lemma A.4, Lemma A.3, Lemma A.2 (Page 102) and then finally Lemma A.1 (Page 105).

Proof. (Lemma A.4)
This Lemma follows directly by the proof of Lemma 4.1. □

Proof. (Lemma A.3)
The statement we want to show is:

$$e_1^\top \mathcal{D}\mathcal{M}_3 = \sqrt{n|H|^{\frac{1}{2}}b^p} \cdot e_1^\top \mathcal{M}_3 \to N\left(0, R_{\mathcal{K}}\nu_K^p f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})\right),$$

where

$$\mathcal{M}_3 := \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})\begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x})\end{pmatrix}\{\epsilon_i\}\right] \in \mathbb{R}^{p+1},$$

with

$$\epsilon_i := K_H(\mathbf{Z}_i - \mathbf{z}) - E\left[K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right].$$

We established in Lemma A.6 that

$$\begin{aligned}
\mathrm{E}\left[\mathcal{M}_3\right] &= \mathbf{0}_{p+1}, \\
\mathrm{Var}\left[e_1^\top \mathcal{M}_3\right] &= \frac{1}{n|H|^{\frac{1}{2}} b^p} \nu_K^p R_{\mathcal{K}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + o\left(\frac{1}{n|H|^{\frac{1}{2}} b^p}\right).
\end{aligned}$$

Usual variance properties yields

$$\mathrm{Var}\left(\sqrt{n|H|^{\frac{1}{2}} b^p} \cdot e_1^\top \mathcal{M}_3\right) = \nu_K^p R_{\mathcal{K}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + o\left(1\right).$$

So the only thing left to show is the asymptotic normality. To show this we use the Liapunov Central Limit Theorem (Theorem 2.3). We have

$$\sqrt{n|H|^{\frac{1}{2}} b^p} \cdot e_1^\top \mathcal{M}_3 = \sum_{i=1}^n \frac{\sqrt{|H|^{\frac{1}{2}} b^p}}{\sqrt{n}} W_b(\mathbf{X}_i - \mathbf{x}) \epsilon_i.$$

Define

$$T_{n,i} := \frac{\sqrt{|H|^{\frac{1}{2}} b^p}}{\sqrt{n}} W_b(\mathbf{X}_i - \mathbf{x}) \epsilon_i$$

and

$$S_n := \sqrt{n|H|^{\frac{1}{2}} b^p} \cdot e_1^\top \mathcal{M}_3 = \sum_{i=1}^n T_{n,i}.$$

We know that $\mathrm{Var}(S_n) = \nu_K^p R_{\mathcal{K}} f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + o\left(1\right) = Const + o(1)$. Using this notation we see that the desired statement follows immediately by the Liapunov CLT, if we show the Liapunov condition (2.7):

$$\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}[|T_{n,i} - E[T_{n,i}]|^3] = 0.$$

We already calculated that $E(T_{n,i}) = 0$ and we know that the $T_{n,i}$ are *i.i.d.*. Therefore we have to show that:

$$\lim_{n\to\infty} n\mathrm{E}[|T_{n,1}|^3] = 0.$$

It holds that

$$
\begin{aligned}
n\mathrm{E}[|T_{n,1}|^3] &= n\mathrm{E}\left[\left|\frac{\sqrt{|H|^{\frac{1}{2}}b^p}}{\sqrt{n}}W_b(\mathbf{X}_1-\mathbf{x})\epsilon_1\right|^3\right] \\
&= \frac{|H|^{\frac{3}{4}}b^{\frac{3}{2}p}}{\sqrt{n}}\mathrm{E}\left[\left|W_b(\mathbf{X}_1-\mathbf{x})\epsilon_1\right|^3\right] \\
&= \frac{|H|^{\frac{3}{4}}b^{\frac{3}{2}p}}{\sqrt{n}}\mathrm{E}\left[\left|W_b(\mathbf{X}_1-\mathbf{x})\{K_H(\mathbf{Z}_1-\mathbf{z})-E\left[K_H(\mathbf{Z}_1-\mathbf{z})|\mathbf{X}_i\right]\}\right|^3\right].
\end{aligned}
$$

We get:

$$
\begin{aligned}
n\mathrm{E}[|T_{n,1}|^3] &= \frac{|H|^{\frac{3}{4}}b^{\frac{3}{2}p}}{\sqrt{n}|H|^{\frac{3}{2}}b^{3p}}\mathrm{E}\left[\left|K^p\left(\frac{\mathbf{X}_1-\mathbf{x}}{b}\right)\left\{\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.-E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1\right]\right\}\right|^3\right] \\
&\leq \frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\mathrm{E}\left[K^p\left(\frac{\mathbf{X}_1-\mathbf{x}}{b}\right)^3\left\{\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)\right.\right. \\
&\qquad\qquad \left.\left.+E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1\right]\right\}^3\right] \\
&= \mathscr{V}_1+\mathscr{V}_2,
\end{aligned}
$$

where

$$
\begin{aligned}
\mathscr{V}_1 := \frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\mathrm{E}\Bigg[K^p\left(\frac{\mathbf{X}_1-\mathbf{x}}{b}\right)^3\Bigg\{&\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)^3 \\
&+3\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)^2 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1\right] \\
&+\ 3\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1\right]^2\Bigg\}\Bigg],
\end{aligned}
$$

$$
\mathscr{V}_2 := \frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\mathrm{E}\left[K^p\left(\frac{\mathbf{X}_1-\mathbf{x}}{b}\right)^3 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1\right]^3\right].
$$

First we analyze $\mathscr{V}_2$:

$$
\mathscr{V}_2 = \frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\int_{R^p}K^p\left(\frac{\tilde{\mathbf{x}}-\mathbf{x}}{b}\right)^3 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1=\tilde{\mathbf{x}}\right]^3 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}.
$$

We already showed in the proof of Lemma A.5 that

$$E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)\middle|\mathbf{X}_1=\tilde{\mathbf{x}}\right]=|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}})+O(|H|^{\frac{1}{2}}).$$

Therefore,

$$\mathscr{V}_2=\frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\int_{R^p}K^p\left(\frac{\tilde{\mathbf{x}}-\mathbf{x}}{b}\right)^3\left[|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}})+O(|H|^{\frac{1}{2}})\right]^3 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}.$$

The next step is the usual substitution with $\tilde{\mathbf{x}}=(b\bar{\mathbf{x}}+\mathbf{x})$:

$$\mathscr{V}_2=\frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\int_{R^p}K^p\left(\bar{\mathbf{x}}\right)^3\left[|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}}+\mathbf{x})+O(|H|^{\frac{1}{2}})\right]^3 f_{\mathbf{X}}(b\bar{\mathbf{x}}+\mathbf{x})b^p d\bar{\mathbf{x}}.$$

Then using the Taylor expansions

$$f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}}+\mathbf{x})=f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right),$$
$$f_{\mathbf{X}}(b\bar{\mathbf{x}}+\mathbf{x})=f_{\mathbf{X}}(\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right),$$

we get

$$\begin{aligned}\mathscr{V}_2&=\frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\int_{R^p}K^p\left(\bar{\mathbf{x}}\right)^3\left[|H|^{\frac{1}{2}}(f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right))+O(|H|^{\frac{1}{2}})\right]^3\\&\qquad\cdot\left(f_{\mathbf{X}}(\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right)\right)b^p d\bar{\mathbf{x}}\\&=\frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}O(|H|^{\frac{3}{2}}b^p)\int_{R^p}K^p\left(\bar{\mathbf{x}}\right)^3\left[\left(f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right)\right)+O(1)\right]^3\\&\qquad\cdot\left(f_{\mathbf{X}}(\mathbf{x})+\sum_{r=1}^{p}O\left(b\bar{x}_r\right)\right)d\bar{\mathbf{x}}\\&=\frac{O(|H|^{\frac{3}{2}}b^p)}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}=O\left(\frac{\sqrt{|H|^{\frac{3}{2}}}}{\sqrt{nb^p}}\right)=o(1).\end{aligned}$$

We used here that $\int_R K\left(\bar{x}\right)^3\bar{x}^4 d\bar{\mathbf{x}}<\infty$ (because we assume that $\int_{\mathbb{R}}|z^5|K(z)dz<\infty$) as well as $nb^{p+2}|H|^{\frac{1}{2}}\to\infty$ and therefore $nb^p\to\infty$.

Finally we analyze $\mathscr{V}_1$:

$$\mathscr{V}_1 := \frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\mathrm{E}\left[K^p\left(\frac{\mathbf{X}_1-\mathbf{x}}{b}\right)^3\left\{\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)^3\right.\right.$$

$$+3\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)^2 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1\right]$$

$$\left.\left.+\ 3\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1\right]^2\right\}\right]$$

$$=\frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\int_{\mathbb{R}^{d+p}}\left[K^p\left(\frac{\tilde{\mathbf{x}}-\mathbf{x}}{b}\right)^3\left\{\mathcal{K}\left(H^{-\frac{1}{2}}(\tilde{\mathbf{z}}-\mathbf{z})\right)^3\right.\right.$$

$$+3\mathcal{K}\left(H^{-\frac{1}{2}}(\tilde{\mathbf{z}}-\mathbf{z})\right)^2 E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1=\tilde{\mathbf{x}}\right]$$

$$\left.\left.+\ 3\mathcal{K}\left(H^{-\frac{1}{2}}(\tilde{\mathbf{z}}-\mathbf{z})\right)E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)|\mathbf{X}_1=\tilde{\mathbf{x}}\right]^2\right\}\right]$$

$$f_{\mathbf{Z},\mathbf{X}}(\tilde{\mathbf{z}},\tilde{\mathbf{x}})d(\tilde{\mathbf{z}},\tilde{\mathbf{x}}).$$

Again we use the that we shown in the proof of Lemma A.5 that

$$E\left[\mathcal{K}\left(H^{-\frac{1}{2}}(\mathbf{Z}_1-\mathbf{z})\right)\middle|\mathbf{X}_1=\tilde{\mathbf{x}}\right]=|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\tilde{\mathbf{x}})+O(|H|^{\frac{1}{2}}).$$

By using this and change of variable $\tilde{\mathbf{x}}=(b\bar{\mathbf{x}}+\mathbf{x})$ and $\tilde{\mathbf{z}}=(H^{\frac{1}{2}}\bar{\mathbf{z}}+\mathbf{z})$ we get

$$\mathscr{V}_1=\frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}}\int_{\mathbb{R}^{d+p}}\left[K^p\left(\bar{\mathbf{x}}\right)^3\left\{\mathcal{K}(\bar{\mathbf{z}})^3\right.\right.$$

$$+3\mathcal{K}(\bar{\mathbf{z}})^2(|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}}+\mathbf{x})+O(|H|^{\frac{1}{2}}))$$

$$\left.\left.+\ 3\mathcal{K}(\bar{\mathbf{z}})(|H|^{\frac{1}{2}}f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}}+\mathbf{x})+O(|H|^{\frac{1}{2}}))^2\right\}\right]$$

$$f_{\mathbf{Z},\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}}+\mathbf{z},b\bar{\mathbf{x}}+\mathbf{x})|H|^{\frac{1}{2}}b^p d(\bar{\mathbf{z}},\bar{\mathbf{x}}).$$

Now we use the Taylor expansions

$$f_{\mathbf{Z},\mathbf{X}}(H^{\frac{1}{2}}\bar{\mathbf{z}}+\mathbf{z}, b\bar{\mathbf{x}}+\mathbf{x}) = f_{\mathbf{Z},\mathbf{X}}(\mathbf{z},\mathbf{x}) + \sum_{\substack{(k,r)\\ \in\{1,\dots,d\}^2}} O\left([H^{\frac{1}{2}}]_{k,r}\bar{z}_r\right) + \sum_{r=1}^{p} O\left(b\bar{x}_r\right),$$

$$f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|b\bar{\mathbf{x}}+\mathbf{x}) = f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \sum_{r=1}^{p} O\left(b\bar{x}_r\right),$$

and get

$$\begin{aligned}\mathscr{V}_1 = \frac{1}{\sqrt{n|H|^{\frac{3}{2}}b^{3p}}} \int_{\mathbb{R}^{d+p}} \Bigg[K^p(\bar{\mathbf{x}})^3 \Bigg\{ &\mathcal{K}(\bar{\mathbf{z}})^3 \\ &+ 3\mathcal{K}(\bar{\mathbf{z}})^2 \left(|H|^{\frac{1}{2}} \left[f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \sum_{r=1}^{p} O\left(b\bar{x}_r\right)\right] + O(|H|^{\frac{1}{2}})\right) \\ &+ 3\mathcal{K}(\bar{\mathbf{z}}) \left(|H|^{\frac{1}{2}} \left[f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) + \sum_{r=1}^{p} O\left(b\bar{x}_r\right)\right] + O(|H|^{\frac{1}{2}})\right)^2 \Bigg\}\Bigg] \\ &\left[f_{\mathbf{Z},\mathbf{X}}(\mathbf{z},\mathbf{x}) + \sum_{\substack{(k,r)\\ \in\{1,\dots,d\}^2}} O\left([H^{\frac{1}{2}}]_{k,r}\bar{z}_r\right) + \sum_{r=1}^{p} O\left(b\bar{x}_r\right)\right] |H|^{\frac{1}{2}} b^p d(\bar{\mathbf{z}},\bar{\mathbf{x}}).\end{aligned}$$

After factoring $|H|^{\frac{1}{2}}b^p$ out and using the kernel properties, we get

$$\mathscr{V}_1 = O\left(\frac{1}{\sqrt{n|H|^{\frac{1}{2}}b^p}}\right) = o(1).$$

Therefore the Liapunov condition is satisfied. □

Proof. (Lemma A.2)
This proof is heavily based on the results in Ruppert and Wand (1994). We want to prove that

$$\mathcal{M}_2 = \begin{pmatrix} f_{\mathbf{X}}(\mathbf{x})\mu_K \sum_{j=1}^{p} \frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + O_p(b^3) \\ O_p(b)\mathbf{1}_p \end{pmatrix}, \tag{A.8}$$

and therefore also,

$$e_1^\top \mathcal{D}\mathcal{M}_2 = f_{\mathbf{X}}(\mathbf{x})\sqrt{n|H|^{\frac{1}{2}}b^p}\mu_K \sum_{j=1}^{p} \frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + o_p(1), \tag{A.9}$$

where $\mathbf{1}_p$ is the vector in $\mathbb{R}^p$ where every entry is one, and

$$\mathcal{M}_2 = \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\}\right] \in \mathbb{R}^{p+1},$$

$$= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\} \\ \frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(\mathbf{X}_i - \mathbf{x}) \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\} \end{pmatrix}$$

and $\mathcal{H}_g(\mathbf{x})$ is the Hessian of $g(\mathbf{x})$.

First note that if we show that equation (A.8) holds we automatically have verified equation (A.9) by just multiplying the matrix product out and using that $\sqrt{n|H|^{\frac{1}{2}} b^p} b^3 \to 0$.

We begin the proof of equation (A.8) with the lower part of the the vector $\mathcal{M}_2$. Multiplying equation (2.13) from Ruppert and Wand (1994) by $\frac{1}{2b^2}$ and taking their matrix $H = b^2 I_{d\times d}$ (where $I_{d\times d}$ is the identity matrix in $R^{d\times d}$) leads to

$$\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) b^{-2}(\mathbf{X}_i - \mathbf{x}) \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\}$$
$$= \frac{1}{2b^2} O_p(b^3)\mathbf{1}_p = O_p(b).$$

So the only thing left to show is that

$$\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\}$$
$$= f_{\mathbf{X}}(\mathbf{x})\mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + O_p(b^3).$$

Using equation (2.14) in Ruppert and Wand (1994) yields

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\}\right]$$
$$= f_{\mathbf{X}}(\mathbf{x})\mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 g(\mathbf{x})}{\partial x_j^2} + O(b^3). \tag{A.10}$$

Remark A.1. *To see why equation (2.14) in Ruppert and Wand (1994) yields (A.10), consider there the (non-conditional-) expectation in the second line. This expectation is derived by a Taylor expansion. Analyzing the non-stochastic part of the term $o_p(\mathrm{tr}(H))$ related to this expectation, unveils that it equals $O(\mathrm{tr}(H) \cdot \mathrm{tr}(\boldsymbol{1}_{d\times d} H^{\frac{1}{2}}))$. Multiplying this result by $f_{\boldsymbol{X}}(\boldsymbol{x})$ and setting their $H = b^2 I_{d\times d}$ results in (A.10).*

We will now show that

$$\mathscr{V}_3 := \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x})\left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\}\right] = o(1).$$

It holds,

$$\begin{aligned}
\mathscr{V}_3 &= \frac{1}{n}\mathrm{Var}\left[W_b(\mathbf{X}_1 - \mathbf{x})\left\{\frac{1}{2}(\mathbf{X}_1 - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_1 - \mathbf{x})\right\}\right] \\
&= \frac{1}{n}\left\{\mathrm{E}\left[\left(W_b(\mathbf{X}_1 - \mathbf{x})\left\{\frac{1}{2}(\mathbf{X}_1 - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_1 - \mathbf{x})\right\}\right)^2\right]\right. \\
&\qquad \left. - \mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})\left\{\frac{1}{2}(\mathbf{X}_1 - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_1 - \mathbf{x})\right\}\right]^2\right\}.
\end{aligned}$$

Using the previous result about the expectation and the linearity we get

$$\begin{aligned}
\mathscr{V}_3 &= \frac{1}{n}\mathrm{E}\left[\left(W_b(\mathbf{X}_1 - \mathbf{x})\left\{\frac{1}{2}(\mathbf{X}_1 - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_1 - \mathbf{x})\right\}\right)^2\right] + O\left(\frac{b^4}{n}\right) \\
&= \frac{1}{4n}\mathrm{E}\left[W_b(\mathbf{X}_1 - \mathbf{x})^2\left\{(\mathbf{X}_1 - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_1 - \mathbf{x})\right\}^2\right] + o(1) \\
&= \frac{1}{4n}\int_{\mathbb{R}^p} W_b(\tilde{\mathbf{x}} - \mathbf{x})^2\left\{(\tilde{\mathbf{x}} - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\tilde{\mathbf{x}} - \mathbf{x})\right\}^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}} + o(1) \\
&= \frac{1}{4nb^{2p}}\int_{\mathbb{R}^p} K^p\left(\frac{\tilde{\mathbf{x}} - \mathbf{x}}{b}\right)^2\left\{(\tilde{\mathbf{x}} - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\tilde{\mathbf{x}} - \mathbf{x})\right\}^2 f_{\mathbf{X}}(\tilde{\mathbf{x}})d\tilde{\mathbf{x}} + o(1).
\end{aligned}$$

Now we substitute with $\tilde{\mathbf{x}} = (b\bar{\mathbf{x}} + \mathbf{x})$

$$\begin{aligned}
\mathscr{V}_3 &= \frac{1}{4nb^p}\int_{\mathbb{R}^p} K^p(\bar{\mathbf{x}})^2\left\{(b\bar{\mathbf{x}})^\top \mathcal{H}_g(\mathbf{x})(b\bar{\mathbf{x}})\right\}^2 f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x})d\bar{\mathbf{x}} + o(1) \\
&= \frac{1}{4nb^{p-2}}\int_{\mathbb{R}^p} K^p(\bar{\mathbf{x}})^2\left\{\bar{\mathbf{x}}^\top \mathcal{H}_g(\mathbf{x})\bar{\mathbf{x}}\right\}^2\left\{f_{\mathbf{X}}(\mathbf{x}) + O\left(b\sum_{j=1}^{p}\bar{x}_j\right)\right\}d\bar{\mathbf{x}} + o(1) \\
&= O\left(\frac{1}{nb^{p-2}}\right) + o(1) = o(1).
\end{aligned}$$

Recall for the linear Taylor expansions of $f_{\mathbf{X}}(b\bar{\mathbf{x}} + \mathbf{x})$ (see Corollary 2.1) that we assume $f_{\mathbf{X}}(\mathbf{x})$ to have bounded continuous first order derivatives with respect to $\mathbf{x}$ (see Remark 3.1). Note also that we used in the last step that

$$\int_{\mathbb{R}^p} K^p(\bar{\mathbf{x}})^2\left\{\bar{\mathbf{x}}^\top \mathcal{H}_g(\mathbf{x})\bar{\mathbf{x}}\right\}^2(1 + \bar{x}_j)d\bar{\mathbf{x}} < \infty \ \forall j \in \{1, \ldots, p\}$$

as well as $nb^{p+2}|H|^{\frac{1}{2}} \to \infty$ and therefore $nb^{p-2} \to \infty$.

Now since a vanishing variance implies convergence in L^2 to the expectation, and convergence in L^2 implies convergence in probability, we get that

$$\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \left\{ \frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) \right\}$$
$$= f_{\mathbf{X}}(\mathbf{x})\mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 g(\mathbf{x})}{\partial x_j^2} + O_p(b^3).$$

With Lemma A.4 we can now conclude that for $j \in \{1, \ldots, p\}$

$$\frac{\partial^2 g(\mathbf{x})}{\partial x_j^2} = \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + O(\mathrm{tr}(\mathbf{1}_{d\times d}H)).$$

Combining this with the equation before leads to

$$\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \left\{ \frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) \right\}$$
$$= f_{\mathbf{X}}(\mathbf{x})\mu_K \sum_{j=1}^{p} \frac{b^2}{2} \frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + O(b^2\mathrm{tr}(\mathbf{1}_{d\times d}H)) + O_p(b^3).$$

Since we assume $\frac{\mathrm{tr}(\mathbf{1}_{d\times d}H)}{b} = O(1)$ to hold we know that $b^2\mathrm{tr}(\mathbf{1}_{d\times d}H) = O(b^3)$ holds. This completes the final part for the proof of Lemma A.2. □

Proof. (Lemma A.1) The statement we want to prove is:

$$e_1^\top \mathcal{D}\mathcal{M}_1^{-1}\left[\mathcal{M}_2 + \mathcal{M}_3\right] + e_1^\top \mathcal{D}(s.o.) = \frac{1}{f_{\mathbf{X}}(\mathbf{x})} e_1^\top \mathcal{D}\left[\mathcal{M}_2 + \mathcal{M}_3\right] + o_p(1),$$

where

$$\mathcal{M}_1 = \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \left(1, (\mathbf{X}_i - \mathbf{x})^\top\right)\right] \in \mathbb{R}^{(p+1)\times(p+1)},$$
$$\mathcal{M}_2 = \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \left\{\frac{1}{2}(\mathbf{X}_i - \mathbf{x})^\top \mathcal{H}_g(\mathbf{x})(\mathbf{X}_i - \mathbf{x})\right\}\right] \in \mathbb{R}^{p+1},$$
$$\mathcal{M}_3 = \left[\frac{1}{n}\sum_{i=1}^{n} W_b(\mathbf{X}_i - \mathbf{x}) \begin{pmatrix} 1 \\ b^{-2}(\mathbf{X}_i - \mathbf{x}) \end{pmatrix} \{\epsilon_i\}\right] \in \mathbb{R}^{p+1},$$

with

$$\epsilon_i = K_H(\mathbf{Z}_i - \mathbf{z}) - E\left[K_H(\mathbf{Z}_i - \mathbf{z})|\mathbf{X}_i\right].$$

Note that the term $e_1^\top \mathcal{D}(s.o.)$ will be of smaller order than $e_1^\top \mathcal{D}\mathcal{M}_1^{-1}\mathcal{M}_2$. Since we will see that the order of $e_1^\top \mathcal{D}\mathcal{M}_1^{-1}\mathcal{M}_2$ is already low, we will disclaim the calculation of $e_1^\top \mathcal{D}(s.o.)$.

Using now Lemma A.7 and blockwise inversion of the matrix $\mathcal{M}_1$ one can show that

$$\mathcal{M}_1^{-1} = \begin{pmatrix} \frac{1}{f_{\mathbf{X}}(\mathbf{x})} & \mathbf{0}_p^\top \\ -\frac{1}{f_{\mathbf{X}}(\mathbf{x})^2}\nabla f_{\mathbf{X}}(\mathbf{x}) & \frac{1}{\mu_K f_{\mathbf{X}}(\mathbf{x})} I_p \end{pmatrix} + \begin{pmatrix} O_p(b^2) & O_p(b^2)\mathbf{1}_p^\top \\ o_p(1)\mathbf{1}_p & o_p(1)\mathbf{1}_{p\times p} \end{pmatrix}.$$

With this we can rewrite

$$\begin{aligned} e_1^\top \mathcal{D}\mathcal{M}_1^{-1}\left[\mathcal{M}_2+\mathcal{M}_3\right] &= e_1^\top \mathcal{D}\begin{pmatrix} \frac{1}{f_{\mathbf{X}}(\mathbf{x})} & \mathbf{0}_p^\top \\ -\frac{1}{f_{\mathbf{X}}(\mathbf{x})^2}\nabla f_{\mathbf{X}}(\mathbf{x}) & \frac{1}{\mu_K f_{\mathbf{X}}(\mathbf{x})} I_p \end{pmatrix}\left[\mathcal{M}_2+\mathcal{M}_3\right] \\ &\quad + e_1^\top \mathcal{D}\left[\mathcal{M}_1^{-1} - \begin{pmatrix} \frac{1}{f_{\mathbf{X}}(\mathbf{x})} & \mathbf{0}_p^\top \\ -\frac{1}{f_{\mathbf{X}}(\mathbf{x})^2}\nabla f_{\mathbf{X}}(\mathbf{x}) & \frac{1}{\mu_K f_{\mathbf{X}}(\mathbf{x})} I_p \end{pmatrix}\right]\left[\mathcal{M}_2+\mathcal{M}_3\right] \\ &= \frac{1}{f_{\mathbf{X}}(\mathbf{x})} e_1^\top \mathcal{D}\left[\mathcal{M}_2+\mathcal{M}_3\right] \\ &\quad + e_1^\top \mathcal{D}\begin{pmatrix} O_p(b^2) & O_p(b^2)\mathbf{1}_p^\top \\ o_p(1)\mathbf{1}_p & o_p(1)\mathbf{1}_{p\times p} \end{pmatrix}\left[\mathcal{M}_2+\mathcal{M}_3\right] \\ &= \frac{1}{f_{\mathbf{X}}(\mathbf{x})} e_1^\top \mathcal{D}\left[\mathcal{M}_2+\mathcal{M}_3\right] \\ &\quad + \sqrt{nb^p|H|^{\frac{1}{2}}}\, O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_2 \\ &\quad + \sqrt{nb^p|H|^{\frac{1}{2}}}\, O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_3, \end{aligned}$$

where $\mathbf{1}_{p+1}$ is the vector in $\mathbb{R}^{p+1}$, where every entry is one. The only thing left to show is

$$\sqrt{nb^p|H|^{\frac{1}{2}}}\, O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_2 + \sqrt{nb^p|H|^{\frac{1}{2}}}\, O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_3 = o_p(1).$$

We split this problem in two parts. First we show that

$$\sqrt{nb^p|H|^{\frac{1}{2}}}\, O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_3 = o_p(1),$$

and afterwards that

$$\sqrt{nb^p|H|^{\frac{1}{2}}}\, O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_2 = o_p(1).$$

Beginning with $\sqrt{nb^p|H|^{\frac{1}{2}}}\, O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_3 = o_p(1)$, recall that we showed in Lemma A.6 for $j \in \{1,\dots,p+1\}$,

$$\operatorname{Var}\left[e_j^\top \mathcal{M}_3\right] = \mathrm{E}\left[|e_j^\top \mathcal{M}_3|^2\right] = O\left(\frac{1}{n|H|^{\frac{1}{2}} b^{p+2}}\right),$$

because $\frac{1}{n|H|^{\frac{1}{2}} b^p} = O\left(\frac{1}{n|H|^{\frac{1}{2}} b^{p+2}}\right)$. Note also, that $E(|X_n|^2) = O(a_n)$ for a random

sequence $\{X_n\}_{n=1}^{\infty}$ and non-negative non-stochastic sequence a_n implies $X_n = O_p(\sqrt{a_n})$ (see, Li and Racine, 2006, Theorem A.7.). Therefore we get for $j \in \{1, \ldots, p+1\}$

$$e_j^\top \mathcal{M}_3 = O_p\left(\frac{1}{b\sqrt{n|H|^{\frac{1}{2}}b^p}}\right).$$

It follows that

$$\sqrt{nb^p|H|^{\frac{1}{2}}}O_p(b^2)\mathbf{1}_{p+1}^\top \mathcal{M}_3 = \sum_{j=1}^{p+1}\sqrt{nb^p|H|^{\frac{1}{2}}}O_p(b^2)e_j^\top \mathcal{M}_3 = O_p(b) = o_p(1).$$

It is only left to show that

$$\sqrt{nb^p|H|^{\frac{1}{2}}}O_p(b^2)\mathbf{1}_{1+p}^\top \mathcal{M}_2 = o_p(1).$$

By Lemma A.2 we know that:

$$\mathcal{M}_2 = \begin{pmatrix} f_{\mathbf{X}}(\mathbf{x})\mu_K \sum_{j=1}^{p} \frac{b^2}{2}\frac{\partial^2 f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x})}{\partial x_j^2} + O_p(b^3) \\ O_p(b)\mathbf{1}_p \end{pmatrix} = O_p(b)\mathbf{1}_{p+1},$$

where $\mathbf{1}_p$ is the vector in $\mathbb{R}^p$ and $\mathbf{1}_{p+1}$ is the vector in $\mathbb{R}^{p+1}$ where every entry is one.

Therefore

$$\begin{aligned}\sqrt{nb^p|H|^{\frac{1}{2}}}O_p(b^2)\mathbf{1}_{1+p}^\top \mathcal{M}_2 &= \sqrt{nb^p|H|^{\frac{1}{2}}}O_p(b^2)\mathbf{1}_{1+p}^\top O_p(b)\mathbf{1}_{p+1} \\ &= \sqrt{nb^p|H|^{\frac{1}{2}}}O_p(b^3).\end{aligned}$$

Since we assume that $\sqrt{nb^p|H|^{\frac{1}{2}}}b^3 \to 0$, we get our desired result. With this we have shown the last needed part for the proof of Lemma A.1. □

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182 – 198.

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *Journal of Multivariate Analysis*, 110:43 – 73. Special Issue on Copula Modeling and Dependence.

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics*, 67(2):445–453.

Bergsma, W. (2011). Nonparametric testing of conditional independence by means of the partial copula. *ArXiv e-prints*.

Billingsley, P. (1995). *Probability and Measure*. Wiley, 3 edition.

Czado, C. and Schmidt, T. (2011). *Mathematische Statistik* -. Springer-Verlag, Berlin Heidelberg New York, 2011. aufl. edition.

Duong, T. (2016). *ks: Kernel Smoothing*. R package version 1.10.4.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Number 66 in Monographs on statistics and applied probability series. Chapman & Hall, London [u.a.].

Fan, J., Yao, Q., and Tong, H. (1993). Estimating measures of sensitivity of initial values to nonlinear stochastic systems with chaos *.

Fan, J., Yao, Q., and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83(1):189–206.

Ghazal, G. (1996). Recurrence formula for expectations of products of quadratic forms. *Statistics & Probability Letters*, 27(2):101 – 109.

Gijbels, I., Omelka, M., and Veraverbeke, N. (2012). Multivariate and functional covariates and conditional copulas. *Electron. J. Statist.*, 6:1273–1306.

Gijbels, I., Omelka, M., and Veraverbeke, N. (2015). Estimation of a copula when a covariate affects only marginal distributions. *Scandinavian Journal of Statistics*, 42(4):1109–1126.

Gijbels, I., Veraverbeke, N., and Omelka, M. (2011). Conditional copulas, association measures and their applications. *Comput. Stat. Data Anal.*, 55(5):1919–1932.

Hayfield, T. and Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5).

Klemelä, J. (2009). *Smoothing of multivariate data: density estimation and visualization*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ.

Li, Q. and Racine, J. S. (2006). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Lin, J.-G., Zhang, K.-S., and Zhao, Y.-Y. (2016). Nonparametric estimation of multivariate multiparameter conditional copulas. *Journal of the Korean Statistical Society*.

Loader, C. R. (1999). Bandwidth selection: classical or plug-in? *Ann. Statist.*, 27(2):415–438.

Nagler, T. (2014). Kernel methods for vine copula estimation. Master's thesis, Technische Universität München.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence*. *International Economic Review*, 47(2):527–556.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.

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

Sklar, A. (1959). Fonstions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Spanhel, F. and Kurz, M. S. (2015a). Simplified vine copula models: Approximations based on the simplifying assumption. *ArXiv e-prints*.

Spanhel, F. and Kurz, M. S. (2015b). The partial copula: Properties and associated dependence measures. *ArXiv e-prints*.

Vatter, T. and Chavez-Demoulin, V. (2015). Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141:147 – 167.

Veraverbeke, N., Omelka, M., and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, 38(4):766–780.

Wand, M. (2015). *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*. R package version 2.23-15.

Wand, M. and Jones, M. (1994). *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Wand, M. P. (1992). Error analysis for general multivariate kernel estimators. *Journal of Nonparametric Statistics*, 2(1):1–15.

Yin, J., Geng, Z., Li, R., and Wang, H. (2010). Nonparametric covariance model. *Statistica Sinica*, 20(1):469–479.