Technische Universität München

Department of Mathematics

Nonparametric Estimation of three dimensional Pair-Copula Constructions

Master's Thesis

by

Amina Boukharrata

Supervisor:	Prof. Claudia Czado, Ph. D.
Advisor:	Dr. Nicole Krämer
Submission date:	May 7, 2013

Hiermit erkläre ich, dass ich die Masterarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 07. Mai 2013

Acknowledgments

First of all I would like to express my gratitude to Prof. Czado for giving me the opportunity to work on this interesting topic. I am thankful for her incentives, advice and time, that we spent on discussions. I would also like to thank Dr. Krämer for helping me with all questions and providing helpful instructions.

Furthermore, I want to thank my friends for making the past years of study so enjoyable. In particular, I want to thank Angelina a lot for her support and encouragement. My sincere thanks goes to my family for their love, patience and support they gave me all my life.

Zusammenfassung

Das Ziel dieser Arbeit ist es, ausgewählte Konzepte nichtparametrischer Schätzverfahren für Pair-Copula Constructions vorzustellen. Der Vorteil von nichtparametrischen Schätzern liegt darin, dass sie asymmetrische Abhängigkeiten erfassen können. Der Schwerpunkt liegt auf der Kerndichteschätzung und Bernstein Copulas. Bei Bernstein Copulas, wird das grundlegende Konzept um eine Penalisierung der Koeffizienten erweitert um glatte Schätzfunktionen zu fördern. Die vorgestellten Schätzmethoden werden in einer empirischen Untersuchung und in einer Simulationsstudie angewendet und die nichtparametrisch geschätzten Höhenlinien der Pair-Copula Dichtefunktionen werden den parametrisch geschätzten gegenübergestellt. Für die Beurteilung der Vorhersagegüte des geschätzten Modells werden kreuzvalidierte Likelihoods herangezogen.

Abstract

The goal of this thesis is to review selected concepts of nonparametric estimation of pair-copula constructions. The main advantage of the nonparametric estimation is its flexibility to capture asymmetric dependence. The focus is on the theory of kernel density estimation and Bernstein copulas. For the Bernstein copulas, the basic concept is extended by a penalization on the coefficients to support the smoothness of the function estimate. The presented concepts are applied in an empirical study and a simulation study and nonparametric fitted pair-copula density contours are compared to parametric fitted contours. For comparison of the model prediction quality, cross-validated likelihoods are consulted.

Contents

List of Figures i

List of Tables ii

1 Introduction 1

2 Theoretical Background 3
2.1 Kernel Density Estimation . . . 3
2.1.1 Univariate Kernel Density Estimation . . . 3
2.1.2 Univariate Bandwidth Selection . . . 5
2.1.3 Multivariate Kernel Density Estimation . . . 10
2.1.4 Multivariate Bandwidth Selection . . . 11
2.1.5 Estimating Distribution Functions . . . 12
2.2 Dependence . . . 14
2.2.1 Copulas . . . 15
2.2.2 Pair-Copula Constructions . . . 17
2.2.3 Dependence Measures . . . 20
2.3 Model Comparison using Cross-Validation . . . 21

3 Nonparametric Pair-Copula Constructions 23
3.1 Estimation by Empirical Copulas . . . 23
3.2 Estimation by Kernel Density Functions . . . 26
3.3 Estimation by Bernstein Polynomials . . . 29
3.4 Estimation by penalized Bernstein Polynomials . . . 33

4 Empirical Study 38
4.1 Exploratory Data Analysis . . . 38
4.2 Nonparametric C-Vine Estimation . . . 39
4.3 Comparison to C-Vine with Parametric Pair-Copulas . . . 46
4.4 Cross-Validation . . . 46

5 Simulation Study 52

6 Conclusion 58

Bibliography 59

List of Figures

2.1 Effect of bandwidth and kernel selection on the density estimate. 7
2.2 Original data and copula data. 14
2.3 Nonparametrically estimated univariate cumulative distribution function and probability density function. 15
2.4 Example of five-dimensional C-Vine. 19

3.1 Illustration of standardized Bernstein polynomials. 30
3.2 Margins of estimated pair-copula densities. 34

4.1 Scatter plots, histograms and empirical contours of original uranium data (left hand side) and copula data (right hand side). 39
4.2 Contour plots of estimated pair-copula densities in a C-Vine with pairs 12, 13, 23;1. 42
4.3 Contour plots of estimated pair-copula densities in a C-Vine with pairs 12, 23, 13;2. 43
4.4 Contour plots of estimated pair-copula densities in a C-Vine with pairs 13, 23, 12;3. 44
4.5 Contour plots of nonparametrically and parametrically estimated pair-copula densities in a C-Vine with pairs 12, 23, 13;2. 47
4.6 Boxplots of cross-validated log-likelihoods for nonparametric fitted C-Vines. 49
4.7 Density plots of cross-validated log-likelihoods. 50
4.8 Density plots of cross-validated log-likelihoods per estimation method. . . . 50

5.1 Contour plots of the true pair-copula density for four scenarios. 53
5.2 Contour plots of the true and estimated pair-copula densities in scenario 1. 55
5.3 Contour plots of the true and estimated pair-copula densities in scenario 2. 56
5.4 Contour plots of the true and estimated pair-copula densities in scenario 3. 57

List of Tables

2.1 Some famous kernel functions. 5
2.2 Integrated kernel functions. 13

4.1 Empirical Kendall's tau and Spearman's rho for three-dimensional uranium copula data. 39
4.2 Overview of considered pair-copula constructions. 39
4.3 Overview of considered pair-copula density estimators. 40
4.4 Estimated bandwidths in the nonparametric C-Vine models. 45
4.5 Optimized penalty parameters. 45
4.6 Model specification of the fitted parametric C-Vine model PCC 2. 46
4.7 Log-likelihoods of the nonparametric and parametric C-Vine models PCC 2. 46
4.8 Summary of cross-validated log-likelihoods of the nonparametric and parametric C-Vine models. 51

5.1 Overview of simulated and fitted pair-copula constructions. 52
5.2 Simulation scenario settings. 53

Chapter 1

Introduction

Starting with the work of Sklar [1959], statistical dependence models for high-dimensional data by means of copulas have extensively been analyzed. For a long time, parametric copula families have been in the focus. Nonparametric methods were not popular due to, among others, the curse of dimensionality and comparatively high computational effort. The recently introduced concept of pair-copula constructions of Joe [1996] helps to overcome the curse of dimensionality, such that nonparametric copula modeling became more famous. Nonparametric copulas are especially useful in the presence of asymmetric bivariate dependence, as parametric copulas do not handle asymmetry properly. Furthermore, with parametric copulas there is a model risk if the family is not chosen properly. However, if a parametric family is chosen correctly, parametric estimation is more efficient than nonparametric estimation.

In general, it is possible to estimate the pair-copula itself nonparametrically and to deduce the pair-copula density by derivation. Alternatively, it is possible to estimate the pair-copula density nonparametrically and to deduce the pair-copula by integration.
Scaillet and Fermanian [2002] investigate kernel estimation of copula functions for time series. Hobaek Haff and Segers [2012] perform estimation of pair-copula constructions with the empirical pair-copula in the style of Deheuvels [1979]. Lopez-Paz et al. [2012] use kernel density estimation to model R-Vines. To avoid boundary bias, the d-dimensional copula data, which is uniformly distributed on the compact unit hypercube $[0,1]^d$, is transformed to normally distributed data on $\mathbb{R}^d$ by applying the univariate standard normal quantile function to each margin. On this level, the bivariate density is estimated and then transformed back to the original scale. A different approach is the usage of Bernstein polynomials to express the pair-copula. This was introduced by Sancetta and Satchell [2004]. The coefficients can be estimated data-driven as probabilities from a contingency table (see Pfeifer et al. [2009]). Diers et al. [2012] follows this approach in the Bernstein copula estimation with application to insurance data. Weiß and Scheffer [2012] also conduct this approach in Vine estimation with application to financial returns data. Schellhase [2012], however, obtains the coefficients from maximum likelihood estimation. He even penalizes the likelihood to influence the coefficients in order to receive a smooth density estimate. Schellhase [2012] shows a further pair-copula density estimation approach using penalized B-splines, based on arguments from the spline smoothing literature. Shen et al. [2008] use

linear B-splines in the pair-copula function estimation.

The goal of this thesis is to review selected concepts of nonparametric estimation of pair-copula constructions. In fact, we focus on the theory of kernel density estimation and Bernstein copulas. We apply these concepts in an empirical study and a simulation study in three dimensions and compare the results to parametric pair-copula constructions. The thesis is outlined as follows. In Chapter 2 we present the theoretical background of kernel density estimation and pair-copula constructions. As a tool for model comparison, we introduce the method of cross-validated log-likelihoods. Nonparametric estimation of bivariate copulas is presented in Chapter 3. We start with the empirical copula defined by indicator functions. We explain the pair-copula estimation by kernel density estimation based on the data, which is transformed by standard normal quantile function in order to avoid boundary bias. Next, we introduce Bernstein copulas and the maximum likelihood estimation of the coefficients. In this context, we also go detailed into the concept of penalization. In Chapter 4 we apply the kernel density estimation approach and the Bernstein copula approach to a data sample and figure out typical behavior of the estimates. We evaluate cross-validated log-likelihoods to get an idea of the out-of-sample prediction quality. We draw a comparison to a parametric pair-copula construction. In Chapter 5, we apply selected nonparametric and parametric pair-copula constructions to simulated data. The simulation study is intended to figure out the flexibility of nonparametric copula estimation over parametric copula estimation in the presence of asymmetric dependence.

Chapter 2

Theoretical Background

2.1 Kernel Density Estimation

The histogram is a basic statistical tool to obtain information about the underlying probability mass function or probability density function of a given sample. It is simply created for univariate data, however, its appearance depends on the bin selection. Since the histogram is not continuous, it is an unsatisfying approximation of continuous probability density functions. Hence, more sophisticated tools are needed. Here, kernel density estimation comes into play.

Kernel density estimation is a basic concept in nonparametric and semi-parametric statistics. It is a flexible method to estimate probability density functions and cumulative distribution functions. The univariate kernel density estimation was proposed by Fix and Hodges [1951]. Cacoullos [1966] and Epanechnikov [1969] extended it to a multivariate setting. Marron and Ruppert [1993] propose a method to reduce boundary bias of kernel density estimators by proper transformation of the data, estimating the density using the transformed data and re-transforming the estimate to the original scale.

In this section, we explain the principle of kernel density estimation for univariate and multivariate probability density functions. From this, we deduce an estimator of the respective cumulative distribution function. Throughout this thesis, we restrict ourselves to the continuous case. As the bandwidth selection is crucial for the smoothness and volatility of the resulting probability density estimate, we discuss this in detail. We introduce the normal reference rule of thumb method, proposed by Silverman [1986], as well as the data-driven methods of least squares cross-validation and maximum likelihood cross-validation.

2.1.1 Univariate Kernel Density Estimation

We introduce univariate kernel density estimation following Silverman [1986] Li and Racine [2007].

For a fixed $n \in \mathbb{N}$, let $x_1, \ldots, x_n$, be observations of a univariate continuous random

variable $X : \Omega \to \mathbb{R}$ with distribution function $F : \mathbb{R} \to [0,1]$ and probability density function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$. We denote the estimates of F and f by $\hat{F}$ and $\hat{f}$, respectively.

The following definition is taken from Zucchini [2003].

Definition 2.1 (General Kernel Function)
Any function $k(x) : \mathbb{R} \to \mathbb{R}$ that satisfies the following conditions is a kernel function:

$$\begin{aligned} \int_{-\infty}^{\infty} k(x)\,dx &= 1 \\ \int_{-\infty}^{\infty} xk(x)\,dx &= 0 \qquad (2.1) \\ \int_{-\infty}^{\infty} x^2 k(x)\,dx &=: b_2 < \infty \qquad (2.2) \end{aligned}$$

In practice, kernel functions are often chosen to be symmetric and nonnegative. Any symmetric probability density function with finite variance is a kernel function. For a kernel function that takes negative values, the estimate itself may take negative values which is not desirable when estimating density functions. Kernel methods with asymmetric kernels can be found in Abadir and Lawford [2004].

Definition 2.1 is very general. We present a further definition taken from Li and Racine [2007], which is more restrictive. From now on, when we speak of kernel functions, we assume them to satisfy the following definition.

Definition 2.2 (Kernel Function)
A nonnegative bounded function $k(x) : \mathbb{R} \to \mathbb{R}_{\geq 0}$ that satisfies

$$\begin{aligned} \int_{-\infty}^{\infty} k(x)\,dx &= 1 \qquad (2.3) \\ k(x) &= k(-x) \qquad (2.4) \\ \int_{-\infty}^{\infty} x^2 k(x)\,dx &=: b_2 > 0 \qquad (2.5) \end{aligned}$$

is a kernel function.

Equation (2.3) together with the nonnegativity of k makes k being a probability density function. Note that if Equation (2.4) holds, then Equation (2.1) holds, too.

In Table 2.1 we define some common kernel functions, namely the second order Gaussian, the second order Epanechnikov and the uniform kernel function. The order of the

Kernel	Definition
Gaussian	$k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}$
Epanechnikov	$k(x) = \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}x^2\right), \quad \lvert x\rvert < \sqrt{5}$
Rectangular or Uniform	$k(x) = \frac{1}{2}, \quad \lvert x\rvert < 1$

Table 2.1: Definition of some famous kernel functions.

first kernel moment that is different from zero is called *order of the kernel*. Formally, $m \in \mathbb{N}$ is the order of k, if

$$\int_{-\infty}^{\infty} x^l k(x)\,dx = 0 \quad \text{for} \quad l = 1, \ldots, m-1 \quad \text{and} \quad \int_{-\infty}^{\infty} x^m k(x)\,dx \neq 0.$$

Gaussian and Epanechnikov kernels also exist in orders higher than two. They can be useful to enhance the goodness of fit, but as Li and Racine [2007] state, the drawback is that the kernel function necessarily takes negative values. Hence, we do not go deeper into this.

Definition 2.3 (Univariate Kernel Density Estimator)
For a probability density function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$, the kernel density estimator is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right),$$

where $x_1, \ldots, x_n$ denote the observations of the random variable X and k is a kernel function as in Definition (2.2). The parameter $h > 0$ is called bandwidth or smoothing parameter.

Sometimes we write $\hat{f}_h$ instead of $\hat{f}$ to emphasize that the estimator depends on the parameter h.

Silverman [1986] points out that if the kernel k is a probability density function, the estimate $\hat{f}$ will be a probability density, too. Furthermore, $\hat{f}$ possesses the same continuity and differentiability properties as k.

2.1.2 Univariate Bandwidth Selection

Obviously, the resulting estimator depends on the choice of the kernel function k and the bandwidth h. They should be chosen in a way such that the distance between the true probability density function f and the estimated probability density function $\hat{f}$ is minimized.

Serfling [1980] showed that in many situations, the second order Epanechnikov kernel is optimal. Many authors emphasize that the bandwidth affects the estimate much more than the kernel function, e. g. see Li and Racine [2007]. We illustrate this in Figure 2.1. We simulate 500 observations from the bimodal probability density function $f(x) = \frac{1}{2} \cdot \phi_{-2;0.5}(x) + \frac{1}{2} \cdot \phi_{2;0.5}(x)$. The function ϕ_{μ,σ^2} denotes the probability density function of the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. We performed several kernel density estimations on that sample and depict the resulting density estimates in Figure 2.1. In each column of the figure, we varied the kernel function and in each row of the figure, we applied different bandwidths. On the one hand, we applied optimal bandwidths from least squares cross-validation, maximum likelihood cross-validation and from the rule of thumb. On the other hand, we determined the bandwidths explicitly to illustrate the effect of undersmoothing (for $h = 0.05$) and oversmoothing (for $h = 1$ and $h = 2.5$). We clearly see the significant effect of bandwidth variation. Kernel variation has only obvious influence in this example, if the bandwidth parameter is far away from its optimal value. In other words, as long as h is chosen properly, the choice of the kernel function is of secondary importance.

The bandwidth h can be fixed or varying. Varying bandwidths are obtained by generalized nearest neighbors method or adaptive nearest neighbors method (cf. Li and Racine [2007]). In this thesis, we concentrate on fixed bandwidths.

Following we define common measures for the distance between a true density f and its estimate $\hat{f}_h$.

The mean squared error (MSE) is defined as

$$MSE\left(\hat{f}_h(x)\right) := E\left[\left(\hat{f}_h(x) - f(x)\right)^2\right] \tag{2.6}$$

and can be rewritten to

$$MSE\left(\hat{f}_h(x)\right) = \left(E\left[\hat{f}_h(x) - f(x)\right]\right)^2 + Var\left[\hat{f}_h(x)\right]. \tag{2.7}$$

An optimal bandwidth h shall minimize the MSE. In Equation (2.7), the first term reflects the squared bias and the second term is the variance of $\hat{f}_h(x)$. There is a tradeoff between squared bias reduction and variance reduction, that is, they cannot be minimized at the same time. Later we introduce cross-validation methods, which are motivated by the need to balance the squared bias and variance of the estimate. The following theorem gives an idea of the explicit form of the MSE, when the number of observations tends to infinity and the bandwidth tends to zero at the same time.

Theorem 2.4
Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with probability density function $f(x)$. Furthermore, let $f(x)$ be three times differentiable and let x be an interior point in the support of $f(x)$. Then, for the kernel density estimator $\hat{f}(x)$ from

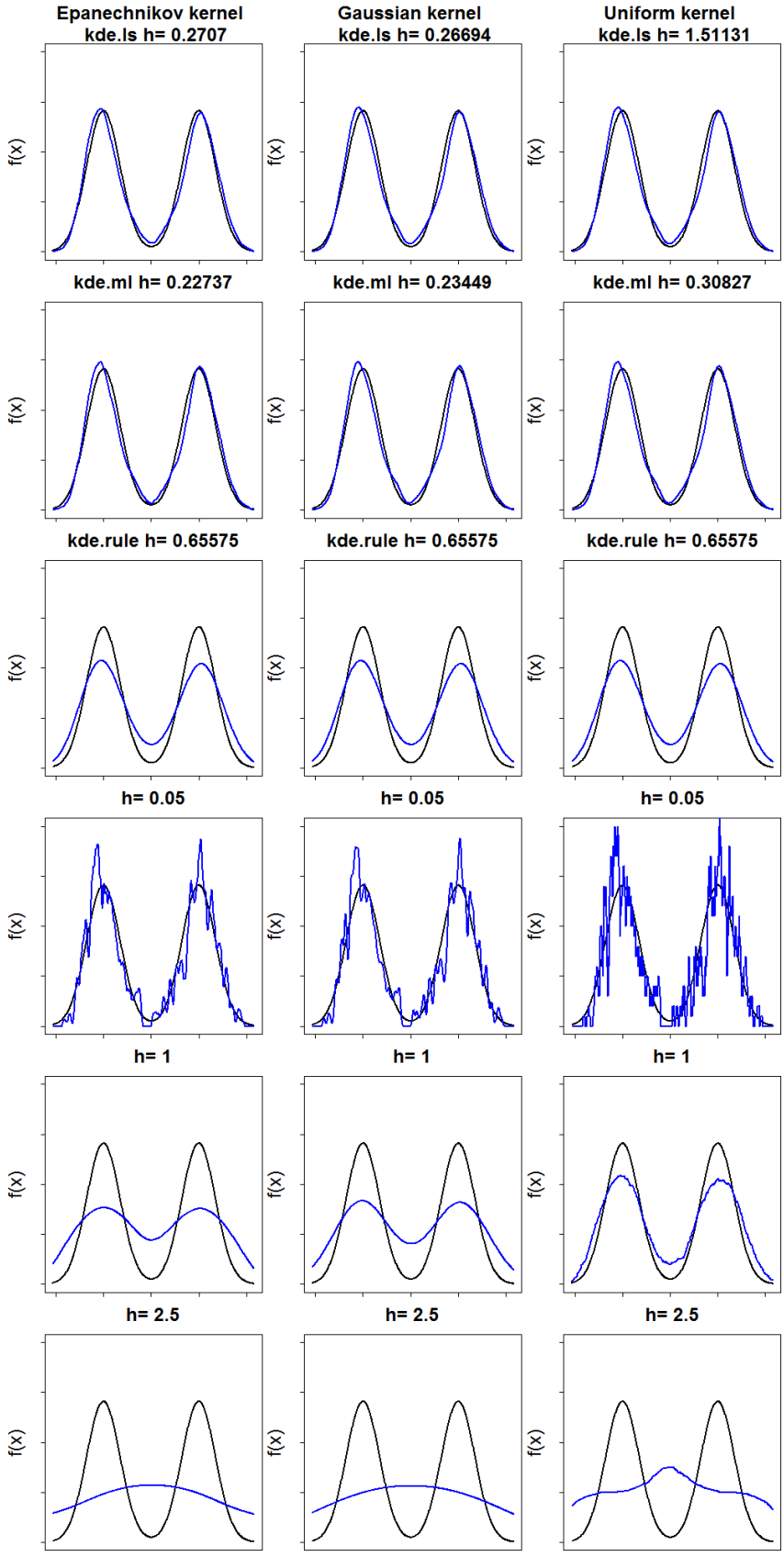

Figure 2.1: Illustration of the effect of bandwidth and kernel selection on the density estimate. The black function is the true probability density function, the blue function is the fitted kernel density estimate for the respective bandwidth and kernel function (bandwidth specification: kde.ls = least-squares cross-validation, kde.ml = maximum likelihood cross-validation, kde.rule = normal reference rule of thumb).

Definition 2.3,

$$MSE\left(\hat{f}_h(x)\right) = \frac{h^4}{4}\left(b_2 \cdot f''(x)\right)^2 + \frac{b \cdot f(x)}{nh} + o\left(h^4 + \frac{1}{nh}\right) \tag{2.8}$$

as $n \to \infty, h \to 0$ *and* $nh \to \infty$*, where* $f''(x)$ *is the second order derivative of* $f(x)$*,* $b_2 := \int v^2 k(v)\, dv$ *and* $b := \int k^2(v)\, dv$*.*

Proof: We refer to Li and Racine [2007].

The MSE measures the distance between the true density and the estimated density in a point x. An advanced distance measure is the integrated mean squared error ($IMSE$). It does not depend on one point x, because it involves all points x in the domain of $\hat{f}_h$. The $IMSE$ is defined as

$$IMSE\left(\hat{f}_h\right) := \int MSE\left(\hat{f}_h(x)\right) dx = \int E\left[\left(\hat{f}_h(x) - f(x)\right)^2\right] dx. \tag{2.9}$$

By Theorem 2.4, one can show that

$$IMSE\left(\hat{f}_h\right) = \frac{h^4 b_2^2}{4}\int \left(f''(x)\right)^2 dx \quad + \frac{b}{nh} \quad + o\left(h^4 + \frac{1}{nh}\right), \tag{2.10}$$

as $n \to \infty, h \to 0$ and $nh \to \infty$, with b and b_2 as in Theorem 2.4.

The optimal bandwidth h_{opt} that minimizes the two leading terms of (2.10) is given by

$$h_{opt} = c_0 n^{-\frac{1}{5}} \tag{2.11}$$

with the positive constant $c_0 = \left[\frac{b_2^2}{b}\int \left(f''(x)\right)^2 dx\right]^{-1/5}$. If $f''(x) = 0$ for almost all x, c_0 is not well-defined.

It can be shown (see Li and Racine [2007]) that $\hat{f}(x)$ converges to $f(x)$ in probability, if k satisfies Definition 2.2. In that sense, $\hat{f}(x)$ is a consistent estimator of $f(x)$. It can also be shown that $\hat{f}(x)$ has an asymptotic normal distribution.

Bandwidth Selection by Rule of Thumb

The rule of thumb method calculates the optimal bandwidth by Equation (2.11). Therefore, we need to calculate the second derivative of the unknown density f. Silverman [1986] suggests to assume a density from a parametric family of distributions. When working with the Gaussian density, it is referred to as the *normal reference* rule of thumb. Then, the second derivative $f''(x)$ is easy to calculate. If additionally one uses the Gaussian kernel estimator, the optimal bandwidth from Equation (2.11) turns out to be

$$h_{opt} = \left(\frac{4}{3n}\right)^{1/5} \cdot \sigma \quad \approx \quad \frac{1.06}{n^5} \cdot \sigma. \tag{2.12}$$

This is proper for distributions that are close to the Normal distribution. It is left to estimate the standard deviation σ. One could consider the sample standard deviation $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, but this is not robust. Hence, Silverman [1986] suggests the estimator

$$\hat{\sigma} = min\left(s, \frac{r}{1.349}\right), \tag{2.13}$$

where r denotes the interquartile range of the data. The constant 1.349 is derived from the fact, that the interquartile range of a normal distribution with variance σ^2 is equal to $1.349 \cdot \sigma$.

Least Squares Cross-Validation

Rudemo [1982] and Bowman [1984] initially proposed the least squares cross-validation method for bandwidth selection. The resulting h is optimal in the sense that it minimizes the integrated squared error (ISE). The following is presented according to Li and Racine [2007].

The integrated squared error is defined as

$$\begin{aligned} ISE\left(\hat{f}\right) &:= \int \left(\hat{f}(x) - f(x)\right)^2 dx \\ &= \int \hat{f}(x)^2\, dx - 2\int \hat{f}(x) f(x)\, dx + \int f(x)^2\, dx. \end{aligned} \tag{2.14}$$

We ignore the last summand of Equation (2.14), as it does not depend on h and obtain the objective function

$$\widetilde{ISE}\left(\hat{f}\right) := \int \hat{f}(x)^2\, dx - 2\int \hat{f}(x) f(x)\, dx. \tag{2.15}$$

Let X denote the continuous random variable corresponding to f and let $\left\{X_j\right\}_{j=1,\ldots,n}$ be the random observations used for the computation of $\hat{f}$. We approximate

$$\int \hat{f}(x) f(x)\, dx \quad = \quad E_X\left[\hat{f}(X)\right] \quad \approx \quad \frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i}(X_i), \tag{2.16}$$

where $E_X[\cdot]$ denotes the expectation with respect to f and

$$\hat{f}_{-i}(X_i) := \frac{1}{(n-1)h}\sum_{j=1, j\neq i}^{n} k\left(\frac{X_i - X_j}{h}\right). \tag{2.17}$$

$\hat{f}_{-i}(X_i)$ is called the *leave-one-out kernel estimator* of $f(X_i)$. Furthermore, we compute

$$\begin{aligned} \int \hat{f}(x)^2\, dx &= \frac{1}{(nh)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int k\left(\frac{X_i - x}{h}\right) k\left(\frac{X_j - x}{h}\right) dx \\ &= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n} \bar{k}\left(\frac{X_i - X_j}{h}\right) \end{aligned} \tag{2.18}$$

with

$$\bar{k}(v) = \int k(u)\, k(v-u)\, du \tag{2.19}$$

being the two-fold convolution kernel of k. Plugging (2.16) and (2.18) into (2.15), we obtain the least squares cross-validation objective function

$$CV_f(h) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{h} \bar{k}\left(\frac{X_i - X_j}{h}\right) - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{h} k\left(\frac{X_i - X_j}{h}\right). \tag{2.20}$$

The optimal bandwidth h_{opt} is chosen to minimize $CV_f(h)$, i. e. the resulting density estimate $\hat{f}_{h_{opt}}$ is optimal with respect to ISE. To this end, numerical search algorithms are applied. In the application part of this thesis, we use the R package `np` (see Hayfield and Racine [2008]) to perform kernel density estimation. The bandwidth estimation routine uses Powell's unconstrained optimization by quadratic approximation (UOBYQA) algorithm as proposed by Powell [2000].

Likelihood Cross-Validation

Bandwidth selection by likelihood cross-validation was proposed by Duin [1976]. With this method, the optimal bandwidth h_{opt} is chosen to maximize the *leave-one-out log-likelihood function* $\mathcal{L}$. Let $X_1, \ldots, X_n$ be independent and identically distributed continuous random variables with probability density function $f(x)$. Then, the leave-one-out log-likelihood function is

$$\mathcal{L}(h; X_1, \ldots, X_n) = \ln L(h; X_1, \ldots, X_n) = \ln\left(\prod_{i=1}^{n} \hat{f}_{-i}(X_i)\right) = \sum_{i=1}^{n} \ln \hat{f}_{-i}(X_i).$$

with $\hat{f}_{-i}(X_i)$ as defined in Equation (2.17). Schuster and Gregory [1981] showed that using compactly supported kernels with likelihood cross-validation bandwidths for estimating infinitely supported probability densities, leads to inconsistent results. Chow et al. [1983] showed that using compactly supported kernels with this bandwidth method for estimating compactly supported probability densities, leads to consistent results. However, Hall [1987] advises not to use compactly supported kernels with likelihood cross-validation bandwidths.

2.1.3 Multivariate Kernel Density Estimation

In this section, we refer to Li and Racine [2007].

For a fixed $n \in \mathbb{N}$, let $\mathbf{x}_1, \ldots, \mathbf{x}_n$, be n observations of a d-dimensional continuous random variable $\mathbf{X} : \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, with distribution function $F : \mathbb{R}^d \to [0,1]$ and probability density function $f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$. That is, each observation $\mathbf{x}_i$ consists of d univariate entries $x_{i1}, \ldots, x_{id}$. We denote the estimates of F and f by $\hat{F}$ and $\hat{f}$, respectively.

Definition 2.5 (Multivariate Kernel Density Estimator)
For a probability density function $f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, the kernel density estimator is defined as

$$\hat{f}(\mathbf{x}) = \frac{1}{nh_1 \ldots h_d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right), \tag{2.21}$$

where

$$K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) := \prod_{j=1}^{d} k\left(\frac{x_{ij} - x_j}{h_j}\right) \tag{2.22}$$

with k being a univariate kernel function as in Definition 2.2 and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ denote the observations of the random variable $\mathbf{X}$. The parameters $h_1, \ldots, h_d > 0$ are called bandwidths.

2.1.4 Multivariate Bandwidth Selection

The optimal bandwidths $h_1, \ldots, h_d$ can be found as a result of the rule of thumb or by cross-validation arguments.

Bandwidth Selection by Rule of Thumb

The optimal bandwidths by the rule of thumb are given by

$$h_j = c_j s_j n^{-\frac{1}{4+d}}, \quad j = 1, \ldots, d,$$

where c_j is a positive constant and s_j denotes the sample standard deviation of $\left\{x_{ij}\right\}_{i=1,\ldots,n}$. The exponent $-\frac{1}{4+d}$ results from our assumption that all considered random variables are continuous and that we use kernel functions of order two. The general form of the exponent is

$$-\frac{1}{2 \cdot (\text{kernel order}) + (\text{number of continuous variables})}.$$

Often, the constants $c_j, j = 1, \ldots, d$ are all set to 1.06, which comes from the univariate normal reference rule of thumb $\left(1.06 = \left(\frac{4}{3}\right)^{-\frac{1}{4+1}}\right)$. However, this choice is not advisable when working with other kernels than the Gaussian. Henderson and Parmeter [2011] derive optimal constants for several multivariate kernel functions of different orders. An enhanced approach is to choose $c_j, j = 1, \ldots, d$ not equally, but rather according to the behavior of the density of interest in the respective dimension j. For example, if the density is quite bumpy in dimension 1 and very smooth in dimension 2, then c_1 should be relatively small and c_2 relatively large. However, the rule of thumb method does not provide this quality.

Least Squares Cross-Validation

A straightforward generalization of the univariate least squares cross-validation objective function from Equation (2.20) to d dimensions results in

$$CV_f(h_1, \ldots, h_d) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{n} \overline{K_h}(X_i, X_l) - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{l=1, l \neq i}^{n} K_h(X_i, X_l),$$

where

$$K_h(X_i, X_l) = \prod_{j=1}^{d} \frac{1}{h_j} k\left(\frac{X_{ij} - X_{lj}}{h_j}\right) \quad \text{and}$$

$$\overline{K_h}(X_i, X_l) = \prod_{j=1}^{d} \frac{1}{h_j} \overline{k}\left(\frac{X_{ij} - X_{lj}}{h_j}\right).$$

Again we assume k to satisfy Definition 2.2 and $\overline{k}$ to be its two-fold convolution kernel as defined in Equation (2.19).

Likelihood Cross-Validation

The likelihood cross-validation in one dimension can directly be extended to the multivariate setting. The likelihood function has to be maximized over all bandwidths. Again we face the problem of inconsistent results with the usage of compactly supported kernels when the underlying distribution is infinitely supported.

2.1.5 Estimating Distribution Functions

So far, we obtained a kernel estimator for the multivariate probability density function $f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ as $\hat{f}(\mathbf{x}) = \frac{1}{nh_1 \ldots h_d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)$ (recall Equation (2.21)). Now we are interested in estimating the corresponding cumulative distribution function $F : \mathbb{R}^d \to [0, 1]$. We do this by replacing the true density function f by $\hat{f}$ in the basic equation $F(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} f(\mathbf{v}) d\mathbf{v}$. This yields

$$\hat{F}(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} \hat{f}(\mathbf{v}) d\mathbf{v} \tag{2.23}$$

$$\overset{(2.21)}{=} \int_{-\infty}^{\mathbf{x}} \frac{1}{nh_1 \ldots h_d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{v}}{h}\right) d\mathbf{v} \tag{2.24}$$

$$\overset{(2.22)}{=} \frac{1}{nh_1 \ldots h_d} \sum_{i=1}^{n} \prod_{j=1}^{d} \int_{-\infty}^{x_j} k\left(\frac{x_{ij} - v_j}{h_j}\right) dv_j. \tag{2.25}$$

In Equation (2.25) we need to integrate the kernel function k. In Table 2.2 we present some common integrated kernels.

In the next section we introduce copulas. When fitting a copula model to d-dimensional data $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$, the first step is to transform the original data to copula data. Copula data $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is expected to be marginally uniformly distributed on the unit interval. This is achieved by a transformation with the marginal distribution functions F_j, $j = 1, \ldots, d$. Recall that for a random variable X with distribution function F we have

Kernel	Integrated Kernel
Gaussian	$H(x) = \Phi(x), \quad x \in \mathbb{R}$, Φ standard normal distribution function
Epanechnikov	$H(x) = \begin{cases} 0, & x \leq -\sqrt{5}; \\ \frac{3}{4\sqrt{5}}\left(x - \frac{x^3}{15}\right) + \frac{1}{2}, & -\sqrt{5} < x < \sqrt{5}; \\ 1, & x \geq \sqrt{5} \end{cases}$
Rectangular or Uniform	$H(x) = \begin{cases} 0, & x \leq -1; \\ \frac{1}{2}x + \frac{1}{2}, & -1 < x < 1; \\ 1, & x \geq 1 \end{cases}$

Table 2.2: Integrated kernel functions.

$Y := F(X) \sim U(0,1)$. As in practice we do not know the true marginal distribution functions, we estimate them as $\hat{F}_j$, $j = 1, \ldots, d$, and set

$$\hat{u}_{ij} := \hat{F}_j\left(x_{ij}\right), \quad i = 1, \ldots, n \quad \text{for all } j.$$

We illustrate this using data from a univariate variable, i. e. for $d = 1$ margin. From the data used later on in the empirical study, we take the $n = 655$ observations of the variable *Co*. We estimate its distribution function $\hat{F}_1$ by kernel density estimation with Epanechnikov kernel and bandwidth from least squares cross-validation. We estimate the optimal bandwidth as $h = 0.0356$ with the function `npudensbw` from the `R` package `np`. This gives

$$\begin{aligned}
\hat{F}_1(x) &\overset{(2.25)}{=} \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x} k\left(\frac{x_{i1} - v}{h}\right) dv \\
&\overset{(2.4)}{=} \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{x} k\left(\underbrace{\frac{v - x_{i1}}{h}}_{=:g(v)}\right)\underbrace{\frac{1}{h}}_{=g'(v)} dv \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{x-x_{i1}}{h}} k(z)\, dz \\
&= \frac{1}{n}\sum_{i=1}^{n} H(z)\Big|_{z=-\infty}^{z=\frac{x-x_{i1}}{h}} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(H\left(\frac{x - x_{i1}}{h}\right) - 0\right) \\
&= \frac{1}{655}\sum_{i=1}^{655} H\left(\frac{x - x_{i1}}{0.0356}\right),
\end{aligned} \tag{2.26}$$

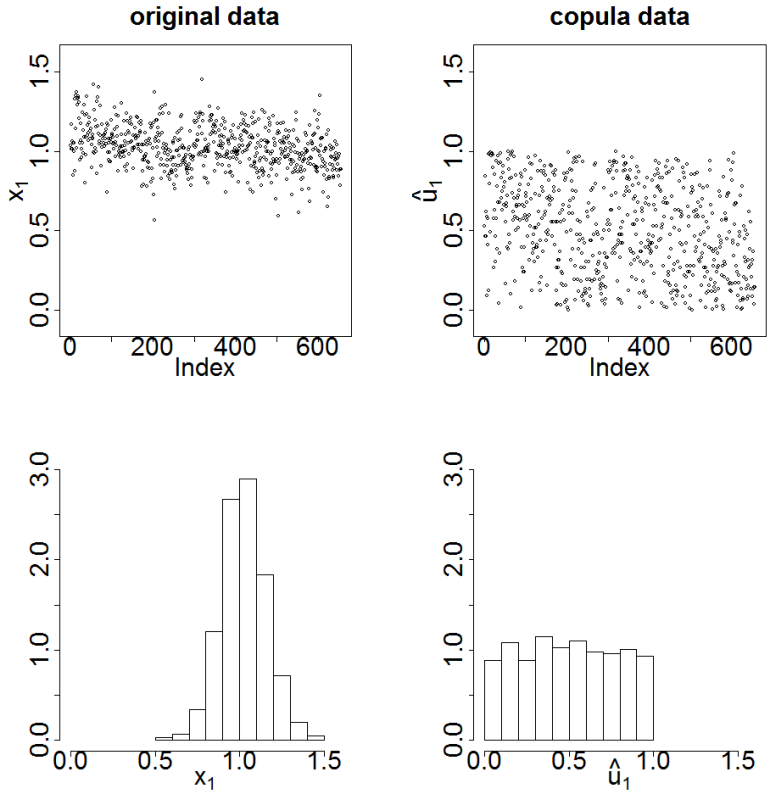

Figure 2.2: Comparison of original and transformed copula data. On the left hand side, we see plots of the original data. On the right hand side, we see the data after it is transformed by a fitted distribution function.

where $x_{i1}, i = 1, \ldots, 655$ are the observations of the variable Co and H is the integrated Epanechnikov kernel as in Table 2.2. We obtain the copula data as

$$\hat{u}_{l1} := \hat{F}_1(x_{l1}) = \frac{1}{655} \sum_{i=1}^{655} H\left(\frac{x_{l1} - x_{i1}}{0.0356}\right), \quad l = 1, \ldots, n.$$

In Figure 2.2 we see scatter plots and histograms of the original data and the transformed data. In Figure 2.3 the estimated distribution function $\hat{F}_1$ is displayed together with its probability density $\hat{f}_1$.

2.2 Dependence

Copulas are a powerful method to model dependencies within multi-dimensional data. Sklar's Theorem guarantees the existence and uniqueness of copulas for continuous distribution functions. Multivariate copulas can be decomposed to a combination of bivariate (conditional) copulas. This resulting decomposition is called pair-copula construction, and was proposed by Joe [1996] and further developed by Bedford and Cooke [2001] and

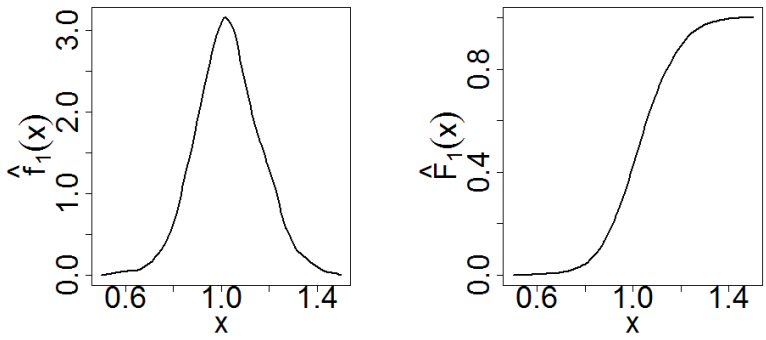

Figure 2.3: Nonparametrically estimated distribution function that is used for data transformation and its probability density.

Kurowicka and Cooke [2006].
The possibility of pair-copula constructions is very valuable. Firstly, it allows to model dependencies separately from the marginal distributions. Secondly, the so-called curse of dimensionality is avoided, because the multi-dimensional copula is decomposed into a bunch of bivariate building blocks. In general, one can choose parametric, semi-parametric or nonparametric pair-copulas as building blocks.
There are several structures of a pair-copula construction available. Bedford and Cooke [2001] and Bedford and Cooke [2002] introduced a graphical organisation of the pair-copula construction by means of trees, called *Vine*.
In Section 2.2.1 we give the definition and basic properties of a copula function as well as the fundamental Theorem of Sklar [1959]. In Section 2.2.2, we introduce C-Vines. We go on defining several dependence measures.

2.2.1 Copulas

Following, we refer to Nelsen [2006].

Definition 2.6 (Copula)
A function $C : [0,1]^d \to [0,1]$ *is called* d*-dimensional copula, if the following properties are satisfied:*

(i) For all $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$, $\quad C(\mathbf{u}) = 0$ *if at least one coordinate of* $\mathbf{u}$ *equals* 0.

(ii) For all $j = 1, \ldots, d$, $\quad C\left(1, \ldots, 1, u_j, 1, \ldots, 1\right) = u_j$.

(iii) C *is* d*-increasing (see Nelsen [2006]).*

The concept of copulas was developed by Sklar [1959] and is based on Sklar's Theorem. Basically it says that for each multivariate distribution function, there exists a copula function that links the joint distribution function to its univariate marginal distribution

functions.

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a vector of univariate random variables $X_1, \ldots, X_d$ with joint probability density function f and cumulative distribution function F. Further, let $f_1, \ldots, f_d$ be the corresponding marginal density functions and let $F_1, \ldots, F_d$ the corresponding marginal distribution functions be strictly increasing and continuous.

Theorem 2.7 (Sklar's Theorem)
Let F be a d-dimensional distribution function with margins $F_1, \ldots, F_d$. Then there exists a d-dimensional copula C such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{2.27}$$

for all $\mathbf{x} = (x1, \ldots, xd) \in (\mathbb{R} \cup \{-\infty, \infty\})^d$.
If all $F_j, j = 1, \ldots, d$ are continuous, then C is unique. On the other hand, if C is a d-dimensional copula and $F_1, \ldots, F_d$ are distribution functions, then F as defined in Equation (2.27) is a d-dimensional distribution function with margins $F_1, \ldots, F_d$.

The copula function can be interpreted as distribution function of d uniformly distributed univariate variables. Recall that if X_j has a distribution function F_j, then $F_j\left(X_j\right) \sim U(0,1)$.

As $F_1, \ldots, F_d$ are strictly increasing and continuous, they are invertible and we obtain the copula by inversion of Equation (2.27) as

$$C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right). \tag{2.28}$$

The copula density $c : [0,1]^d \to \mathbb{R}_{\geq 0}$ can be obtained by partial derivation.

$$f(\mathbf{x}) = \frac{\partial^d C(F_1(x_1), \ldots, F_d(x_d))}{\partial x_1 \ldots \partial x_d} = \frac{\partial^d C(F_1(x_1), \ldots, F_d(x_d))}{\partial F_1(x_1) \ldots \partial F_d(x_d)} f_1(x_1) \ldots f_d(x_d)$$
$$\Rightarrow c(F_1(x_1), \ldots, F_d(x_d)) = \frac{\partial^d C(F_1(x_1), \ldots, F_d(x_d))}{\partial F_1(x_1) \ldots \partial F_d(x_d)} = \frac{f(\mathbf{x})}{f_1(x_1) \ldots f_d(x_d)}$$

A copula $C : [0,1]^d \to [0,1]$ can be used to express multivariate dependence between d variables $X_1, \ldots, X_d, d \in \mathbb{N}$. Within pair-copula constructions, the d-dimensional copula can be expressed by several bivariate copulas. This is why we restrict ourselves to bivariate copulas without limiting the application to $d = 2$.

Definition 2.8 (Bivariate Copula)
A function $C : [0,1]^2 \to [0,1]$ is called bivariate copula, if the following properties are satisfied:

(i) $C(u,0) = 0$ *for all* $u \in [0,1]$ *and* $C(0,v) = 0$ *for all* $v \in [0,1]$

(ii) $C(u,1) = u$ *for all* $u \in [0,1]$ *and* $C(1,v) = v$ *for all* $v \in [0,1]$

(iii) C *is 2-increasing, i. e.* $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ *for all* $u_1, u_2, v_1, v_2 \in [0,1]$ *with* $u_1 \leq u_2$ *and* $v_1 \leq v_2$.

Copulas are invariant under strictly increasing transformations of the margins, as shown in Genest and Favre [2007], whereas the linear correlation coefficient is invariant under strictly increasing linear transformations only. One can work with ranks R_{ij} of the observations $x_{ij}, i = 1, \ldots, n$, of the j-th variable $X_j, j = 1, \ldots, d$, instead of the sample itself. The advantage is that ranks are invariant under strictly increasing transformations. Genest and Favre [2007] illustrate this fact in two dimensions simply with scatter plots of the ranks of the original sample and the ranks of the transformed sample.

2.2.2 Pair-Copula Constructions

From Sklar's Theorem, we can deduce the d-dimensional probability density function as

$$f(x_1, \ldots, x_d) = \left(\prod_{j=1}^{d} f_j\left(x_j\right) \right) c\left(F_1(x_1), \ldots, F_d(x_d)\right) \tag{2.29}$$

With pair-copula constructions one can express the joint copula density c as a product of several bivariate pair-copulas. As commonly used, we assume for simplification that the copula density does not depend on the value of the conditioning variables. Formally, we write

$$\begin{aligned} c_{jk|D}\left(F\left(x_j|x_D\right), F\left(x_k|x_D\right)|x_D\right) &= c_{jk|D}\left(F\left(x_j|x_D\right), F\left(x_k|x_D\right)\right) \\ &=: c_{jk;D}\left(F\left(x_j|x_D\right), F\left(x_k|x_D\right)\right), \end{aligned}$$

where $D \subset \{1, \ldots, d\}$ is a subset of indices, $j, k \in \{1, \ldots, d\} \setminus D, j \neq k$ and $x_D := (x_l : l \in D)$. We demonstrate the concept of pair-copula constructions following Aas et al. [2009].

Let $\mathbf{X} = (X_1, \ldots, X_d)$ have the joint distribution function F, marginal distribution functions $F_1, \ldots, F_d$ and corresponding density functions f and $f_1, \ldots, f_d$. By recursive conditioning,

$$f(x_1, \ldots, x_d) = f_1(x_1) f(x_2|x_1) f(x_3|x_2, x_1) \ldots f(x_d|x_{d-1}, \ldots, x_1). \tag{2.30}$$

Then, the conditional densities in Equation (2.30) can be expressed with the help of pair-copula densities as

$$\begin{aligned} f(x_2|x_1) &= \frac{f(x_1, x_2)}{f_1(x_1)} = \frac{c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_1(x_1) \cdot f_2(x_2)}{f_1(x_1)} \\ &= c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_2(x_2) \end{aligned}$$

$$\begin{aligned} f(x_3|x_2, x_1) &= \frac{f(x_2, x_3|x_1)}{f(x_2|x_1)} = \frac{c_{23;1}\left(F(x_2|x_1), F(x_3|x_1)\right) \cdot f(x_2|x_1) \cdot f(x_3|x_1)}{f(x_2|x_1)} \\ &= c_{23;1}\left(F(x_2|x_1), F(x_3|x_1)\right) \cdot f(x_3|x_1) \\ &= c_{23;1}\left(F(x_2|x_1), F(x_3|x_1)\right) \cdot c_{13}\left(F_1(x_1), F_3(x_3)\right) \cdot f_3(x_3) \end{aligned}$$

and so on,

where $c_{12}(\cdot,\cdot)$ is the pair-copula density for the pair of variables $F_1(X_1)$ and $F_2(X_2)$ and $c_{23;1}(\cdot,\cdot)$ is the pair-copula density for the pair of variables $F(x_2|x_1)$ and $F(x_3|x_1)$.

In general,

$$f\left(x_j|\mathbf{x}_D\right) = c_{jv;D_{-v}}\left(F\left(x_j|\mathbf{x}_{D_{-v}}\right), F\left(x_v|\mathbf{x}_{D_{-v}}\right)\right) \cdot f\left(x_j|\mathbf{x}_{D_{-v}}\right), \quad j = 1,\ldots,d,$$

where $\mathbf{x}_D$ is a subvector of $\mathbf{x}$ with elements from D as indices, $D \subset \{1,\ldots,d\} \setminus \{j\}$ and $D_{-v} := D \setminus \{v\}$. For different choices of x_v the representation of $f\left(x_j|\mathbf{x}_D\right)$ results in a different decomposition.
Finally, the joint density function f is decomposed into a product of its marginal densities and bivariate pair-copulas.

The input arguments of $c_{jv;D}$ require the calculation of conditional marginal distribution functions, e. g. $F\left(x_j|\mathbf{x}_D\right)$. For this, we use the following recursive identity, introduced by Joe [1996],

$$F\left(x_j|\mathbf{x}_D\right) = \frac{\partial C_{jv;D}\left(F\left(x_j|\mathbf{x}_{D_{-v}}\right), F\left(x_v|\mathbf{x}_{D_{-v}}\right)\right)}{\partial F\left(x_v|\mathbf{x}_{D_{-v}}\right)}.$$

We have seen that the pair-copula construction of a d-dimensional probability density is not unique. The number of possible decompositions rather increases significantly with increasing dimension d. For example, Aas et al. [2009] state that there are 240 different decompositions for a five-dimensional density. Bedford and Cooke [2001] and Bedford and Cooke [2002] introduced a graphical organisation of the pair-copula construction, called *Vines*. There are Regular Vines (R-Vines), Canonical Vines (C-Vines) and Drawable Vines (D-Vines). For details, we refer to Kurowicka and Cooke [2006] and Kurowicka and Joe [2011]. Following we concentrate on C-Vines.

For a pair-copula construction in d dimensions, the vine consists of $d-1$ trees. With tree, we mean an acyclic connected graph with nodes and edges. The nodes in the first tree represent the variables $X_1,\ldots,X_d$, in the second tree the transformed variables $F(X_l|X_{r_1}), l = 1,\ldots,d, l \neq r_1$, in the third tree the transformed variables $F(X_l|X_{r_2},X_{r_1}), l = 1,\ldots,d, l \neq r_1, r_2$, and so on. An edge between two nodes represents the bivariate pair-copula, that enters the model. C-Vines are characterized by the fact, that in each tree there exists one root node which is connected to all other nodes. That is, if X_{r_1} is the root node in the first tree, this implies that, among others, the pair-copula densities $c_{1r_1}, c_{2r_1},\ldots,c_{(r_1-1)r_1}, c_{(r_1+1)r_1},\ldots,c_{dr_1}$ enter the model. Then, conditional distribution functions conditioned on the root node $F(X_l|X_{r_1}), l = 1,\ldots,d, l \neq r_1$ are calculated. These serve as variables, i. e. nodes, in the next tree and a new root node r_2 is determined. This implies, that the pair-copula densities $c_{1r_2;r_1},, c_{2r_2;r_1},\ldots,c_{(r_2-1)r_2;r_1}, c_{(r_2+1)r_2;r_1},\ldots,c_{dr_2;r_1}$ enter the model. Repeating these steps, the conditioning is on all previous root nodes. This results in a star structure of the C-Vine, see Figure 2.4 for an example. Assuming the root nodes $1\ldots,d-2$, the C-Vine density is given by

$$f(\mathbf{x}) = \prod_{r=1}^{d} f_r(x_r) \prod_{i=1}^{d-1} \prod_{j=i+1}^{d} c_{ij;1,\ldots,i-1}\left(F\left(x_i|x_1,\ldots,x_{i-1}\right), F\left(x_j|x_1,\ldots,x_{i-1}\right)\right). \tag{2.31}$$

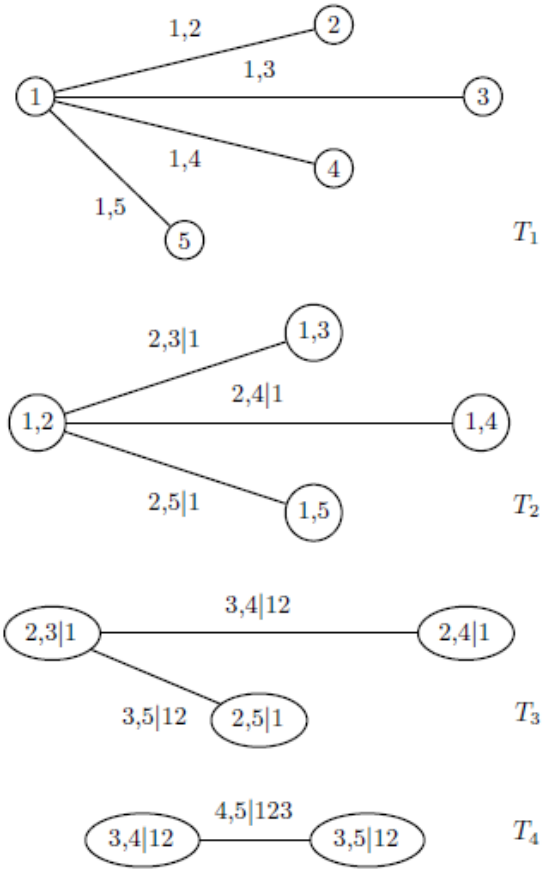

Figure 2.4: Example of five-dimensional C-Vine.

For a method of selecting the root nodes, we follow the idea of Dissmann et al. [2013], who present a sequential selection of R-Vine trees as maximum spanning trees. We only deal with C-Vines which is much more simple. We choose that variable as root, such that the sum of absolute values of empirical Kendall's τ of all resulting pairs are maximized for that tree. The higher $|\tau_n|$ is for a bivariate sample, the stronger the dependence is assumed between the underlying random variables. The idea is that the first tree shall contain the pair-copulas with the strongest dependence.

2.2.3 Dependence Measures

A basic measure of dependence is the Pearson correlation coefficient, which measures the linear dependence between two random variables X_1 and X_2. In the case $\text{var}\left(X_j\right) \neq 0$ and $\text{var}\left(X_j\right) < \infty$, $j = 1, 2$,

$$\text{corr}\left(X_1, X_2\right) = \frac{\text{cov}\left(X_1, X_2\right)}{\sqrt{\text{var}\left(X_1\right)}\sqrt{\text{var}\left(X_2\right)}}.$$

We say that X_1 and X_2 are perfectly correlated, if $|\text{corr}\left(X_1, X_2\right)| = 1$. Then X_2 is a linear function of X_1 or vice versa. Uncorrelatedness $\text{corr}\left(X_1, X_2\right) \neq 0$ has two meanings: either, X_1, X_2 are independent or X_1, X_2 are dependent, but not linearly. This shows the limited power of the linear correlation coefficient. By means of copulas, one can express much more sophisticated dependence structures. We refer to Nelsen [2006].

There are two famous dependence measures based on ranks: Spearman's ρ and Kendall's τ. They inherit the favorable property of invariance under monotone transformations, since they are based on ranks.

Definition 2.9 (Spearman's ρ - sample version)
Let $(x_{i1}, x_{i2}), i = 1, \ldots, n$ be a bivariate sample from the random variable (X_1, X_2) and let $R_{i1}, R_{i2}, i = 1, \ldots, n$, be the ranks of x_{i1}, x_{i2}. Then, the empirical Spearman's ρ is defined as

$$\rho_n = \frac{\sum_{i=1}^{n}\left(R_{i1} - \overline{R_1}\right)\left(R_{i2} - \overline{R_2}\right)}{\sqrt{\sum_{i=1}^{n}\left(R_{i1} - \overline{R_1}\right)^2 \sum_{i=1}^{n}\left(R_{i2} - \overline{R_2}\right)^2}},$$

where $\overline{R_j} = \frac{1}{n}\sum_{i=1}^{n} R_{ij} = \frac{n+1}{2}, \quad j = 1, 2.$

Empirical Spearman's ρ can be interpreted as the empirical Pearson correlation of the ranks, i. e. $\rho_n = corr_n\left(R_1, R_2\right)$ for R_1, R_2 rank variables of X_1, X_2.

Definition 2.10 (Kendall's τ - sample version)
Let $(x_{i1}, x_{i2}), i = 1, \ldots, n$ be a bivariate sample from the random variable (X_1, X_2). Let P_n denote the number of concordant pairs and Q_n the number of discordant pairs. Then Kendall's τ is defined as

$$\tau_n = \frac{P_n - Q_n}{\binom{n}{2}}.$$

Proposition 2.11 (Spearman's ρ)
Let X_1, X_2 be random variables with distribution functions F_1, F_2, joint distribution function F and copula C, such that $F = C(F_1, F_2)$. Then, Spearman's ρ can be expressed as

$$\rho(X_1, X_2) = 12 \int_0^1 \int_0^1 u_1 u_2 \; dC(u_1, u_2) - 3.$$

Proposition 2.12 (Kendall's τ)
Let X_1, X_2 be random variables with distribution functions F_1, F_2, joint distribution function F and copula C, such that $F = C(F_1, F_2)$. Then, Kendall's τ can be expressed as

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) \; dC(u_1, u_2) - 1.$$

The above expressions make clear that ρ and τ only depend on the copula C. The higher $|\rho(X_1, X_2)|$ and $|\tau(X_1, X_2)|$ are, the more dependent X_1, X_2 are.

2.3 Model Comparison using Cross-Validation

We present the method of log-likelihood cross-validation with repeated random subsampling, as we use it in the empirical study to compare the fitted models. We refer to Smyth [2000].

Let $\mathbf{X}$ be a d-dimensional random vector with probability density function f. Let $N := \{\mathbf{x}_i, i = 1, \ldots, n\}$ be a set of n independent observations of $\mathbf{X}$. We divide that available data into two complementary data sets $N^{test} \subsetneq N$ and $N^{train} := N \setminus N^{test}$. N^{train} contains n^{train} data points and N^{test} contains n^{test} data points.
We estimate $\hat{f}$ using the training data N^{train}. From this, we obtain the fitted model parameters, let us call them $\hat{\boldsymbol{\Theta}}^{train}$. Then, we evaluate the log-likelihood of the model on the test data N^{test}, also called validation set, using the fitted model parameters $\hat{\boldsymbol{\Theta}}^{train}$. The test log-likelihood L^{test} is also referred to as *log predictive score*. It gives us information about the out-of-sample prediction quality of the model.

$$L^{test} := L\left(\hat{\boldsymbol{\Theta}}^{train}; N^{test}\right) := \sum_{\mathbf{x} \in N^{test}} \ln \hat{f}\left(\mathbf{x}; \hat{\boldsymbol{\Theta}}^{train}\right). \tag{2.32}$$

For each model of interest we estimate L^{test} as described above with the same training data and test data. The model with the highest test log-likelihood is interpreted as the model with the highest prediction quality. This can be motivated by connecting it to the Kullback-Leibler distance. The Kullback-Leibler distance between two continuous probability densities f and g is defined as

$$KL(f, g) := \int_{-\infty}^{\infty} f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x} = E\left[\ln \frac{f(\mathbf{X})}{g(\mathbf{X})}\right],$$

where $E[\cdot]$ denotes the expectation with respect to f. It measures the discrepancy between the two probability densities. It is zero if the two probability densities are equal and nonzero if they are different.
Smyth [2000] shows that the expected value of the negative test log-likelihood per sample is equal to the Kullback-Leibler distance between the fitted and the true density plus some constant, which is independent of the model.

One single test log-likelihood per model may be not significant for model comparison, as it depends on several factors. For example, it depends concretely on the data N^{train}, N^{test} and the robustness of the parameter estimation. Hence, we are interested in the cross-validated estimate L^{cv} of L^{test}. We repeat the data partitioning, model fitting and likelihood evaluation T times for a fixed $T \in \mathbb{N}$, and receive T test log-likelihoods as in Equation (2.32) as

$$L^{test,t} := L\left(\hat{\mathbf{\Theta}}^{train,t}, N^{test,t}\right), \quad t = 1, \ldots, T.$$

The parameters $\hat{\mathbf{\Theta}}^{train,t}$ are obtained from the t-th parameter fitting using the training data $N^{train,t}$ from the t-th data partitioning. Then, the cross-validated estimate is defined as the arithmetic mean

$$L^{cv} := \frac{1}{T} \sum_{t=1}^{T} L^{test,t}.$$

The model with the highest cross-validated log-likelihood is interpreted as the model with the highest prediction quality.

Smyth [2000] states that T between 20 and 50 is appropriate in most problems. There is no general rule how to choose n_{test}, but Smyth [2000] showed that the choice $n^{test} = 0.5 \cdot n$ is reasonable in many applications. Cross-validation is computationally expensive, because T parameter fittings and evaluations for each model of interest have to be performed.

Chapter 3

Nonparametric Pair-Copula Constructions

So far we learned the basic theoretical concept of pair-copula constructions. Now we present an assortment of nonparametric estimation methods for pair-copulas and pair-copula densities. The drawback of parametric copula families is that they provide a certain degree of symmetry or are restricted to certain correlation structures, which may not be desirable. Nonparametric copula approaches are not bound to these limitations and thus can provide a more adequate estimate of the underlying dependence structure. Following we present pair-copula constructions with the empirical copula, kernel density estimation and with Bernstein polynomials, both unpenalized and penalized.
There are further nonparametric estimation methods of copulas and pair-copula constructions, for example by means of Wavelets (Genest et al. [2009]) and B-splines (Shen et al. [2008], Schellhase [2012]), just to name some. However, we do not treat these, as they go beyond the scope of this thesis.

Recall Equation (2.29), where we express the d-dimensional probability density f as a product of the d marginals $f_j, j = 1, \dots, d$ and the d-dimensional copula density c. Our goal is to estimate the copula density c. Hence we assume the marginals $f_j, j = 1, \dots, d$ to be known or being estimated separately. This is a conventional approach to separate univariate marginal density estimation from copula density estimation. We estimate the copula density as pair-copula construction. Hence, we restrict our considerations to the bivariate case.

3.1 Estimation by Empirical Copulas

The simplest nonparametric copula is the empirical copula, which was initially defined by Deheuvels [1979]. We give the definition in two dimensions.

Let $(x_{i1}, x_{i2}), i = 1, \dots, n$, be n observations of a bivariate random vector (X_1, X_2) with joint distribution function $F : \mathbb{R}^2 \to [0, 1], \quad (x_1, x_2) \to F(x_1, x_2)$ and strictly increasing and continuous marginal distribution functions $F_1(x_1)$ and $F(x_2)$. Sklar's Theorem in

Equation (2.28) implies that

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right) \quad \text{for } u_1, u_2 \in [0,1].$$

We define the empirical estimators of F, F_1 and F_2 by means of indicator functions as

$$\begin{aligned}
\hat{F}(x_1, x_2) &:= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_{(i)1}\leq x_1}\mathbf{1}_{x_{(i)2}\leq x_2} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_{(i)1}\leq x_1, x_{(i)2}\leq x_2} \\
\hat{F}_1(x_1) &:= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_{(i)1}\leq x_1} \\
\hat{F}_2(x_2) &:= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_{(i)2}\leq x_2}
\end{aligned}$$

for $x_1, x_2 \in \mathbb{R}$ and $x_{(i)1}, x_{(i)2}$ the i-th order statistics of the samples $x_{i1}, x_{i2}, i = 1, \ldots, n$. The inverted marginal distribution functions F_1^{-1} and F_2^{-1} are estimated by the pseudo-inverse functions

$$\begin{aligned}
\hat{F}_1^{\leftarrow}(u_1) &:= \inf\left\{t \in \mathbb{R} : \hat{F}_1(t) \geq u_1\right\}, u_1 \in [0,1] \\
\hat{F}_2^{\leftarrow}(u_2) &:= \inf\left\{t \in \mathbb{R} : \hat{F}_2(t) \geq u_2\right\}, u_2 \in [0,1].
\end{aligned}$$

Following Deheuvels [1979], the empirical copula function is defined as

$$C_n(u_1, u_2) = \hat{F}\left(\hat{F}_1^{\leftarrow}(u_1), \hat{F}_2^{\leftarrow}(u_2)\right) \quad \text{for } u_1, u_2 \in [0,1].$$

We tend to the approach given in Hobaek Haff and Segers [2012], hence we modify the factor in the definition of the empirical marginal distribution functions to $\frac{1}{n+1}$ in order to obtain $\hat{F}_j\left(x_j\right) \in [0,1)$ for all $x_j \in \mathbb{R}$ for $j = 1, 2$. Hence, we work with the estimators

$$\begin{aligned}
\hat{F}_1(x_1) &:= \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}_{x_{(i)1}\leq x_1} \\
\hat{F}_2(x_2) &:= \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}_{x_{(i)2}\leq x_2}.
\end{aligned}$$

Setting

$$\hat{u}_{i1} := \hat{F}_1(x_{i1}) \quad \text{and} \quad \hat{u}_{i2} := \hat{F}_2(x_{i2}) \quad \text{for} \quad i = 1, \ldots, n,$$

and denoting the i-th order statistics of $\hat{u}_{i1}, \hat{u}_{i2}$ by $\hat{u}_{(i)1}$ and $\hat{u}_{(i)2}, i = 1, \ldots, n$, we estimate the bivariate copula function as

$$\hat{C}(u_1, u_2) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{u_{(i)1}\leq u_1, u_{(i)2}\leq u_2},$$

We bring this into the context of d-dimensional pair-copula constructions.

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{id}), i = 1, \ldots, n$, be a random sample from a distribution with distribution function $F : \mathbb{R}^d \to [0,1]$ and probability density $f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ that admits a pair-copula construction as in Equation (2.31) with root nodes $1, \ldots, d-2$. We estimate the corresponding pair-copulas following the presentation in Hobaek Haff and Segers [2012].

First, we compute the normalized ranks

$$\hat{u}_{ij} = \frac{1}{n+1} \sum_{t=1}^{n} \mathbf{1}_{x_{(t)j} \leq x_{(i)j}}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, d.$$

Then, the empirical pair-copula for all pairs connected to the first root 1 are estimated as

$$\hat{C}_{1j}\left(u_1, u_j\right) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\hat{u}_{(t)1} \leq u_1, \hat{u}_{(t)j} \leq u_j}, \quad j \in \{2, \ldots, d\}$$

The fitted conditional distribution function is obtained by finite differencing. With the definition $\hat{C}_{1j}(A) := \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{(\hat{u}_{(t)1}, \hat{u}_{(t)j}) \in A}$ for $A \subset \mathbb{R}^2$ and for $h > 0$,

$$\begin{aligned} \hat{C}_{1j}^{(2)}\left(u_1, u_j\right) : &= \frac{\hat{C}_{1j}\left([0, u_1] \times \left[u_j - h, u_j + h\right]\right)}{\hat{C}_{1j}\left([0,1] \times \left[u_j - h, u_j + h\right]\right)} \\ &= \frac{\sum_{t=1}^{n} \mathbf{1}_{\hat{u}_{(t)1} \leq u_1, \left|\hat{u}_{(t)j} - u_j\right| \leq h}}{\sum_{t=1}^{n} \mathbf{1}_{\left|\hat{u}_{(t)j} - u_j\right| \leq h}} \\ \hat{F}_{j|1}\left(x_j | x_1\right) &= \hat{C}^{(2)}\left(\hat{u}_j, \hat{u}_1\right), \quad j \in \{2, \ldots, d\} \end{aligned} \tag{3.1}$$

Then, we compute the normalized ranks of the estimated conditional probabilities as

$$\hat{u}_{i,j|1,\ldots,k} = \frac{1}{n+1} \sum_{t=1}^{n} \mathbf{1}_{\hat{F}_{j|1,\ldots,k}\left(x_{tj} | x_{t1}, \ldots, x_{tk}\right) \leq \hat{F}_{j|1,\ldots,k}\left(x_{ij} | x_{i1}, \ldots, x_{ik}\right)},$$

where $k \in 1, \ldots, d$, $j \in \{k+1, \ldots, d\}$, $i \in 1, \ldots, n$.

The choice of the bandwidth parameter $h > 0$ in Equation (3.2) is very crucial. Hobaek Haff and Segers [2012] suggest to choose $h = c \cdot n^{-1/p}$ for some constant $c > 0$ and a reasonable $p \in \mathbb{N}$. In a simulation study, they sample from a Gaussian D-Vine and investigate the empirical pair-copula estimation with the bandwidths $h = n^{-1/5}$ (this is proportional to the normal reference rule of thumb, see Equation (2.12)), $h = n^{-1/4}$ and $h = 0.5 \cdot n^{-1/3}$. They point out that the optimal bandwidth depends on the dependence structure of the pair-copula construction and not on the margins. The simulation study shows that for a low dependence, in fact Spearman's $\rho = 0.2$, all three bandwidths perform well, but for stronger dependencies, $\rho = 0.5$ and $\rho = 0.8$, the bandwidth $h = 0.5 \cdot n^{-1/3}$ is the only suitable choice. They suppose that undersmoothing, i. e. a rather too small h should be preferred. Undersmoothing (see Figure 2.1) results in a reduced bias and a larger variance. They state that this leads to estimated conditional distribution values far from the true conditional distribution values, but the differences will be averaged out in the copula estimator.

However, an empirical copula is nonsmooth due to its step function characterization. We overcome this problem by using smooth approximation functions. Kolbjørnsen and Stien [2008] estimate the density empirically using Gaussian kernel functions rather than indicator functions. This coincides with the kernel density estimation approach we explain in the following section.

3.2 Estimation by Kernel Density Functions

In this section, we follow Lopez-Paz et al. [2012]. They use R-Vines to decompose the multivariate copula density into a bunch of bivariate pair-copula densities. They estimate each pair-copula density function $c_{jk;D}$ indirectly, i. e. after a proper transformation is applied to the data, a bivariate probability density g is estimated and then re-transformed to the pair-copula density c. An example for a proper transformation is the standard normal quantile function. It transforms the bounded support $[0,1]^2$ of $c_{jk;D}$ to the infinite support $\mathbb{R}^2$, i. e. g is infinitely supported. We follow this approach to avoid the problem of boundary bias in the estimation of g. However, Silverman [1986] points out that in general the estimate on the original scale, i. e. the function c, will be less smooth near the boundaries. Furthermore, we observe that it is not guaranteed that the estimated pair-copula density margins will coincide with the theoretical margins.

We consider kernel density estimation of pair-copulas using Epanechnikov kernel functions k_E or Gaussian kernel functions k_G as defined in Table 2.1, i. e.

$$k_E(x) = \frac{3}{4\sqrt{5}}\left(1-\frac{1}{5}x^2\right), \quad |x| < \sqrt{5} \tag{3.2}$$

$$k_G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R} \tag{3.3}$$

We focus on a three-dimensional C-Vine, assuming the variable X_1 as root, i. e.

$$\begin{aligned} f(x_1,x_2,x_3) =& f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3) \\ &\cdot c_{12}\left(F_1(x_1),F_2(x_2)\right)\cdot c_{13}\left(F_1(x_1),F_3(x_3)\right) \\ &\cdot c_{23;1}\left(F_{2|1}(x_2|x_1),F_{3|1}(x_3|x_1)\right). \end{aligned} \tag{3.4}$$

Let $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}), i = 1,\ldots,n$, be observations from a three-dimensional continuous random variable $\mathbf{X} = (X_1, X_2, X_3)$. We estimate the marginal univariate cumulative distribution functions $\hat{F}_j$, as integral of the univariate kernel density estimate $\hat{f}_j, j = 1,2,3$ as described in Equation (2.25). Now we can transform the observations $\mathbf{x}_i$ to so called *pseudo-observations* or *copula data*

$$\hat{\mathbf{u}}_i = (\hat{u}_{i1}, \hat{u}_{i2}, \hat{u}_{i3}) := \left(\hat{F}_1(x_{i1}), \hat{F}_2(x_{i2}), \hat{F}_3(x_{i3})\right), \quad i = 1,\ldots,n.$$

The random variables $U_j := F_j\left(X_j\right), j = 1,2,3$, are uniformly distributed on $[0,1]$. The copula $C_{jk}\left(u_j,u_k\right)$ is the cumulative distribution function of the bivariate random variable

$\left(U_j, U_k\right), j \neq k$, its corresponding copula density is

$$c_{jk}\left(u_j, u_k\right) = \frac{\partial^2 C_{jk}\left(u_j, u_k\right)}{\partial u_j \partial u_k}.$$

Let us assume $j = 1, k = 2$. We want to transform the support of the copula density $c_{12}(u_1, u_2)$, which is $[0,1]^2$, to $\mathbb{R}^2$. To this end, we take the quantile function of a univariate continuous distribution on $\mathbb{R}$ with strictly positive probability density function. We decide to work with the standard normal distribution $N(0,1)$ and denote the probability density function with ϕ, the cumulative distribution function with Φ and the quantile function with Φ^{-1}. Of course, $\phi > 0$. We transform the random variables U_1 and U_2 to

$$Z_1 := \Phi^{-1}(U_1) \quad \text{and} \quad Z_2 := \Phi^{-1}(U_2).$$

Then, $Z_1, Z_2 \sim N(0,1)$ and its bivariate density g_{12} satisfies

$$g_{12}(z_1, z_2) = \phi(z_1)\,\phi(z_2)\,c_{12}(\Phi(z_1), \Phi(z_2)), \quad z_1, z_2 \in \mathbb{R}. \tag{3.5}$$

This follows from the density transformation theorem. The copula density c_{12} is identical to the original one, but the support is now $\mathbb{R}^2$.

We perform a bivariate kernel density estimation for g_{12}, based on the data $\hat{\mathbf{z}}_i = (\hat{z}_{i1}, \hat{z}_{i2})$, $i = 1, \ldots, n$, where $\hat{z}_{ij} := \Phi^{-1}\left(\hat{u}_{ij}\right)$ for $j = 1, 2$ and $i = 1, \ldots, n$. We set

$$\hat{g}_{12}(z_1, z_2) = \frac{1}{n\hat{h}_1\hat{h}_2} \sum_{i=1}^{n} k\left(\frac{\hat{z}_{i1} - z_1}{\hat{h}_1}\right) k\left(\frac{\hat{z}_{i2} - z_2}{\hat{h}_2}\right), \tag{3.6}$$

where k denotes the kernel function as in Definition 2.2 and $\hat{h}_1, \hat{h}_2$ the optimal bandwidths. Plugging Equation (3.6) into Equation (3.5) and reordering gives

$$\begin{aligned} \hat{c}_{12}(u_1, u_2) &= \frac{\hat{g}\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)}{\phi\left(\Phi^{-1}(u_1)\right)\phi\left(\Phi^{-1}(u_2)\right)} \\ &= \frac{\sum_{i=1}^{n} k\left(\frac{\Phi^{-1}(\hat{u}_{i1}) - \Phi^{-1}(u_1)}{\hat{h}_1}\right) k\left(\frac{\Phi^{-1}(\hat{u}_{i2}) - \Phi^{-1}(u_2)}{\hat{h}_2}\right)}{n\hat{h}_1\hat{h}_2\ \ \phi\left(\Phi^{-1}(u_1)\right)\ \ \phi\left(\Phi^{-1}(u_2)\right)} \end{aligned} \tag{3.7}$$

The copula $C_{12} : [0,1]^2 \to [0,1]$ can be estimated by

$$\hat{C}_{12}(u_1, u_2) := \int_0^{u_1}\int_0^{u_2} \hat{c}_{12}(s,t)\,dt\,ds.$$

and the estimator of the conditional copula $C_{2|1}$ is given by

$$\hat{C}_{2|1}(u_2|u_1) := \int_0^{u_2} \hat{c}_{2|1}(t|u_1)\,dt = \int_0^{u_2} \hat{c}_{12}(u_1, t)\,dt.$$

This comes from the fact that

$$c_{2|1}(u_2|u_1) = \frac{c_{12}(u_1,u_2)}{\underbrace{c_1(u_1)}_{\equiv 1, \text{ since } U(0,1) \text{ density}}} = c_{12}(u_1,u_2).$$

Hence,

$$\begin{aligned}
\hat{C}_{2|1}(u_2|u_1) &:= \int_0^{u_2} \hat{c}_{12}(u_1,t)\,dt \\
&\overset{(3.7)}{=} \frac{1}{n\hat{h}_1\phi(\Phi^{-1}(u_1))}\sum_{i=1}^{n} k\left(\frac{\Phi^{-1}(\hat{u}_{i1}) - \Phi^{-1}(u_1)}{\hat{h}_1}\right) \\
&\quad \cdot \int_0^{u_2} \frac{1}{\hat{h}_2\phi(\Phi^{-1}(t))} k\left(\frac{\Phi^{-1}(\hat{u}_{i2}) - \Phi^{-1}(t)}{\hat{h}_2}\right) dt \\
&\overset{(2.26)}{=} \frac{1}{n\hat{h}_1\phi(\Phi^{-1}(u_1))}\sum_{i=1}^{n} k\left(\frac{\Phi^{-1}(\hat{u}_{i1}) - \Phi^{-1}(u_1)}{\hat{h}_1}\right) \\
&\quad \cdot H\left(\frac{\Phi^{-1}(t) - \Phi^{-1}(\hat{u}_{i2})}{\hat{h}_2}\right)\Bigg|_{t=0}^{t=u_2}
\end{aligned}$$

where $H(x) = \int_{-\infty}^{x} k(t)\,dt$ is the integrated kernel function. For details of the calculation we refer to Section 2.1.

Example 1 (Integrated kernels)
For the Epanechnikov kernel function k_E,

$$H_E(x) = \begin{cases} 0, & x \le -\sqrt{5}; \\ \frac{3}{4\sqrt{5}}\left(x - \frac{x^3}{15}\right) + \frac{1}{2}, & -\sqrt{5} < x < \sqrt{5}; \\ 1, & x \ge \sqrt{5} \end{cases}$$

For the Gaussian kernel function k_G, $H_G(x)$ coincides with the standard normal cumulative distribution function Φ.

We see that we can calculate the conditional distribution functions explicitly. Set

$$\begin{aligned}
\hat{u}_{i,2|1} &:= \hat{F}_{2|1}(\hat{u}_{i2}|\hat{u}_{i1}) = \hat{C}_{2|1}(\hat{u}_{i2}|\hat{u}_{i1}), \\
\hat{u}_{i,3|1} &:= \hat{F}_{3|1}(\hat{u}_{i3}|\hat{u}_{i1}) = \hat{C}_{3|1}(\hat{u}_{i3}|\hat{u}_{i1}) \text{ and} \\
\hat{z}_{i,2|1} &:= \Phi^{-1}(\hat{u}_{i,2|1}), \quad \hat{z}_{i,3|1} := \Phi^{-1}(\hat{u}_{i,3|1}).
\end{aligned}$$

With this at hand, we can proceed to estimate the copula density $c_{23;1}$.

From the theory, the marginal densities $g_1, g_2, g_3, g_{2;1}, g_{3;1}$ are standard normal probability densities. In Figure 3.2, we display the estimated marginal probability density functions which we obtained by integration of the bivariate density estimate, for example,

$$\hat{g}_1(z_1) = \int_{-\infty}^{\infty} \hat{g}_{12}(z_1, t)\, dt.$$

We see that it is not assured that the estimated margins reveal the theoretical shape. In contrast, for the following Bernstein copulas it is guaranteed that the marginal pair-copula densities behave as expected from the theory.

3.3 Estimation by Bernstein Polynomials

Bernstein copulas were introduced by Sancetta and Satchell [2004]. Compared to other approximating polynomials, they state that Bernstein polynomials are advantageous in the copula framework for several reasons: they are closed under differentiation, they lead to proper copula functions under simple restrictions on the coefficients and the resulting estimator has a lower variance than the empirical copula estimator. Sancetta and Satchell [2004] use Bernstein polynomials to estimate the copula function C directly. The copula density c is then obtained as the derivative of C. Schellhase [2012] shows a different approach following the idea of Marx and Eilers [2005]. He estimates the copula density c by a combination of tensor products of univariate B-splines. We perform this estimation method, but with standardized Bernstein polynomials instead of B-splines. The copula C is then deduced by integration.
Weiß and Scheffer [2012] use bivariate Bernstein copulas in the construction of smooth nonparametric vine models. They perform a simulation study and a financial data analysis and show that their model approach is superior to parametric vine models with respect to AIC.
In this section, we refer to Schellhase [2012].

Definition 3.1 (Bernstein Polynomial)
The function $\tilde{P}_{v,m} : [0,1] \to \mathbb{R}_{\geq 0}$,

$$\tilde{P}_{v,m}(u) = \binom{m}{v} u^v (1-u)^{m-v}, \quad u \in [0,1],$$

with $m \in \mathbb{N}, v \in \{0, 1, \ldots, m\}$, *is called Bernstein Polynomial of degree* m.

It is advantageous to standardize the Bernstein polynomials in a way, such that they become probability densities. They already fulfill the property of being nonnegative, so we need them to integrate to one over the unit interval. This is done in the following definition.

Definition 3.2 (Standardized Bernstein Polynomial)
The function $P_{v,m} : [0,1] \to \mathbb{R}_{\geq 0}$,

$$P_{v,m}(u) = (m+1)\binom{m}{v} u^v (1-u)^{m-v}, \quad u \in [0,1],$$

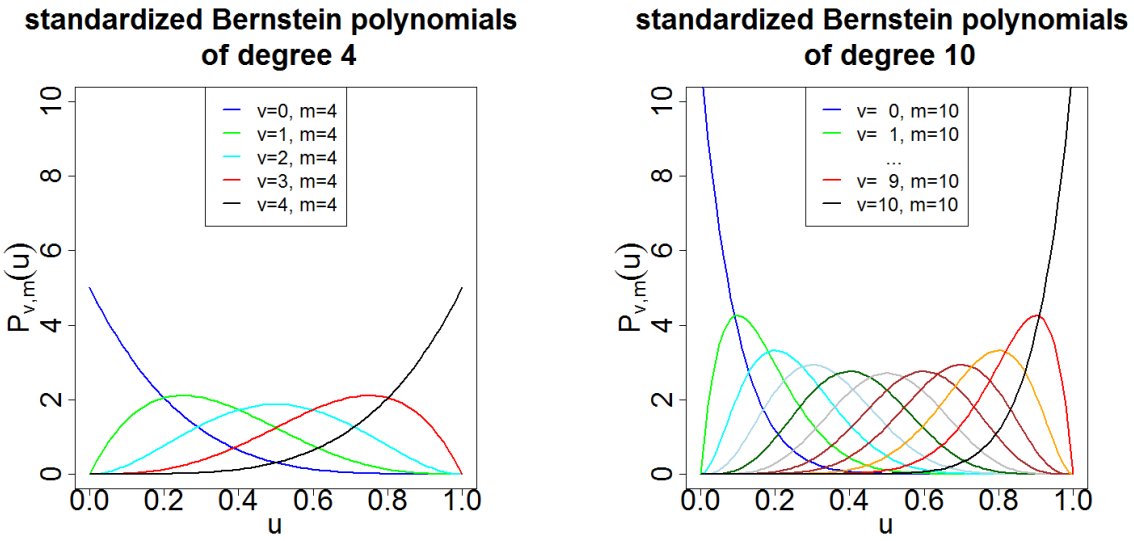

Figure 3.1: Illustration of standardized Bernstein polynomials of degrees four and ten.

with $m \in \mathbb{N}, v \in \{0, 1, \ldots, m\}$*, is called standardized Bernstein Polynomial of degree* m*.*

In fact, the standardized Bernstein polynomial $P_{v,m}$ coincides with the probability density of a Beta distribution $Beta(\alpha, \beta)$ with shape parameter $\alpha, \beta > 0$, if the parameter are set to $\alpha = v+1$ and $\beta = m-v+1$. In Figure 3.1, we see standardized Bernstein polynomials of degree $m = 4$ in the left panel and of degree $m = 10$ in the right panel. We notice that $P_{v,m}(u)$ has high values for $v = 0$ or $v = m$, when u is near 0 or 1, respectively. We guess this to have an impact on the tail-behavior of the pair-copula density estimate.

Let $\mathbf{P}_m(u) = (P_{0,m}(u), \ldots, P_{m,m}(u))$ be a set of standardized Bernstein polynomials. We call $\mathbf{P}_m(u)$ a univariate Bernstein basis with basis functions $P_{0,m}(u), \ldots, P_{m,m}(u)$. We define a bivariate tensor product of the Bernstein bases $\mathbf{P}_m(u_1), \mathbf{P}_m(u_2)$ as

$$\begin{aligned} \mathbf{P}_m(u_1, u_2) &:= \mathbf{P}_m(u_1) \otimes \mathbf{P}_m(u_2) \\ = \{&P_{0,m}(u_1) P_{0,m}(u_2), P_{1,m}(u_1) P_{0,m}(u_2), \ldots, P_{m,m}(u_1) P_{0,m}(u_2), \\ &P_{0,m}(u_1) P_{1,m}(u_2), \ldots, P_{m,m}(u_1) P_{m,m}(u_2)\}. \end{aligned}$$

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})$ be a sample of independent and identically distributed random variables $\mathbf{X}_i = (X_{i1}, \ldots, X_{id}), i = 1, \ldots, n$. Let $D \subset \{1, \ldots, d\}$ be a subset of indices, $j, k \in \{1, \ldots, d\} \setminus D, j \neq k$ and $x_D := (x_l : l \in D)$. With the notation

$$\begin{aligned} c_{jk;D} &:= c_{jk;D}\left(F\left(x_j | x_D\right), F\left(x_k | x_D\right)\right), \\ u_{j|D} &:= F\left(x_j | x_D\right), \end{aligned}$$

the copula density estimator is defined by means of the tensor product as

$$\begin{aligned}\hat{c}_{jk;D}\left(u_{j|D}, u_{k|D}; \boldsymbol{\alpha}^{(jk|D)}\right) &:= \left[\mathbf{P}_m\left(u_{j|D}\right) \otimes \mathbf{P}_m\left(u_{k|D}\right)\right] \boldsymbol{\alpha}^{(jk|D)} \\ &= \sum_{\nu_1=0}^{m}\sum_{\nu_2=0}^{m} P_{\nu_1,m}\left(u_{j|D}\right) P_{\nu_2,m}\left(u_{k|D}\right) \alpha_{\nu_1,\nu_2}^{(jk|D)}.\end{aligned} \tag{3.8}$$

The coefficient vector $\boldsymbol{\alpha}^{(jk|D)} := \left(\alpha_{0,0}^{(jk|D)}, \ldots, \alpha_{0,m}^{(jk|D)}, \ldots, \alpha_{m,m}^{(jk|D)}\right) \in \mathbb{R}^{(m+1)^2}$ is required to satisfy the constraints

$$\begin{aligned}\alpha_{\nu_1,\nu_2}^{(jk|D)} &\geq 0 \quad \text{for all} \quad \nu_1, \nu_2 = 0, 1, \ldots, m \quad \text{and} \\ \sum_{\nu_1=0}^{m}\sum_{\nu_2=0}^{m} \alpha_{\nu_1,\nu_2}^{(jk|D)} &= 1\end{aligned} \tag{3.9}$$

in order to ensure that the resulting estimate is indeed a density. Furthermore, we need the resulting copula density estimate to have uniform margins. We reach this by further restrictions on the coefficient vector, which we derive subsequently.

The bivariate copula density estimator $\hat{c}_{jk;D}$ has uniform marginal densities $\hat{c}_{j;D}\left(u_{j|D}\right) = \int \hat{c}_{jk;D}\left(u_{j|D}, u_{k|D}\right) du_{k|D}$ and $\hat{c}_{k;D}\left(u_{k|D}\right) = \int \hat{c}_{jk;D}\left(u_{j|D}, u_{k|D}\right) du_{j|D}$ if $\hat{c}_{j;D} \equiv 1$ and $\hat{c}_{k;D} \equiv 1$.

$$\begin{aligned}\hat{c}_{j;D}\left(u_{j|D}\right) &= \int_0^1 \hat{c}_{jk;D}\left(u_{j|D}, u_{k|D}\right) \quad du_{k|D} \\ &= \int_0^1 \sum_{\nu_1=0}^{m}\sum_{\nu_2=0}^{m} P_{\nu_1,m}\left(u_{j|D}\right) P_{\nu_2,m}\left(u_{k|D}\right) \alpha_{\nu_1,\nu_2}^{(jk|D)} \quad du_{k|D} \\ &= \sum_{\nu_1=0}^{m}\sum_{\nu_2=0}^{m} P_{\nu_1,m}\left(u_{j|D}\right) \alpha_{\nu_1,\nu_2}^{(jk|D)} \underbrace{\int_0^1 P_{\nu_2,m}\left(u_{k|D}\right) \quad du_{k|D}}_{=1, \quad \text{since density}} \\ &= \sum_{\nu_1=0}^{m} (m+1)\, \tilde{P}_{\nu_1,m}\left(u_{j|D}\right) \sum_{\nu_2=0}^{m} \alpha_{\nu_1,\nu_2}^{(jk|D)}.\end{aligned}$$

Assume that $\sum_{\nu_2=0}^{m} \alpha_{\nu_1,\nu_2}^{(jk|D)} = \frac{1}{m+1}$ holds for all $\nu_1 = 0, 1, \ldots, m$ and use the property, that the sum of Bernstein polynomials is a partition of one, i. e. $\sum_{\nu=0}^{m} \tilde{P}_{\nu,m}(u) = 1$ (this follows directly from the binomial theorem). Then,

$$\begin{aligned}\hat{c}_{j;D}\left(u_{j|D}\right) &= \sum_{\nu_1=0}^{m} (m+1)\, \tilde{P}_{\nu_1,m}\left(u_{j|D}\right) \underbrace{\sum_{\nu_2=0}^{m} \alpha_{\nu_1,\nu_2}^{(jk|D)}}_{\equiv \frac{1}{m+1} \quad \text{by ass.}} \\ &= \sum_{\nu_1=0}^{m} \tilde{P}_{\nu_1,m}\left(u_{j|D}\right) = 1.\end{aligned}$$

Hence, we impose the following constraint on the coefficients:

$$\alpha_{v_1,\cdot}^{(jk|D)} := \sum_{v_2=0}^{m} \alpha_{v_1,v_2}^{(jk|D)} = \frac{1}{m+1} \quad \text{for all} \quad v_1 = 0, 1, \ldots, m. \tag{3.10}$$

The conditional distribution function estimator is obtained by integration of the estimated copula density as

$$\begin{aligned}
\hat{F}_{j|D}\left(x_j|x_D\right) &= \frac{\partial \hat{C}_{jr|D_{-r}}\left(\hat{F}\left(x_j|x_{D_{-r}}\right), \hat{F}\left(x_r|x_{D_{-r}}\right)\right)}{\partial \hat{F}\left(x_r|x_{D_{-r}}\right)} \\
&= \int_0^{u_{j|D_{-r}}} \hat{c}_{jr;D_{-r}}\left(t, u_{r|D_{-r}}\right) \quad dt \\
&= \sum_{v_1=0}^{m} \sum_{v_2=0}^{m} \overline{P}_{v_1,m}\left(u_{j|D_{-r}}\right) P_{v_2,m}\left(u_{r|D_{-r}}\right) \alpha_{v_1,v_2}^{(jr|D_{-r})} \\
&=: \hat{u}_{j|D}.
\end{aligned}$$

The function $\overline{P}_{v_1,m}$ denotes the integrated standardized Bernstein polynomial, which coincides with the cumulative distribution function of the Beta distribution.

The coefficient vector $\boldsymbol{\alpha}^{(jk|D)}$ is obtained by maximum likelihood estimation with side constraints. Assume that $\hat{u}_{ij|D} = \hat{F}\left(x_{ij}|x_D\right), \hat{u}_{ik|D} = \hat{F}\left(x_{ik}|x_D\right), i = 1, \ldots, n$ are already estimated. The log-likelihood contribution of the fitted pair-copula density $\hat{c}_{jk;D}$ is

$$l_{jk;D}\left(\boldsymbol{\alpha}^{(jk|D)}\right) = \sum_{i=1}^{n} \ln\left\{\left[\mathbf{P}_m\left(\hat{u}_{ij|D}\right) \otimes \mathbf{P}_m\left(\hat{u}_{ik|D}\right)\right] \boldsymbol{\alpha}^{(jk|D)}\right\}. \tag{3.11}$$

Maximization of expression (3.11) with respect to $\boldsymbol{\alpha}^{(jk|D)}$ under the side constraints (3.9) and (3.10) yields a proper estimate for $\boldsymbol{\alpha}^{(jk|D)}$. Schellhase [2012] states that this is a quadratic programming problem with linear side constraints and hence, he implements the maximization routine in his `R` package `penDvine` using the `R` package `quadprog`. The `R` package `penDvine` is designed for penalized pair-copula estimation, which we discuss in the next section. However, for suitable parameter settings one can also perform unpenalized estimation. We explain this subsequently in more detail.

It is worth to emphasize that, in contrast to the kernel density estimation approach, in the estimation with Bernstein polynomials, the coefficients are required to fulfill certain constraints in order to guarantee uniform margins. For kernel density estimation approach it is not guaranteed that the margins indeed coincide with their theoretical distribution. In our case in Section 3.2 it is the normal distribution since we perform the estimation on the normal quantiles of uniform. We illustrate this in Figure 3.2 by means of fitted pair-copula densities from a three dimension C-Vine with pairs 12, 13, 23;1. We fitted the densities to the *uranium* data from the empirical study.
In the empirical study we see that the kernel density approach is more flexible than the

Bernstein polynomial approach. We claim that, among others, this is also up to the coefficient constraints.

In the following section, we introduce a penalization to enhance the smoothness of the resulting copula density estimate.

3.4 Estimation by penalized Bernstein Polynomials

The estimation approach by Bernstein polynomials, as described in Section 3.3, leads to a density estimator with $(m+1)^2$ coefficients, where $m \in \mathbb{N}$ is the prespecified order of Bernstein polynomials to be used. The result may be wiggly or *not smooth enough*. We extend that approach to overcome this undesired fact, following the presentation in Schellhase [2012].

A penalization is performed in order to support a smooth fit of copula density estimate. The penalty term is added to the log-likelihood equation in (3.11), such that the log-likelihood is of the form

$$l_{jk;D}\left(\boldsymbol{\alpha}^{(jk|D)}, \lambda^{(jk|D)}\right) = l_{jk;D}\left(\boldsymbol{\alpha}^{(jk|D)}\right) - \frac{1}{2}\lambda^{(jk|D)}\left(\boldsymbol{\alpha}^{(jk|D)}\right)^t \boldsymbol{Q}\boldsymbol{\alpha}^{(jk|D)}, \tag{3.12}$$

where $\lambda^{(jk|D)} \geq 0$ is called *penalty parameter* and $\boldsymbol{Q}$ is a penalty matrix. Assuming $\boldsymbol{Q}$ to be positive definite implies that $-\frac{1}{2}\lambda^{(jk|D)}\left(\boldsymbol{\alpha}^{(jk|D)}\right)^t \boldsymbol{Q}\boldsymbol{\alpha}^{(jk|D)} \leq 0$, i. e. it is in fact a penalty on the likelihood. The penalty parameter controls the amount of smoothing, the penalty matrix is chosen according to the penalty method. There are two versions of the penalty matrix for Bernstein polynomials. Either penalization on coefficients or penalization on the curvature of the pair-copula density.

The idea behind the penalization on the curvature is that a smooth pair-copula density shall have small integrated squared second order derivatives. Recall Equation (3.8), where we defined the Bernstein pair-copula density estimator (now with unconditional notation for convenience) as

$$\begin{aligned}\hat{c}_{jk}\left(u_j, u_k; \boldsymbol{\alpha}^{(jk)}\right) &:= \left[\mathbf{P}_m\left(u_j\right) \otimes \mathbf{P}_m\left(u_k\right)\right]\boldsymbol{\alpha}^{(jk)} \\ &= \sum_{v_1=0}^{m}\sum_{v_2=0}^{m} P_{v_1,m}\left(u_j\right) P_{v_2,m}\left(u_k\right)\alpha_{v_1,v_2}^{(jk)}.\end{aligned}$$

The second order derivative of the standardized Bernstein polynomial is, see Doha et al. [2011],

$$\frac{\partial^2 P_{v,m}(u)}{\partial u^2} = \frac{(m+1)!}{(m-2)!}\sum_{l=max(0,v+2-m)}^{min(v,2)} (-1)^{l+2}\binom{2}{l} P_{v-l,m-2}(u), \quad v = 0, 1, \ldots, m. \tag{3.13}$$

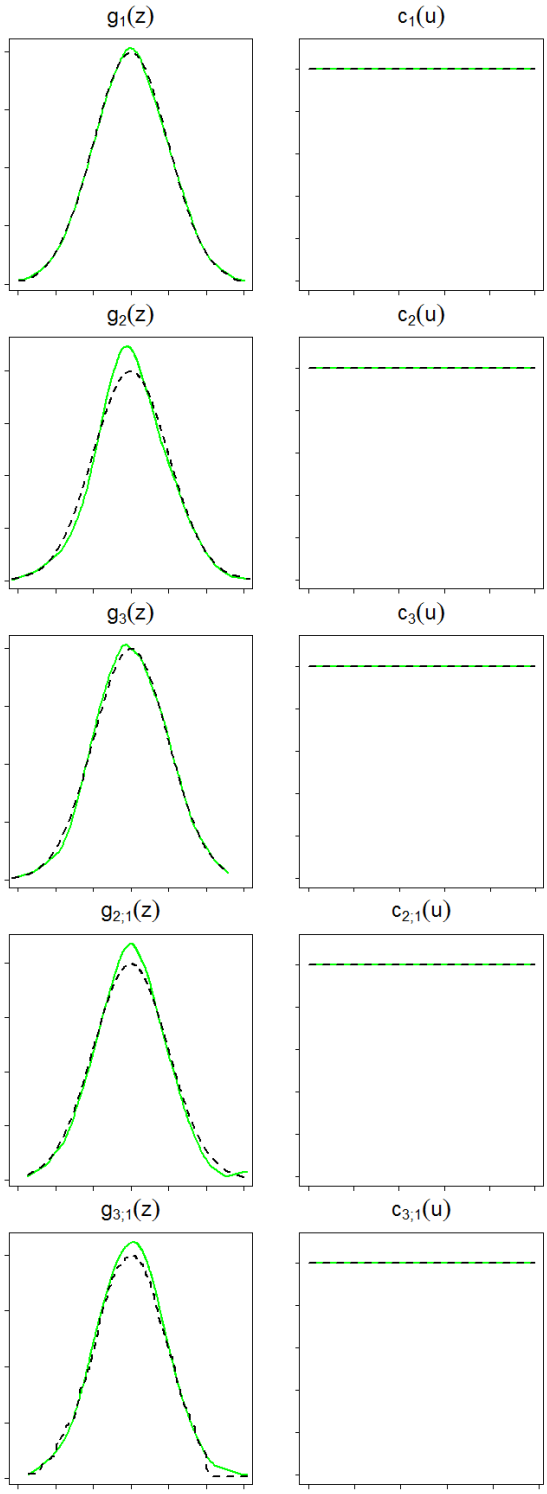

Figure 3.2: Estimated pair-copula density margins from a C-Vine with pairs 12, 13, 23;1 fitted to the uranium data. The green function is the theoretical probability density function, the black dotted function is the fitted marginal density function. On the left hand side, the estimation is done by kernel density estimation, on the right hand side, the estimation is done by Bernstein polynomials.

Example 2 (Second Order Derivatives of standardized Bernstein polynomials)

We illustrate Equation (3.13) for $m = 4$ *and* $v = 0, 1, 2, 3, 4$.

$$\begin{aligned}
\frac{\partial^2 P_{0,4}(u)}{\partial u^2} &= \frac{5!}{2!}\sum_{l=0}^{0}(-1)^{l+2}\binom{2}{l}P_{0-l,2}(u) = \frac{5!}{2!}\left(P_{0,2}(u)\right)\\
\frac{\partial^2 P_{1,4}(u)}{\partial u^2} &= \frac{5!}{2!}\sum_{l=0}^{1}(-1)^{l+2}\binom{2}{l}P_{1-l,2}(u) = \frac{5!}{2!}\left(P_{1,2}(u) - 2P_{0,2}(u)\right)\\
\frac{\partial^2 P_{2,4}(u)}{\partial u^2} &= \frac{5!}{2!}\sum_{l=0}^{2}(-1)^{l+2}\binom{2}{l}P_{2-l,2}(u) = \frac{5!}{2!}\left(P_{2,2}(u) - 2P_{1,2}(u) + P_{0,2}(u)\right)\\
\frac{\partial^2 P_{3,4}(u)}{\partial u^2} &= \frac{5!}{2!}\sum_{l=1}^{2}(-1)^{l+2}\binom{2}{l}P_{3-l,2}(u) = \frac{5!}{2!}\left(-2P_{2,2}(u) + P_{1,2}(u)\right)\\
\frac{\partial^2 P_{4,4}(u)}{\partial u^2} &= \frac{5!}{2!}\sum_{l=2}^{2}(-1)^{l+2}\binom{2}{l}P_{4-l,2}(u) = \frac{5!}{2!}\left(P_{2,2}(u)\right)
\end{aligned}$$

In matrix vector notation, we get

$$\begin{aligned}
\frac{\partial^2}{\partial u^2}\mathbf{P}_4(u) &= \left(\frac{\partial^2 P_{0,4}(u)}{\partial u^2}, \frac{\partial^2 P_{1,4}(u)}{\partial u^2}, \frac{\partial^2 P_{2,4}(u)}{\partial u^2}, \frac{\partial^2 P_{3,4}(u)}{\partial u^2}, \frac{\partial^2 P_{4,4}(u)}{\partial u^2}\right)\\
&= \frac{5!}{2!}\left(P_{0,2}(u), P_{1,2}(u), P_{2,2}(u)\right)\begin{pmatrix} 1 & -2 & 1 & 0 & 0\\ 0 & 1 & -2 & 1 & 0\\ 0 & 0 & 1 & -2 & 1\end{pmatrix}.
\end{aligned}$$

In general, let

$$w := \frac{(m+1)!}{(m-2)!} \quad \text{and} \quad B = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0\\ 0 & 1 & -2 & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & 1 & -2 & 1\end{pmatrix} \in \mathbb{R}^{(m-2)\times(m+1)}.$$

Then, the vector of second order derivatives in Equation (3.13) can be expressed as

$$\frac{\partial^2}{\partial u^2}\mathbf{P}_m(u) = w\left(\mathbf{P}_{m-2}(u)\cdot B\right). \tag{3.14}$$

Now we can construct the penalty matrix. The integrated squared second order derivatives of the pair-copula density estimator from Equation (3.8) is

$$\int\int\left(\frac{\partial^2 \hat{c}_{jk}\left(u_j, u_k\right)}{\partial u_j^2}\right)^2 + \left(\frac{\partial^2 \hat{c}_{jk}\left(u_j, u_k\right)}{\partial u_k^2}\right)^2 du_j\ du_k.$$

Schellhase [2012] states that this can be expressed as

$$\int\int\left(\frac{\partial^2 \hat{c}_{jk}\left(u_j, u_k\right)}{\partial u_j^2}\right)^2 + \left(\frac{\partial^2 \hat{c}_{jk}\left(u_j, u_k\right)}{\partial u_k^2}\right)^2 du_j\ du_k = \left(\boldsymbol{\alpha}^{(jk)}\right)^t\left(\mathbf{Q}_{u_j} + \mathbf{Q}_{u_k}\right)\boldsymbol{\alpha}^{(jk)}$$

for some matrices $\mathbf{Q}_{u_j}, \mathbf{Q}_{u_k}$. For details we refer to Schellhase [2012]. Set

$$\mathbf{Q} := \mathbf{Q}_{u_j} + \mathbf{Q}_{u_k}$$

as the penalty matrix for penalization on the curvature of $\hat{c}_{jk}$. This can be used in the penalized likelihood equation (3.12).

Alternatively, one can penalize the r-th order differences of the coefficients. This is commonly used in spline smoothing, see Eilers and Marx [1996]. The penalty matrix is of the form

$$\mathbf{Q} := (\mathbf{1}_{m+1} \otimes L)^t \, (L \otimes \mathbf{1}_{m+1}),$$

where

$$L = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(m-1)\times(m)} \quad \text{for first order differences, or}$$

$$L = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{(m-2)\times(m)} \quad \text{for second order differences.}$$

Regarding the penalty parameter $\lambda^{(jk|D)}$, Schellhase [2012] applies spline smoothing arguments and Bayesian arguments to view the coefficient vector $\boldsymbol{\alpha}^{(jk|D)}$ as related to the penalization as follows

$$\boldsymbol{\alpha}^{(jk|D)} \sim N\left(\mathbf{0}, \frac{1}{\lambda^{(jk|D)}} \mathbf{Q}^{\leftarrow}\right), \tag{3.15}$$

where $\mathbf{Q}^{\leftarrow}$ is the generalized inverse of the penalty matrix. From this point of view, the coefficient vector is a priori normally distributed and the penalty parameter is involved in the distribution parameter. Schellhase [2012] states that the prior in (3.15) is degenerated, so he works with a singular value decomposition to deduce a log-likelihood expression $l_{jk;D}\left(\lambda^{(jk|D)}, \hat{\boldsymbol{\alpha}}^{(jk|D)}\right)$. Here, $\hat{\boldsymbol{\alpha}}^{(jk|D)}$ denotes the maximum likelihood estimate from the penalized likelihood in Equation (3.12) with a prespecified $\lambda^{(jk|D)}$. The likelihood is maximized iteratively. The maximization assumes a starting value $\lambda_0^{(jk|D)}$ and calculates $\hat{\boldsymbol{\alpha}}^{(jk|D)}$. This coefficient vector is plugged into $l_{jk;D}\left(\lambda^{(jk|D)}, \hat{\boldsymbol{\alpha}}^{(jk|D)}\right)$, such that $\lambda^{(jk|D)}$ can be estimated. With this $\lambda^{(jk|D)}$, we can calculate the next maximum likelihood estimate of the coefficient vector, and so on. The estimation stops when certain tolerance criteria are satisfied or when the maximum number of iterations is reached. For details, see source code of `R` package `penDvine`, Version xy, especially functions `my.loop` and `new.lambda`.

In our empirical study, we applied both, penalized and unpenalized Bernstein estimation, for pair-copula estimation. We implemented these analyses by means of the `R`

package `penDvine`. As already mentioned, the package is designed for the penalized approach. We performed the unpenalized estimation by fixing the penalty parameter λ at zero. However, this does not work with Bernstein functions of arbitrary degree. We had to keep the degree of the Bernstein polynomials under five, otherwise numerical problems occured.

Chapter 4

Empirical Study

In this section, we aim at studying the behavior of nonparametric C-Vines on the basis of a three-dimensional data set. We focus on the example in three dimensions as we want to concentrate on the nonparametric estimation methods and not on tree selection. For model comparison we use cross-validated likelihoods as described in Section 2.3.

4.1 Exploratory Data Analysis

The Uranium Exploration Data of Cook and Johnson [1986] is available in the `R` package `copula` (Hofert et al. [2012]). It consists of log concentration measurements of seven chemicals in 655 water samples collected near Grand Junction, Colorado. The chemicals are Uranium (U), Lithium (Li), Cobalt (Co), Potassium (K), Cesium (Cs), Scandium (Sc) and Titanium (Ti). For simplicity, we focus on C-Vines with the variables Co, Ti and Sc, and relabel them with numbers 1, 2 and 3. Acar et al. [2012] already performed parametric investigations on the same data subset.

By definition, copulas have uniformly distributed margins on $[0, 1]$. In order to fit Vine copula models to the uranium data, the first step is to transform the data, such that each margin is uniformly distributed on $[0, 1]$. Traditionally, the marginal data is transformed by an estimated cumulative distribution function. We decide to estimate the univariate cumulative distribution function by means of kernel density estimation. Alternatively, transformation to relative ranks is possible, too.

For $n \in \mathbb{N}$ observations $x_{1j}, \ldots, x_{nj}$ of the continuous univariate random variable X_j, we estimate the corresponding cumulative distribution function $\hat{F}_j(x)$ by kernel density estimation with Epanechnikov kernels and bandwidth from least squares cross-validation. Then, we calculate the copula data as $u_{ij} = \hat{F}_j\left(x_{ij}\right)$ for $i = 1, \ldots, n$. We apply this procedure to all random variables X_j, $j = 1, 2, 3$. For the estimation of $\hat{F}_j(x)$, $j = 1, 2, 3$ we use the `R` function `npudist` from the `np` package. We call the resulting data *copula* data or *u data*.

We apply these marginal transformations to the uranium data. In Figure 4.1, we depict

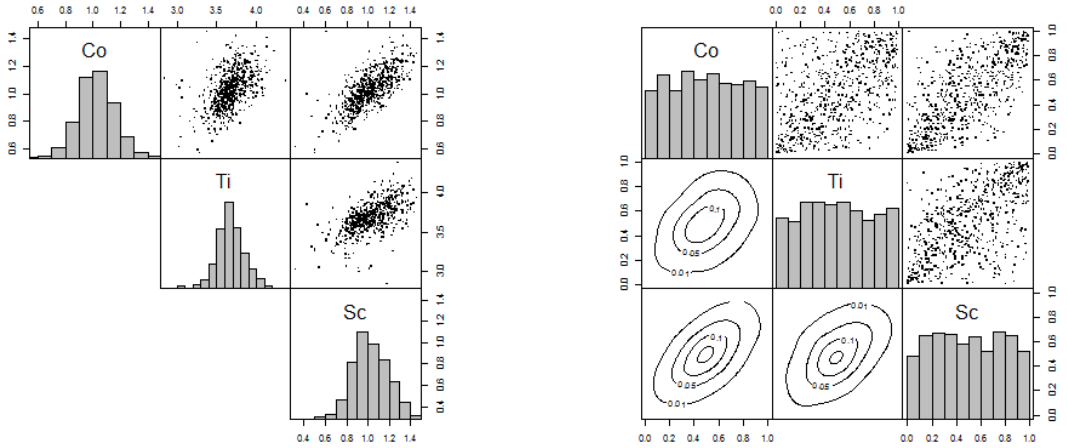

Figure 4.1: Scatter plots, histograms and empirical contours of original uranium data and copula data.

	(Co, Ti)	(Ti, Sc)	(Co, Sc)
$\tau_n(\cdot,\cdot)$	0.365	0.436	0.535
$\rho_n(\cdot,\cdot)$	0.517	0.600	0.718

Table 4.1: Empirical Kendall's tau and Spearman's rho for three-dimensional uranium copula data.

the original uranium data as well as the copula data, which is approximately uniformly distributed on $[0,1]$, as desired. Empirical Kendall's τ and Spearman's ρ in Table 4.1 indicate that the variable pair (Co, Sc) has the highest dependence, followed by (Ti, Sc). The tree selection algorithm by Dissmann et al. [2013] applied to this C-Vine with three variables would suggest to choose Sc as root.

4.2 Nonparametric C-Vine Estimation

Notation	First Tree	Pair-Copula Densities
PCC 1	2-1-3	$c_{12}, c_{13}, c_{23;1}$
PCC 2	1-2-3	$c_{12}, c_{23}, c_{13;2}$
PCC 3	1-3-2	$c_{13}, c_{23}, c_{12;3}$

Table 4.2: Overview of considered pair-copula constructions. (1=Co, 2=Ti, 3=Sc).

Notation	Density Estimator
kde.ls	$\hat{c}^{kde.ls}(u_1,u_2) = \frac{1}{n\hat{h}_1\hat{h}_2}\sum_{i=1}^{n} k\left(\frac{\Phi^{-1}(u_{i1})-\Phi^{-1}(u_1)}{\hat{h}_1}\right) k\left(\frac{\Phi^{-1}(u_{i2})-\Phi^{-1}(u_2)}{\hat{h}_2}\right)$
kde.ml	$\hat{c}^{kde.ml}(u_1,u_2) = \frac{1}{n\hat{h}_1\hat{h}_2}\sum_{i=1}^{n} k\left(\frac{\Phi^{-1}(u_{i1})-\Phi^{-1}(u_1)}{\hat{h}_1}\right) k\left(\frac{\Phi^{-1}(u_{i2})-\Phi^{-1}(u_2)}{\hat{h}_2}\right)$
kde.rule	$\hat{c}^{kde.rule}(u_1,u_2) = \frac{1}{n\hat{h}_1\hat{h}_2}\sum_{i=1}^{n} k\left(\frac{\Phi^{-1}(u_{i1})-\Phi^{-1}(u_1)}{\hat{h}_1}\right) k\left(\frac{\Phi^{-1}(u_{i2})-\Phi^{-1}(u_2)}{\hat{h}_2}\right)$
bern.unpen	$\hat{c}^{bern.unpen}(u_1,u_2) = \sum_{v_1=0}^{m}\sum_{v_2=0}^{m} P_{v_1,m}(u_1)\, P_{v_2,m}(u_2)\, \alpha_{v_1,v_2}, \quad m = 4$
bern.penfix	$\hat{c}^{bern.pen.fix}(u_1,u_2) = \sum_{v_1=0}^{m}\sum_{v_2=0}^{m} P_{v_1,m}(u_1)\, P_{v_2,m}(u_2)\, \alpha_{v_1,v_2}, \quad m = 10$
bern.pen	$\hat{c}^{bern.pen}(u_1,u_2) = \sum_{v_1=0}^{m}\sum_{v_2=0}^{m} P_{v_1,m}(u_1)\, P_{v_2,m}(u_2)\, \alpha_{v_1,v_2}, \quad m = 10$
par	$\hat{c}^{par}(u_1,u_2)$ from parametric family

Notation	Kernel Specification	Bandwidth Specification	Penalization
kde.ls	Epanechnikov	least squares cross-validation	-
kde.ml	Epanechnikov	maximum likelihood cross-validation	-
kde.rule	Gauss	normal ref. rule of thumb	-
bern.unpen	-	-	no, $\boldsymbol{\lambda} \equiv (0,0,0)$
bern.penfix	-	-	yes, on curvature, $\boldsymbol{\lambda} \equiv (1,1,1)$
bern.pen	-	-	yes, on curvature, $\boldsymbol{\lambda}$ optimized
par	-	-	-

Table 4.3: Overview of considered pair-copula density estimators.

We model the dependency present in the uranium data by C-Vines with kernel density estimation performed on standard normal quantiles of the copula data (cf. Section 3.2). We involve three different bandwidth selection methods. Furthermore, we fit C-Vines with Bernstein polynomials, both with and without penalization (cf. Sections 3.3 and 3.4), and compare the contours to the contours of the kernel estimation approaches. In Table 4.2 we define the possible C-Vine decompositions and label them for conventient notation. In three dimensions, there is a total of three possible variable orderings. In Table 4.3 we specify the pair-copula density estimators used in the model fittings.

The fitted pair-copula density contours are displayed in Figures 4.2, 4.3 and 4.4. Each model, i. e. for each decomposition from Table 4.2 and each estimator from Table 4.3, is fitted twice. The first contour estimate is evaluated on a 60 by 60 grid, the second contour estimate is evaluated on a 100 by 100 grid. We can say that there is no remarkable influence of the grid size on the contours' appearances. All methods lead to similar results for the two runs of contour fitting. In Table 4.4 we report the estimated bandwidths and in Table 4.5 the optimized penalty parameters. In foregoing investigations we fitted maximum likelihood bandwidths in combination with Epanechnikov kernels. This indeed leads to inconsistent results as already remarked in Section 2.1.2. Rerunning the model estimation with the same settings led, sometimes, to substantially different bandwidths and therefore to substantially different contours. We then changed the kernel to the infinitely supported Gaussian kernel and obtained more robust estimates.
It is striking that *kde.ls* always returns the most wiggled pair-copula density contours. *kde.rule* and *kde.ml* perform quite similar. The appearance of the Bernstein estimates on the first (unconditioned) level reminds of a Frank copula. We guess this is a consequence of the coefficient constraints. Recalling the shape of standardized Bernstein polynomials from Figure 3.1, we see that $P_{v,m}(u)$ has high values for $v = 0$ or $v = m$, when u is near 0 or 1. Consequently, if the relevant coefficient estimates are not *small enough*, these high values have impact on the density estimate $\hat{c}(u_1, u_2)$, especially near $(0, 0)$ and $(1, 1)$. Related to the graph, this is the region in the lower left corner and the upper right corner.
With Bernstein copulas we have three setups. *bern.unpen* is with polynomials of degree four and no penalization, *bern.penfix* is with polynomials of degree ten and a small fixed penalization, *bern.pen* is with polynomials of degree ten and a optimal estimated penalization. In foregoing investigations, we compared *bern.pen* and *bern.unpen* for the same degree of Bernstein polynomials, namely $m = 4$. The result was, as expected, that *bern.unpen* fits slightly more flexible, *bern.pen* slightly smoother. But this was only remarkable for the contours of unconditioned pair-copula densities c_{jk} and not for the contours of the pair-copula density based on conditional values $c_{jk;D}$.
We changed the setting as defined in Table 4.3 and work with *bern.unpen* with $m = 4$ and *bern.penfix* $m = 10$. On the one hand, *bern.unpen* can be interpreted as more flexible than *bern.penfix* and *bern.pen* due the absence of penalization. On the other hand, the advantage of *bern.penfix* and *bern.pen* over *bern.unpen* is the richnes of the Bernstein basis $\mathbf{P}_m(u) = \{P_{0,m}(u), \ldots, P_{m,m}(u)\}$. Note that for a larger m, we automatically work with *more* Bernstein polynomials of *higher* degrees. As a result, in our setting, *bern.pen* fits more flexible than *bern.unpen*.

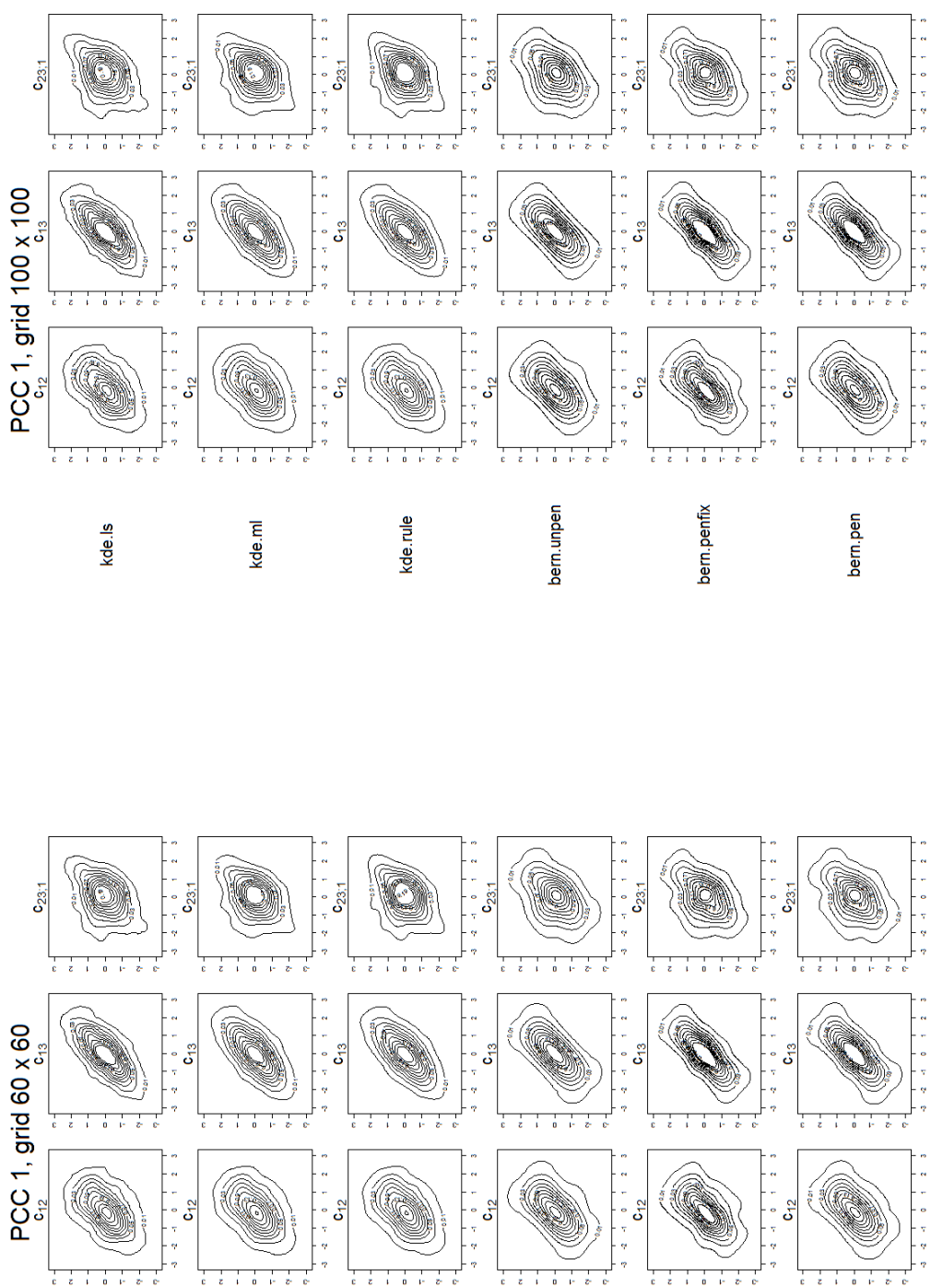

Figure 4.2: Contours of estimated pair-copula densities in a C-Vine with pairs 12, 13, 23;1 (1=Co, 2=Ti, 3=Sc). In each row, a different estimation method is applied. On the left hand side, the pair-copula densities are evaluated on a 60x60 grid, on the right hand side on a 100x100 grid.

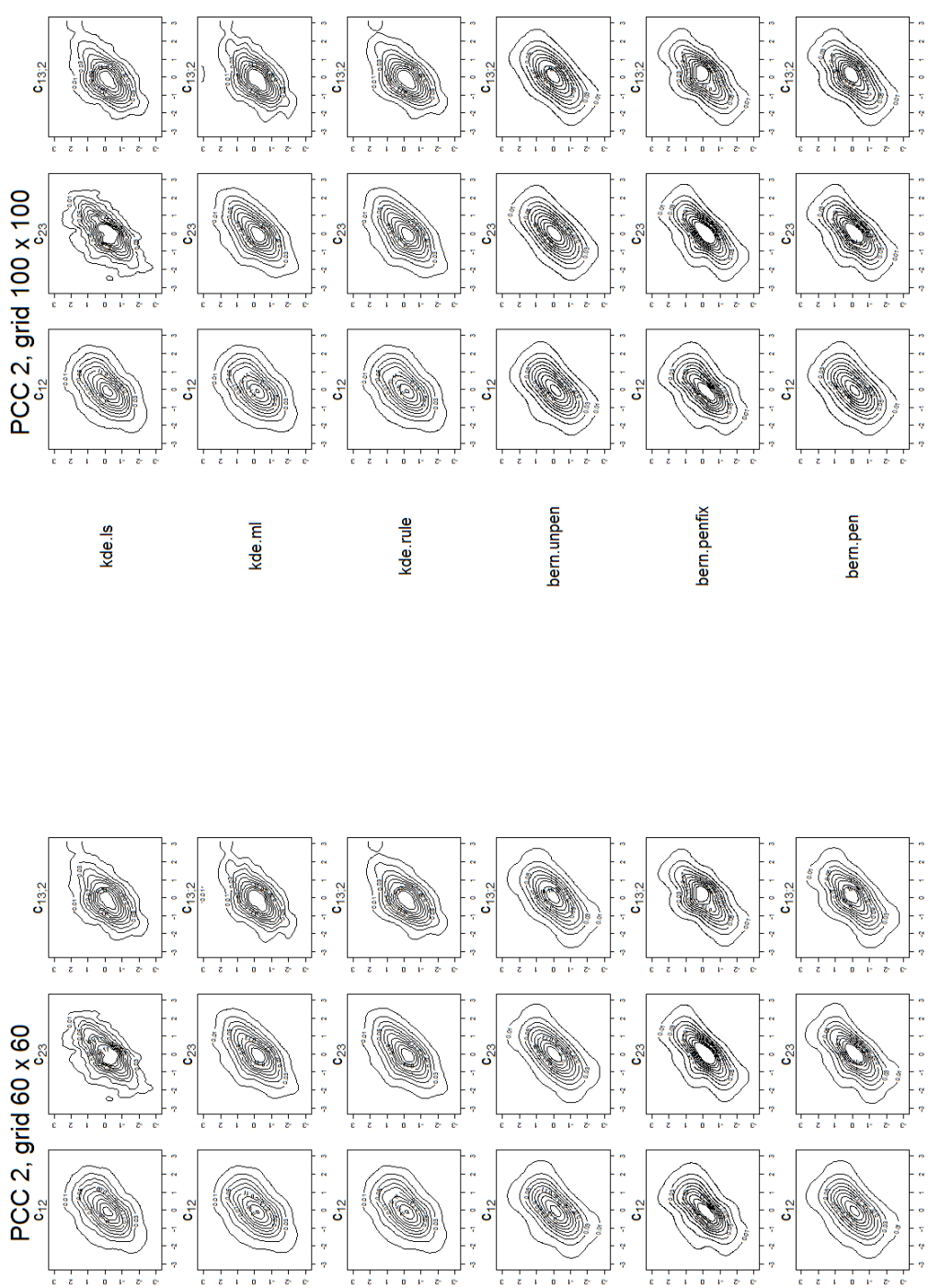

Figure 4.3: Contours of estimated pair-copula densities in a C-Vine with pairs 12, 23, 13;2 (1=Co, 2=Ti, 3=Sc). In each row, a different estimation method is applied. On the left hand side, the pair-copula densities are evaluated on a 60x60 grid, on the right hand side on a 100x100 grid.

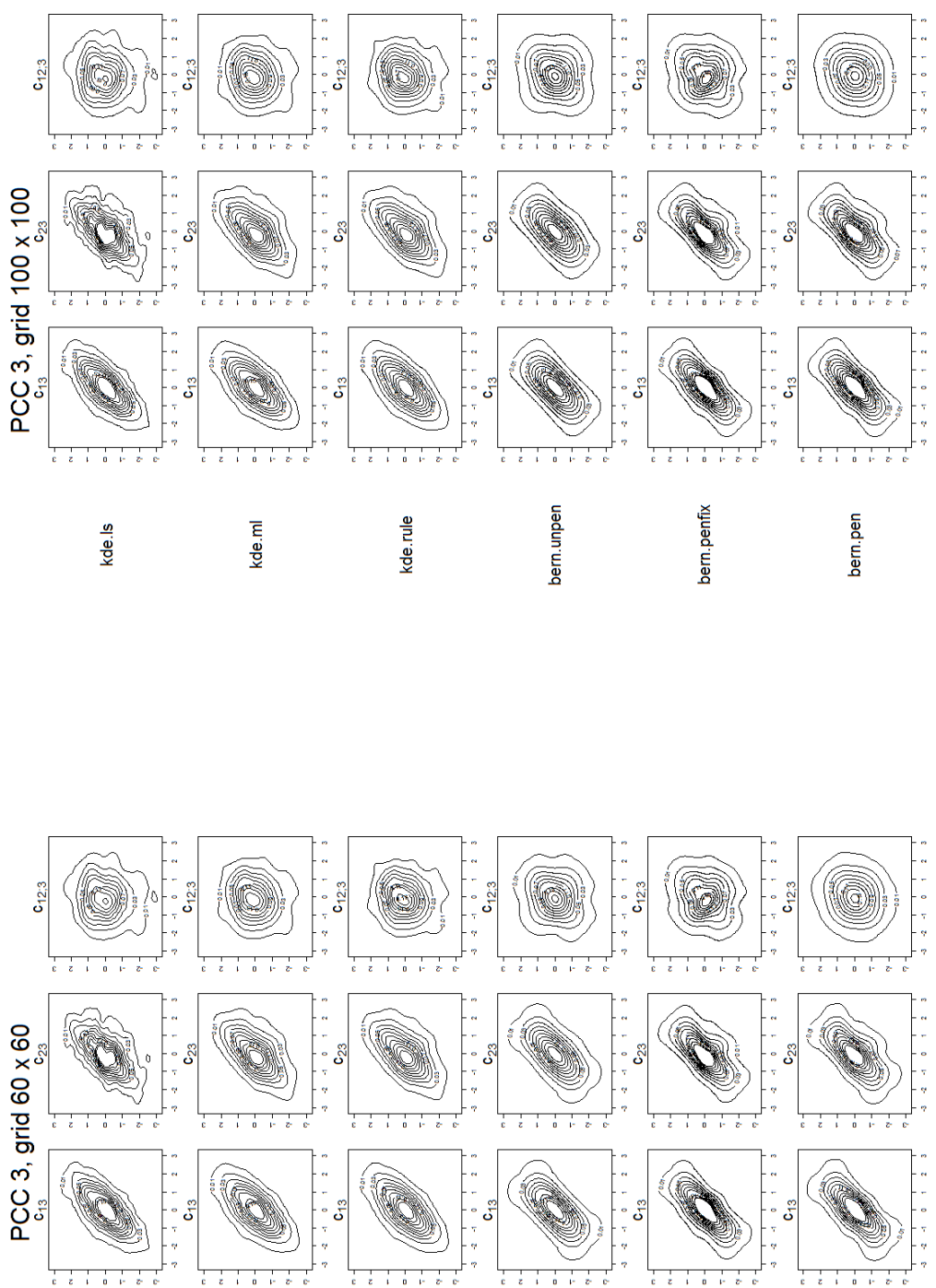

Figure 4.4: Contours of estimated pair-copula densities in a C-Vine with pairs 13, 23, 12;3 (1=Co, 2=Ti, 3=Sc). In each row, a different estimation method is applied. On the left hand side, the pair-copula densities are evaluated on a 60x60 grid, on the right hand side on a 100x100 grid.

		first contour fitting			second contour fitting		
		pair 1	pair 2	pair 3	pair 1	pair 2	pair 3
PCC 1	**kde.ls**	(0.2598, 0.3065)	(0.2597, 0.3185)	(0.2419, 0.2708)	(0.2419, 0.2708)	(0.3242, 0.2495)	(0.3227, 0.2494)
	kde.ml	(0.3453, 0.3387)	(0.3453 0.3387)	(0.3691 0.2800)	(0.3691 0.2800)	(0.2690 0.3896)	(0.2691 0.3894)
	kde.rule	(0.3368, 0.3425)	(0.3368 0.3425)	(0.3368 0.3365)	(0.3368 0.3365)	(0.3028 0.2899)	(0.3028 0.2899)
PCC 2	**kde.ls**	(0.3185, 0.2597)	(0.3185, 0.2597)	(0.1474, 0.2631)	(0.1474, 0.2631)	(0.2764, 0.2528)	(0.2764, 0.2528)
	kde.ml	(0.3387, 0.3453)	(0.3387, 0.3453)	(0.3093, 0.3403)	(0.3092, 0.3403)	(0.3192, 0.2029)	(0.3192, 0.2029)
	kde.rule	(0.3425, 0.3368)	(0.3425, 0.3368)	(0.3425, 0.3365)	(0.3425, 0.3365)	(0.3011, 0.3315)	(0.3011, 0.3315)
PCC 3	**kde.ls**	(0.2708, 0.2419)	(0.2708, 0.2419)	(0.2631, 0.1474)	(0.2631, 0.1474)	(0.3890, 0.2505)	(0.3890, 0.2505)
	kde.ml	(0.2800, 0.3691)	(0.2800, 0.3691)	(0.3403, 0.3092)	(0.3403, 0.3092)	(0.3882, 0.3226)	(0.3882, 0.3226)
	kde.rule	(0.3365, 0.3368)	(0.3365, 0.3368)	(0.3365, 0.3425)	(0.3365, 0.3425)	(0.3033, 0.3146)	(0.3033, 0.3146)

Table 4.4: Estimated bandwidths in the nonparametric C-Vine models.

		first contour fitting	second contour fitting
PCC 1	**bern.pen**	(23.9360, 4.7980, 8.2010)	(23.9360, 4.7980, 8.2010)
PCC 2	**bern.pen**	(23.9360, 7.4740, 7.6780)	(23.9360, 7.4740, 7.6780)
PCC 3	**bern.pen**	(4.7980, 7.4740, 77.8330)	(4.7980, 7.4740, 77.8330)

Table 4.5: Optimized penalty parameters in the nonparametric C-Vine models.

4.3 Comparison to C-Vine with Parametric Pair-Copulas

We focus on the C-Vine PCC 2, i. e. with pairs 12, 23, 13;2 and compare the kernel estimated vine with least-squares bandwidths and the Bernstein vine with fixed penalization to parametric vines.

	$\hat{C}_{12}$	$\hat{C}_{23}$	$\hat{C}_{13;2}$
copula family	t	Frank	t
parameter	0.64	3.87	0.63
degrees of freedom	5.89		5.98

Table 4.6: Model specification of the fitted parametric C-Vine model PCC 2.

	kde.ls	bern.pen.fix	par	kde.rule
log-likelihood	435.74	362.47	428.75	386.53

Table 4.7: Log-likelihoods of the nonparametric and parametric C-Vine models PCC 2.

We fit a parametric C-Vine using the *R* function *CDVineCopSelect* from package *CDVine*. We allow the following copula families: Gaussian, Student t, Clayton and its rotations, Gumbel and its rotations, Frank, Joe and its rotations. The optimal copula family is selected to be optimal with respect to AIC. The fitted parametric C-Vine specification is reported in Table 4.6. The fitted contours are given in Figure 4.5. We see that there is some slight asymmetry in the nonparametric contours which cannot be captured with parametric copula density contours. However, the parametric model log-likelihood is competetive to the nonparametric approaches, see Table 4.7.

4.4 Cross-Validation

We compare the fitted models by means of cross-validated log-likelihoods, see Section 2.3. To this end, we randomly choose a subset of the uranium copula data as test data, the rest of the data serves as training data. We fix the cardinality of the test data at 20 %. Then we fit our models to the training data and evaluate the log-likelihood of the fitted models on the test data. This gives us information about the out-of-sample predictive performance of the models. We do this $T = 50$ times for each model and compare the respective means and medians (because of outliers) of the obtained test log-likelihoods for each model. We interprete the model with the highest cross-validated log-likelihood as the model with highest prediction quality.

The total number of observations of uranium copula data is $n = 655$. Let N be the full data, i. e. n observations of dimension $d = 3$. We assume a test data size of 20 %, that is $n^{test} = 0.2 \cdot n = 131$ and $n^{train} = 0.8 \cdot n = 524$. In 50 runs $t = 1, \ldots, 50$, we randomly choose the test data $N^{test,t} \subset N$ and build the training data $N^{train,t} = N \setminus N^{test,t}$. On $N^{train,t}$ we fit 21

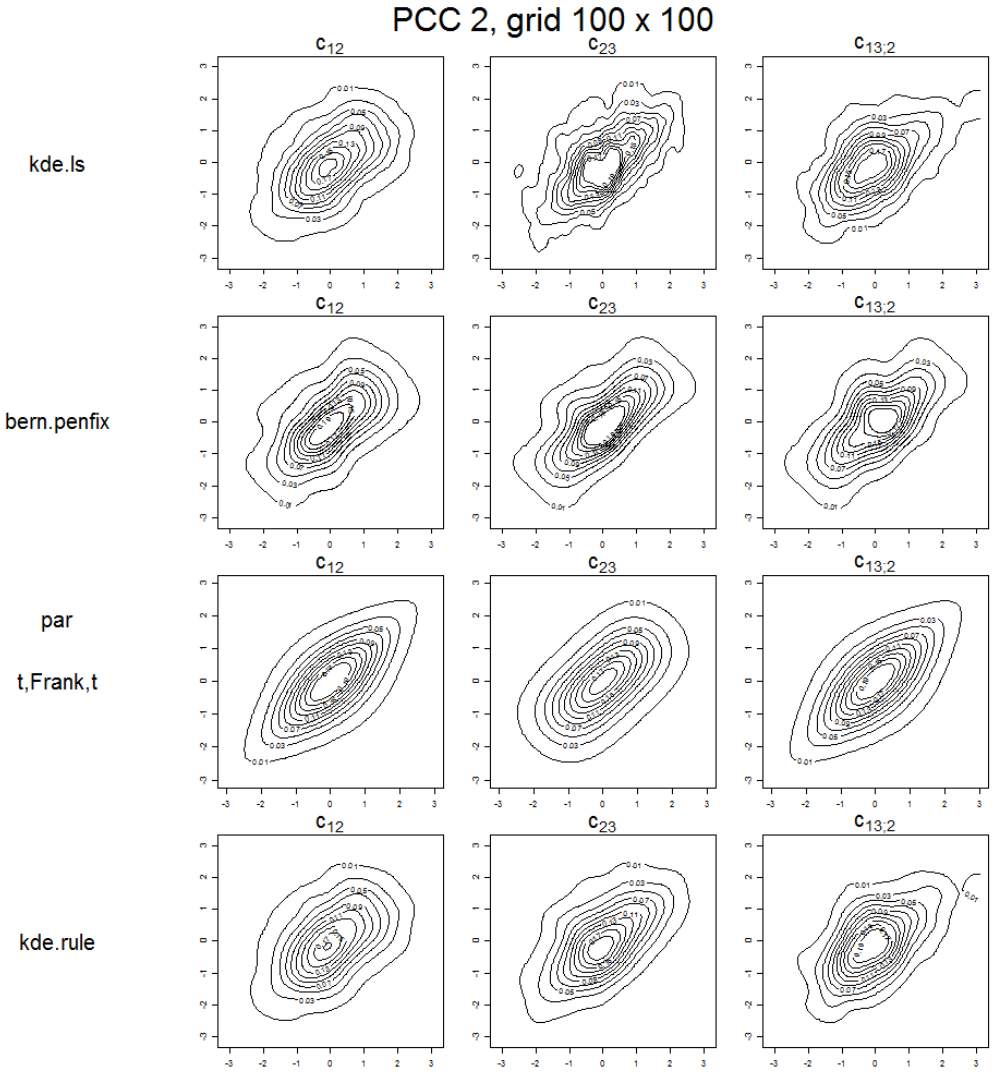

Figure 4.5: Contours of estimated pair-copula densities in a C-Vine with pairs 12, 23, 13;2 (1=Co, 2=Ti, 3=Sc). We compare nonparametric and parametric estimation techniques.

C-Vine models: for each of the three C-Vine decompositions PCC 1, PCC 2, PCC 3 (see Table 4.2) we apply six nonparametric and one parametric estimation method (see Table 4.3). On $N^{test,t}$ we evaluate the log-likelihood of the copula density for each PCC and each estimation method separately, that is

$$L^{test,t}_{M,PCCi} = \sum_{\hat{\mathbf{u}} \in N^{test,t}} \ln c^{M}_{PCCi}\left(\hat{\mathbf{u}}; \hat{\boldsymbol{\Theta}}^{train}\right),$$

where $M \in \{kde.ls, kde.ml, kde.rule, bern.unpen, bern.penfix, bern.pen\}$ and $i = 1, 2, 3$. Furthermore,

$$\begin{aligned} \hat{\boldsymbol{\Theta}}^{train} &:= \left(\hat{h}_1, \hat{h}_2, \hat{\mathbf{u}} \in N^{train,t}\right) \quad \text{for } M \in \{kde.ls, kde.ml, kde.rule\} \quad \text{and} \\ \hat{\boldsymbol{\Theta}}^{train} &:= \hat{\boldsymbol{\alpha}} \in [0,1]^{(m+1)^2} \quad \text{for } M \in \{bern.unpen, bern.penfix, bern.pen\}. \end{aligned}$$

In Figure 4.6, we see in each of the three rows for PCC 1, 2 and 3 the distribution of the test log-likelihoods per estimation method. We see for *kde.ls* some extreme outliers. However, even when excluding those test runs t with $L^{test,t}_{kde.ls,PCCi} > 1000$, *kde.ls* still gives the highest mean and median, but also the highest variation of the test log-likelihoods. The Bernstein approaches behave comparably among themselves. In PCC 1 and PCC 3, *kde.ml* and *kde.rule* are quite similar, but for PCC 2, *kde.ml* performs better. In the Figures 4.7 and 4.8, we display the empirical densities of the test log-likelihoods. In Figure 4.7 we compare the densities per PCC, whereas in Figure 4.8 we compare the densities per estimation method. Figure 4.7 yields a similar interpretation as above for Figure 4.6. From Figure 4.8 we further get the information that for *kde.ls*, *kde.ml* and *kde.rule*, PCC 2 is preferred, whereas *bern.unpen*, *bern.penfix* and *bern.pen* prefer PCC 3.
In Table 4.8 we summarize test log-likelihoods for each of the 21 models. The cross-validated log-likelihood is usually defined as the mean of $L^{test,t}_{M}, t = 1, \ldots, 50$ for a model M. We further show the respective median, because it is more robust than the mean and we experience some extreme outliers with *kde.ls*. In Table 4.8, the standard deviation is also effected by the outliers with *kde.ls*, so we also show the robust estimate of the standard deviation as proposed by Silverman [1986] as $\hat{\sigma} = min\left(s, \frac{r}{1.349}\right)$ (see Equation (2.13)), where s denotes the sample standard deviation and r denotes the interquartile range.
The parametric model performs at least as good as the Bernstein approaches, in PCC 1 and 2 even better. But compared to the kernel density approaches, the parametric model shows significantly smaller test log-likelihoods.

All in all we can say that for all C-Vine decompositions, the kernel estimation methods are preferred over the Bernstein copula methods with respect to cross-validated log-likelihoods. On the other hand, the Bernstein copula methods have the advantage to return smooth denisities and the estimation procedure is computationally less expensive than the kernel estimation methods. But the Bernstein copula methods perform even suboptimal compared the parametric approach when considering the cross-validated log-likelihoods. We suggest the advantage of estimation with Bernstein polynomials compared to parametric estimation is that asymmetric dependencies are captured better, but the disadvantage is the behavior in the tails.

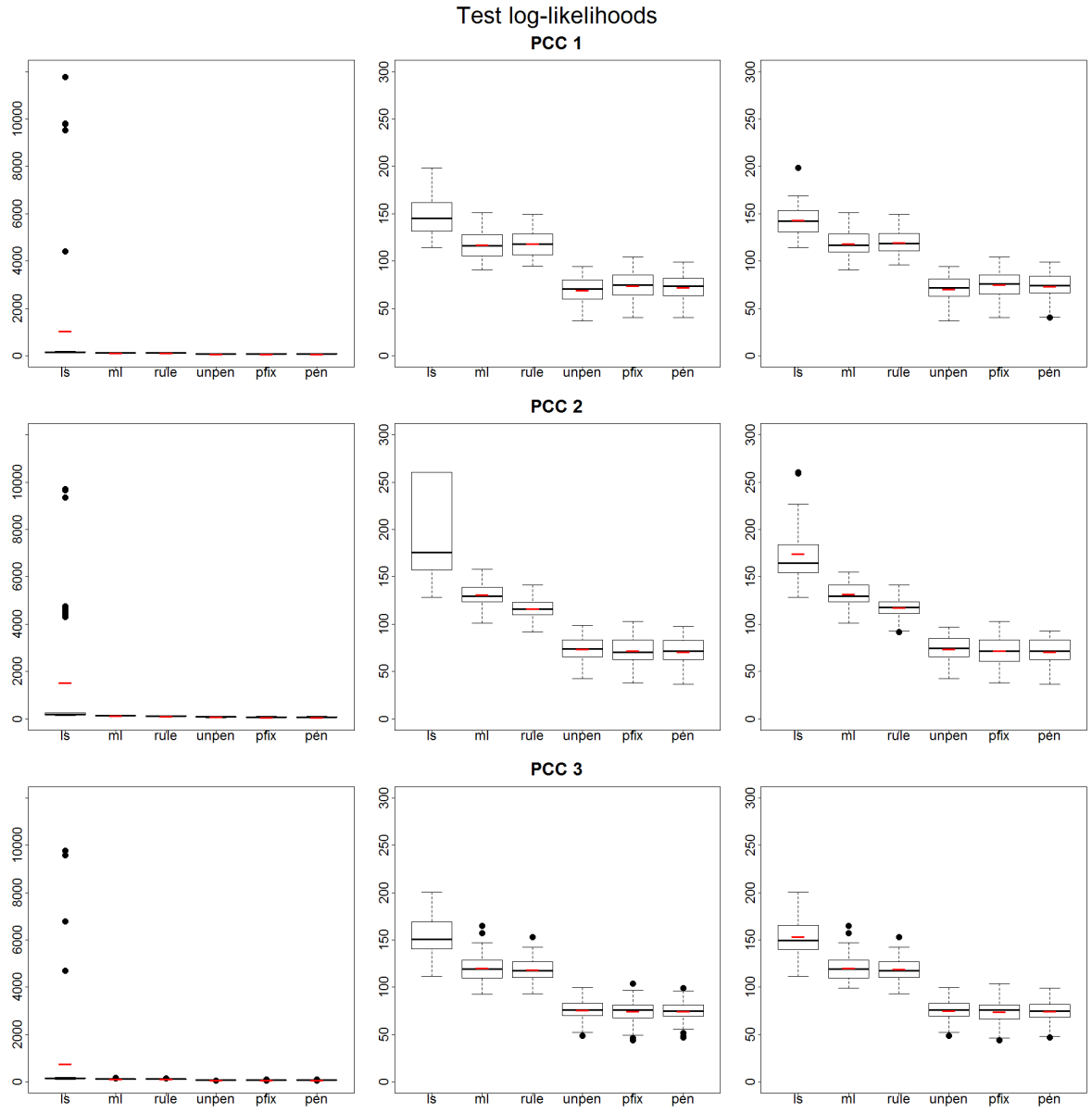

Figure 4.6: Boxplots of cross-validated log-likelihoods for nonparametric fitted C-Vines. For each PCC and each estimation method we fitted parameters for 50 randomly drawn subsamples and calculated the log-likelihoods on the remaining test data. On the left hand side, we see some extreme outliers for the kernel estimation with least-squares bandwidths. The middle column plots are a cut of the left column plots, where we display values of maximal 300. For the right hand column boxplots we excluded those test runs, where the least-squares methods resulted in test log-likelihoods higher than 1000. The red lines show the respective means of the 50 log-likelihoods. (ls=kde.ls, ml=kde.ml, rule=kde.rule, unpen=bern.unpen, pfix=bern.penfix, pen=bern.pen).

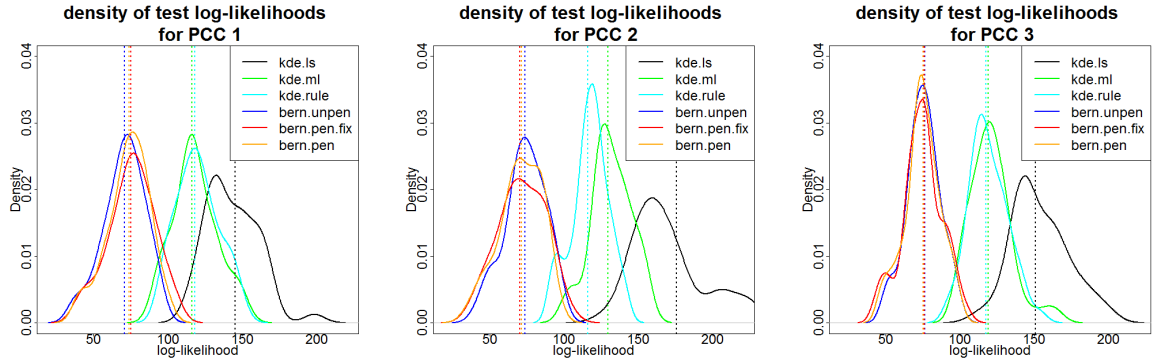

Figure 4.7: Density plots of cross-validated log-likelihoods. For each C-Vine order and each estimation method we fitted parameters for 50 randomly drawn subsamples and calculated the log-likelihoods on the remaining test data. We excluded those test runs, where the least-squares methods resulted in test log-likelihoods higher than 1000. The dotted lines show the respective medians of the 50 log-likelihoods.

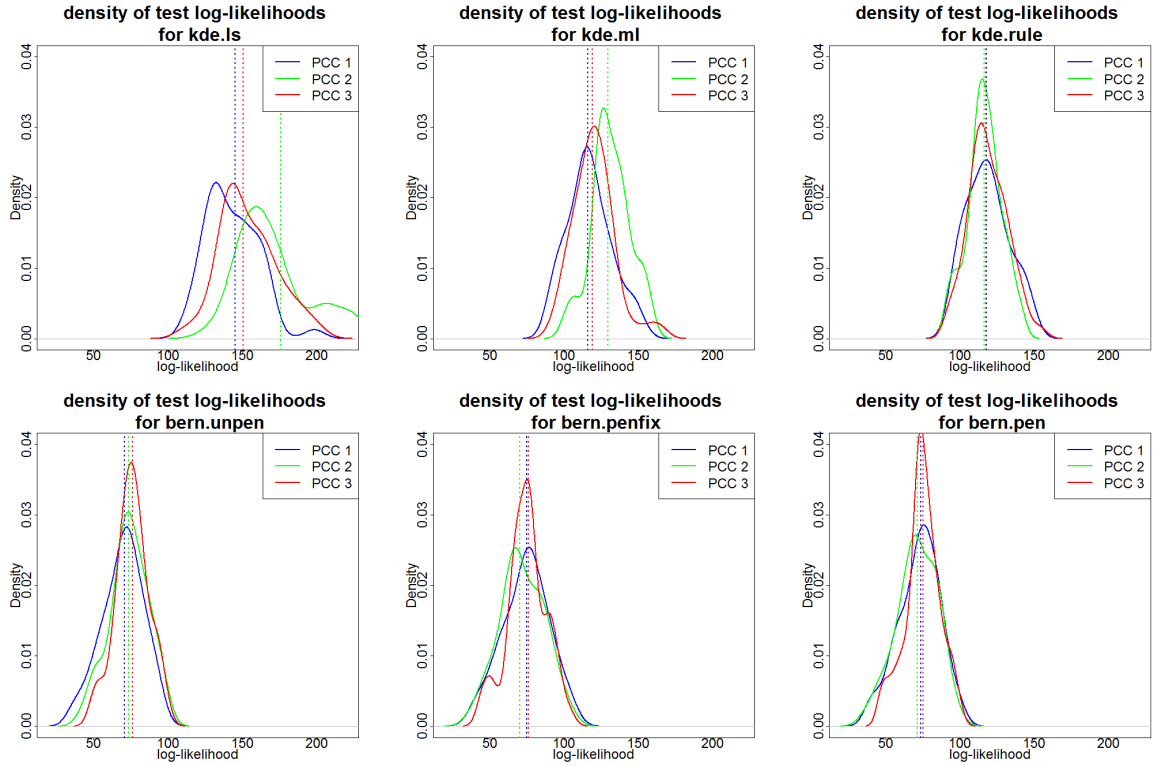

Figure 4.8: Density plots of cross-validated log-likelihoods per estimation method. For kde.ls, we excluded those test runs, where the resulting test log-likelihoods are higher than 1000. The dotted lines show the respective medians of the 50 log-likelihoods.

		Minimum	Mean	Median	Maximum	Std. Dev.	Robust Dev.
	kde.ls	114.19	1034.69	145.15	11777.49	2814.75	22.39
	kde.ml	90.55	117.05	115.88	151.29	14.88	14.88
	kde.rule	94.97	118.49	117.84	149.24	14.35	14.35
PCC 1	**bern.unpen**	36.65	69.53	70.62	93.93	13.87	13.87
	bern.penfix	40.18	74.23	74.84	104.05	15.48	15.40
	bern.pen	40.17	72.40	73.65	99.10	13.96	13.00
	par	46.72	80.05	79.25	111.55	14.21	12.22
	kde.ls	128.05	1519.26	175.52	9705.92	2650.75	75.96
	kde.ml	100.79	131.35	129.54	157.90	12.98	11.56
	kde.rule	91.57	116.40	116.13	141.14	11.58	9.41
PCC 2	**bern.unpen**	42.17	73.84	73.64	98.50	13.07	12.72
	bern.penfix	37.82	71.87	70.22	102.98	14.94	14.94
	bern.pen	36.40	70.91	71.24	97.70	13.68	13.68
	par	53.69	81.11	81.05	105.21	12.80	12.47
	kde.ls	111.41	757.56	150.60	9766.05	2153.85	20.59
	kde.ml	92.62	120.18	119.02	164.86	14.06	13.64
	kde.rule	93.40	118.84	117.18	153.15	13.00	12.19
PCC 3	**bern.unpen**	48.77	75.74	76.23	99.39	11.20	9.65
	bern.penfix	44.17	74.77	75.85	103.59	13.12	9.73
	bern.pen	47.03	74.63	75.06	99.29	11.74	8.47
	par	42.48	76.03	75.66	104.72	14.52	14.52

Table 4.8: Summary of cross-validated log-likelihoods of the nonparametric and parametric C-Vine models.

Chapter 5

Simulation Study

We define three scenarios of a three-dimensional C-Vine copula and simulate for each scenario 1000 observations. The settings for scenario 1 and 2 are taken from Mai and Scherer [2012] with some modified parameter. The third scenario is simulated from Tawn copulas. The scenarios are displayed in Table 5.2. We adopt the notation from Tables 4.2 and 4.3.

The C-Vine structure, from which we simulate, is PCC 2, i. e. it has the pairs 12, 23, 13;2. Then we will fit C-Vines with structure PCC 1, i. e. with pairs 12, 13, 23;1. The intention is to compare the fitted pair-copula density contours $\hat{c}^{kde.ls}, \hat{c}^{bern.penfix}, \hat{c}^{par}$ and $\hat{c}^{kde.rule}$ to the simulated pair-copula density $\hat{c}^{sim}$.

The Tawn copula (see Tawn [1988]) listed in Table 5.2 is an asymmetric extreme value copula. The Gumbel copula can be obtained from the Tawn copula with special parameter selection. The Tawn copula is used in the simulation with the following dependence function.

$$A(t) = 1 - \psi_2 + (\psi_2 - \psi_1)\, t + \left[\psi_1^{\theta} t^{\theta} + \psi_2^{\theta} (1-t)^{\theta}\right]^{1/\theta}$$

with parameter vector (θ, ψ_1, ψ_2) where $\psi_1 \geq 0, \psi_2 \leq 1$ and $\theta \geq 1$.

We simulate PCC 2 for three scenarios according to Table 5.2 and fit PCC 1 with nonparametric and parametric estimation techniques. We allow the following copula families: Gaussian, Student t, Clayton and its rotations, Gumbel and its rotations, Frank, Joe and its rotations. The optimal copula family is selected to be optimal with respect to AIC. We look at the pair-copula densities c_{12} and c_{13}, see Figure 5.1. We are especially interested in c_{13} as it shows asymmetric contours and we are interested in seeing how the known copula density c_{12} (see Table 5.2) is estimated by the nonparametric methods. For scenario 1, 2 and 3 we compare the the *true* pair-copula density contours to the nonparametrically and parametrically estimated contours in Figure 5.2, 5.3 and 5.4. Since the true c_{13} is unknown, as it is not explicitly given in the simulation of PCC 2, we have to estimate

	PCC	First Tree	Pair-Copula Densities	Copula Specification
simulated	PCC 2	1-2-3	$c_{12}, c_{23}, c_{13;2}$	cf. Table 5.2
fitted	PCC 1	2-1-3	$c_{12}, c_{13}, c_{23;1}$	kde.ls, bern.penfix, par, kde.rule

Table 5.1: Overview of simulated and fitted pair-copula constructions.

	C_{12}	par	τ	C_{23}	par	τ	$C_{13;2}$	par	τ
1	G	5.00	0.80	C	0.70	0.26	C	0.70	0.26
2	t	(0.80, 2.10)	0.59	G	1.75	0.43	t	(-0.95, 2.50)	-0.80
3	T	(2.00, 1.00, 0.50)	0.31	T	(2.00, 1.00, 0.50)	0.31	T	(2.00, 1.00, 0.50)	0.31

Table 5.2: Simulation scenario settings. t: Student-t C: Clayton G: Gumbel, T: Tawn.

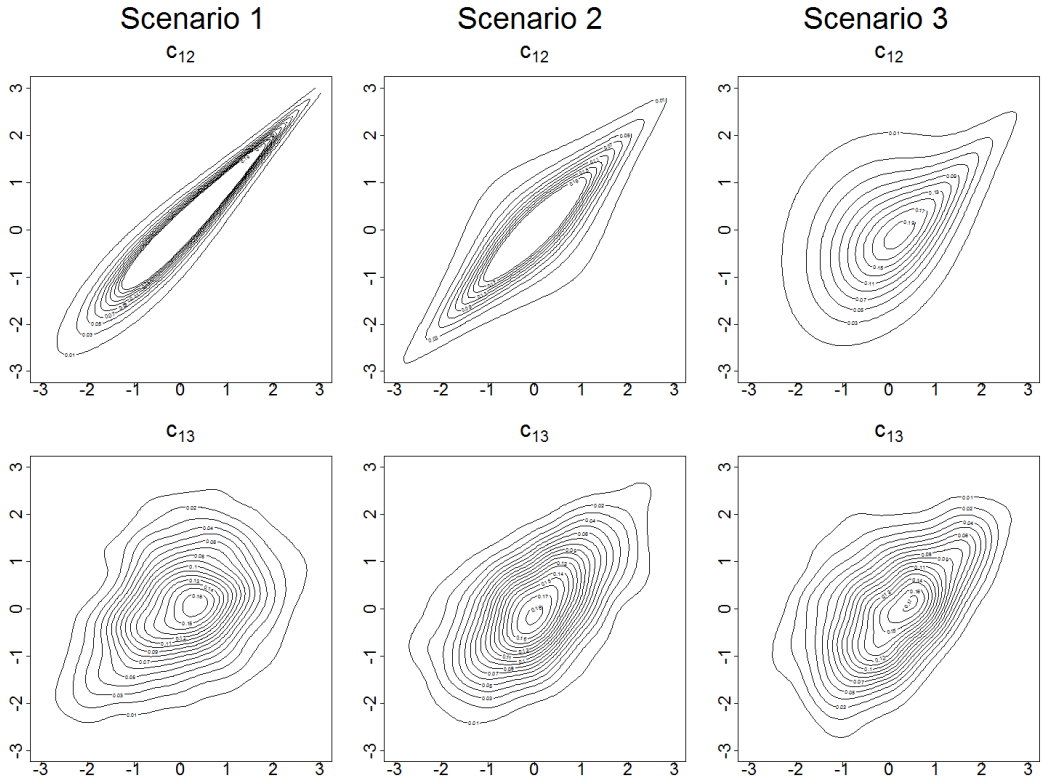

Figure 5.1: True contours of the pair-copula density c_{12} and estimated contours of the pair-copula density c_{13}. The estimation is performed by kernel density estimation with Gaussian kernel functions and bandwidths by rule of thumb.

the *true* contour. We decide to apply kernel density estimation with Gaussian kernels and rule of thumb bandwidths.

For scenario 1, the true contours are captured best by ***kde.ls*** and worst by ***bern.penfix***. ***bern.penfix*** does not describe the tails correctly throughout all scenarios. However, it captures the asymmetry. The contrary is true for the parametric estimation. It captures the tail behavior, but no asymmetric behavior. ***kde.ls*** is, as usual, quite wiggly, especially in the Tawn copula scenario 3.

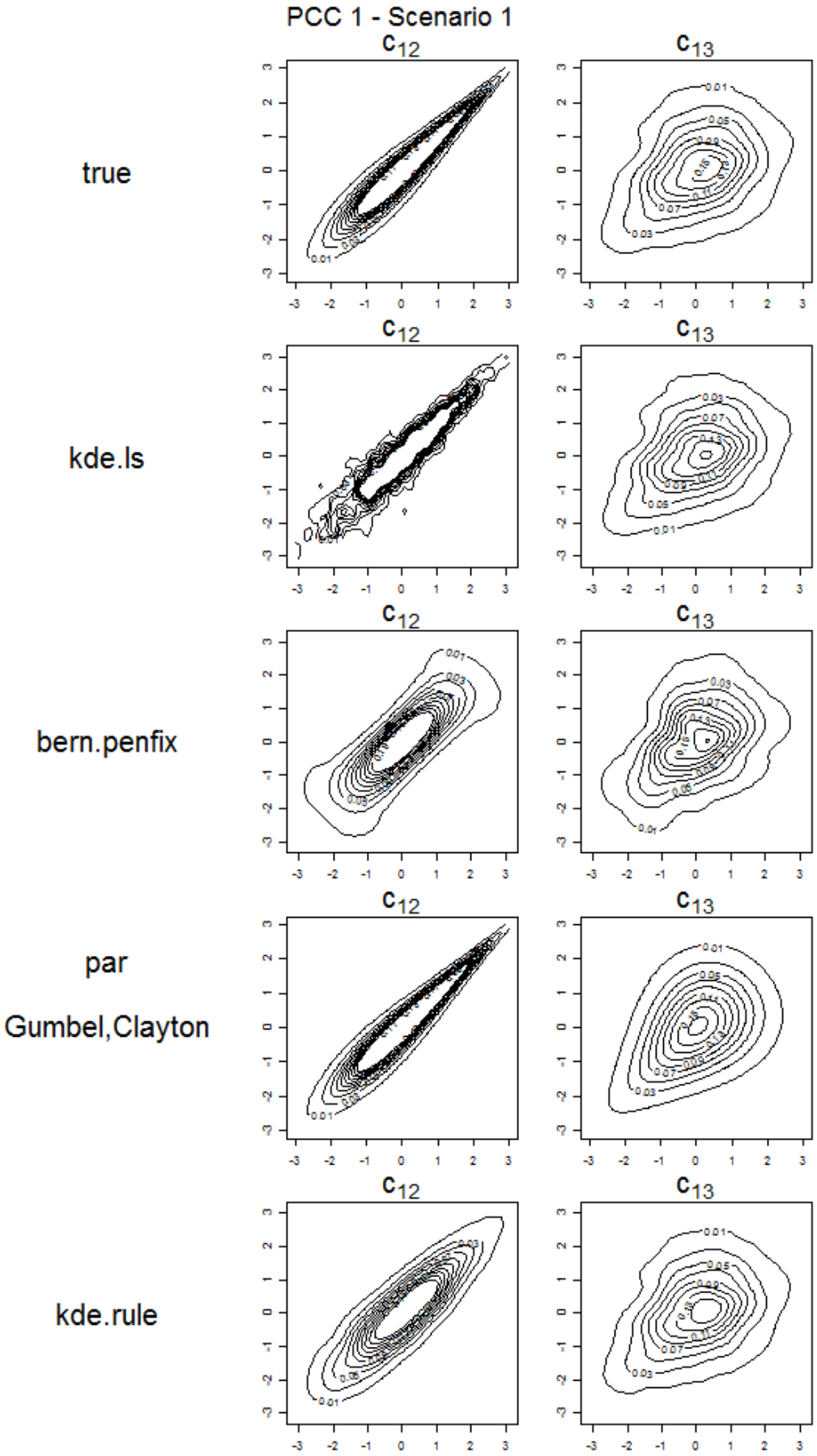

Figure 5.2: Contours of estimated pair-copula densities in PCC 1. In the second to fourth row of the Figure, different estimation methods are applied. We can compare them to the true densities in the first row.

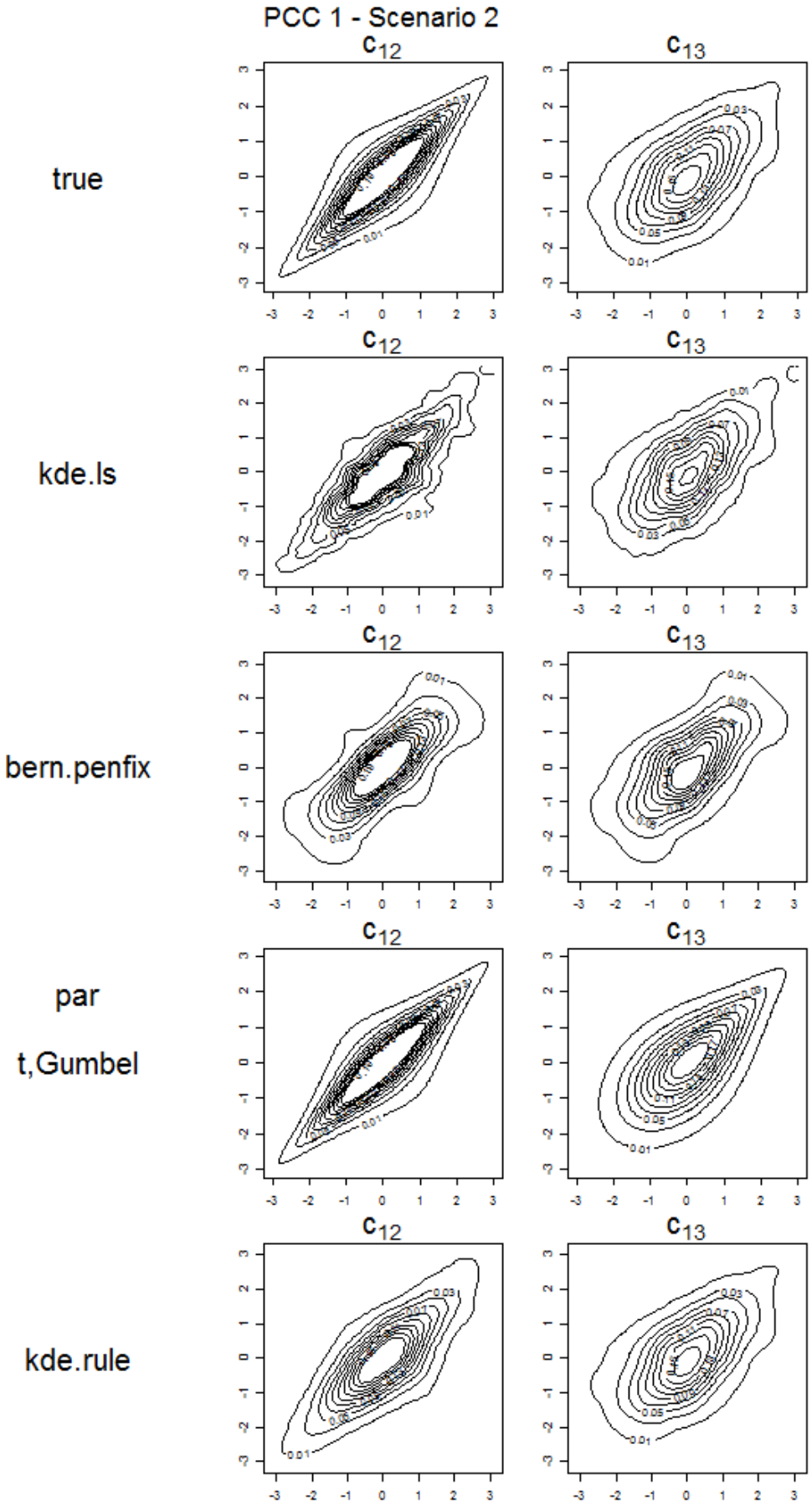

Figure 5.3: Contours of estimated pair-copula densities in PCC 1 for Scenario 2. In the second to fourth row of the Figure, different estimation methods are applied. We can compare them to the true densities in the first row.

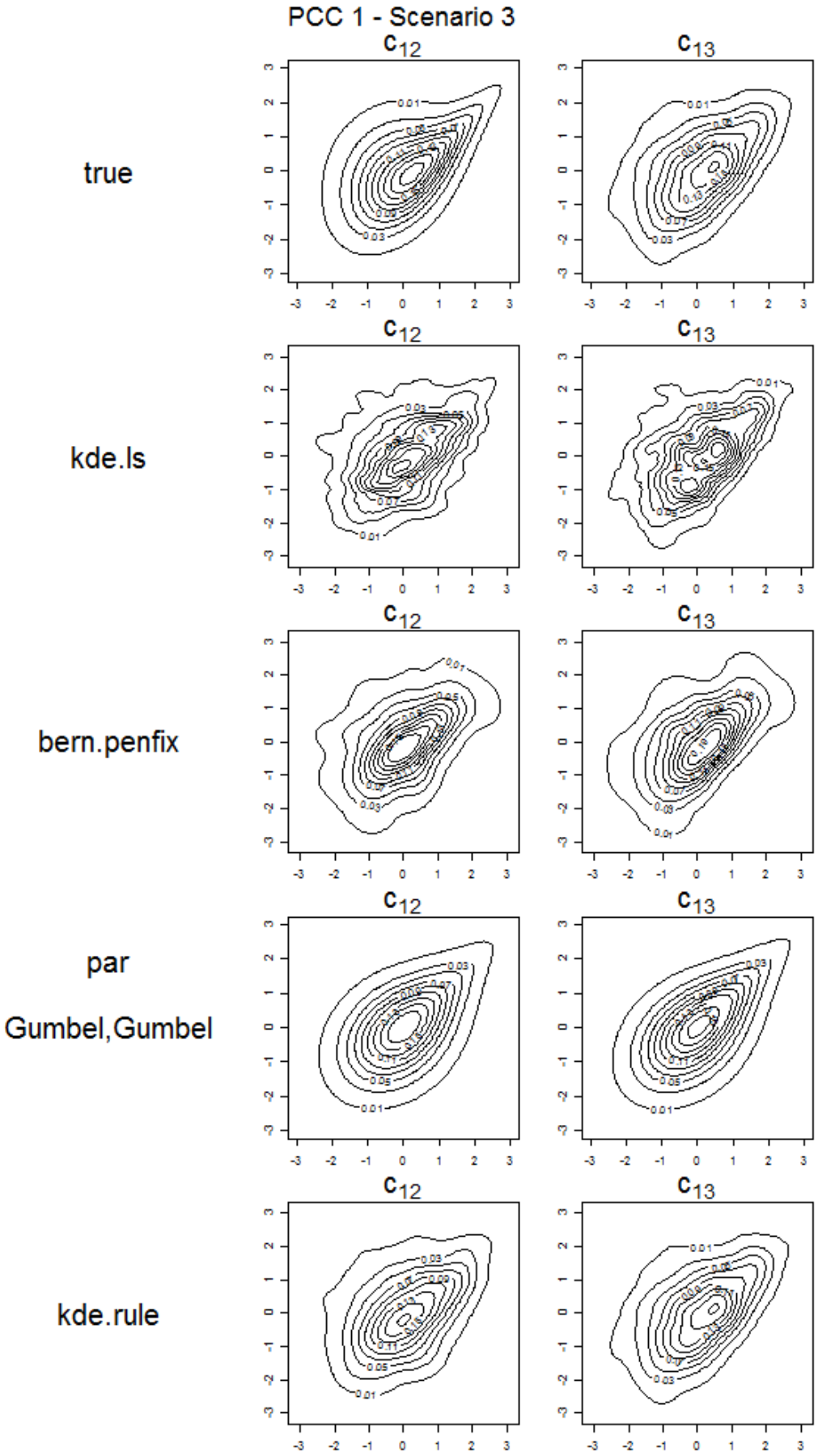

Figure 5.4: Contours of estimated pair-copula densities in PCC 1 for Scenario 3. In the second to fourth row of the Figure, different estimation methods are applied. We can compare them to the true densities in the first row.

Chapter 6

Conclusion

In this thesis, we intended to review nonparametric estimation methods of pair-copula constructions. We focused on the method of kernel density estimation with fixed bandwidths and Bernstein estimation with and without penalization.

For the bandwidth selection in the kernel density estimation, we presented least-squares cross-validated bandwidths, maximum likelihood cross-validated bandwidths and normal reference rule of thumb bandwidths. In an empirical study, we fitted three-dimensional C-Vine copulas to the uranium water samples data, restricted to three chemicals. We experienced that the kernel estimation with least-squares cross-validated bandwidths tends to undersmooth the density. In general, the kernel estimated densities are less smooth compared to the Bernstein estimated densities. With Bernstein estimation, there is a lack of flexibility in the tails. However, the nonparametric methods express a present asymmetric dependence in the data better than parametric models. This overall picture was validated in a simulation study with three scenarios.
We compared the fitted models by means of likelihood cross-validation. From this point of view, the kernel estimation performed far better than the parametric estimation, and this was still slightly better than the Bernstein estimation.

Based on the results of this thesis, investigations could be extended regarding the smoothness of the kernel density estimation approach with least-squares cross-validated bandwidths, maybe by imposing constraints or introducing a kind of penalization. The usage of Bernstein estimation could be improved by an `R` implementation for vine copula estimation with Bernstein polynomials of arbitrary degree without penalization. If the Bernstein estimation becomes competitive, one could consider vine estimation with differently estimated building blocks, for example one pair-copula is estimated by kernel density estimation, the next pair-copula is estimated by Bernstein polynomials and a further pair-copula is estimated parametrically. General statistical concepts could be adopted to the nonparametric copula estimation context, for example hypothesis testing and confidence intervals. One could explore the model performances with respect to other criteria such as AIC. Here, the number of model parameters plays a role next to the likelihood of the model. Furthermore, data simulation from nonparametric vines could be of interest.

Bibliography

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance Mathematics and Economics*, 44(2):182–198, 2009.

Karim M. Abadir and Steve Lawford. Optimal asymmetric kernels. *Economics Letters*, 83(1):61, 2004. URL `http://ssrn.com/abstract=1985504`.

Elif F. Acar, Christian Genest, and Johanna Nešlehová. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90, 2012.

T. Bedford and R. Cooke. Probabilistic density decomposition for conditionally dependent random variables modeled by vines. *Annals of mathematics and Artificial Intelligence*, 32:245–268, 2001.

T. Bedford and R. Cooke. Vines: A new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

Adrian W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):pp. 353–360, 1984. ISSN 00063444. URL `http://www.jstor.org/stable/2336252`.

Theophilos Cacoullos. Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18:179–189, 1966. ISSN 0020-3157. URL `http://dx.doi.org/10.1007/BF02869528`.

Y.-S. Chow, S. Geman, and L.-D. Wu. Consistent cross-validated density estimation. *The Annals of Statistics*, 1:25–38, 1983.

R. D. Cook and M. E. Johnson. Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. *Technometrics*, 28:123 – 131, 1986.

Paul Deheuvels. La fonction de dépendance empirique et ses propriétés. *Bulletin de la Classe des Sciences, Académie Royale de Belgique*, 65:274 – 292, 1979.

Dorothea Diers, Martin Eling, and Sebastian D. Marek. Dependence modeling in non-life insurance using the Bernstein copula. *Insurance: Mathematics and Economics*, 50 (3):430 – 436, 2012. ISSN 0167-6687. doi: 10.1016/j.insmatheco.2012.02.007. URL `http://www.sciencedirect.com/science/article/pii/S0167668712000236`.

J. Dissmann, E. C. Brechmann, C. Czado, and D. Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69, 2013.

E.H. Doha, A.H. Bhrawy, and M.A. Saker. On the derivatives of Bernstein polynomials: an application for the solution of high even-order differential equations. *Boundary Value Problems*, 2011. doi: 10.1155/2011/829543.

R. P W Duin. On the choice of smoothing parameters for Parzen estimators of probability density functions. *Computers, IEEE Transactions on*, C-25(11):1175–1179, 1976. ISSN 0018-9340. doi: 10.1109/TC.1976.1674577.

P. H. C. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.

V.K. Epanechnikov. Non-parametric estimation of a multivariate probabitity density. *Theory of Probability & Its Applications*, 14:153–158, 1969.

E. Fix and J.L. Hodges. *Nonparametric Discrimination: Consistency Properties*. USAF School of Aviation Medicine, 1951.

Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, pages 347–368, 2007.

Christian Genest, Esterina Masiello, and Karine Tribouley. Estimating copula densities through wavelets. *Insurance: Mathematics and Economics*, 44(2):170 – 181, 2009. ISSN 0167-6687. URL `http://www.sciencedirect.com/science/article/pii/S0167668708000978`.

P. Hall. On the use of compactly supported density estimates in problems of discrimination. *Journal of Multivariate Analysis*, 23:131–158, 1987.

Tristen Hayfield and Jeffrey S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5), 2008. URL `http://www.jstatsoft.org/v27/i05/`.

Daniel J. Henderson and Christopher F. Parmeter. Normal reference bandwidths for the general order, multivariate kernel density derivative estimator. Technical Report 2011-15, University of Miami, Department of Economics, 2011. URL `http://ideas.repec.org/p/mia/wpaper/2011-15.html`.

I. Hobaek Haff and J. Segers. Nonparametric estimation of pair-copula constructions with the empirical pair-copula. *ArXiv e-prints*, 2012.

Marius Hofert, Ivan Kojadinovic, Martin Maechler, and Jun Yan. *copula: Multivariate Dependence with Copulas*, 2012. URL `http://CRAN.R-project.org/package=copula`. R package version 0.999-5.

H. Joe. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In Ludger Rüschendorf, Berthold Schweizer, and Michael D. Taylor, editors, *IMS Lecture Notes - Monograph Series*, volume 28, chapter Distributions with Fixed Marginals and Related Topics, pages 120–141. Institute of Mathematical Statistics, Hayward, CA, 1996.

O. Kolbjørnsen and M. Stien. D-vine creation of non-gaussian random field. In *In Proceedings of the Eighth International Geostatistics Congress*, pages 399–408. GECAMIN Ltd., 2008.

D. Kurowicka and R. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, New York, 2006.

D. Kurowicka and H. Joe. *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing Co., Singapore, 2011.

Qi Li and Jeffrey S. Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 2007.

D. Lopez-Paz, J. M. Hérnandez-Lobato, and B. Schölkopf. Semi-supervised domain adaptation with non-parametric copulas, 2012.

Jan-Frederik Mai and Matthias Scherer. *Simulating Copulas*. World Scientific, 2012. URL `http://www.worldscientific.com/doi/abs/10.1142/9781848168756-0005`.

J. S. Marron and D. Ruppert. Transformations to reduce boundary bias in kernel density estimation, 1993.

B. Marx and P. H. C. Eilers. Multidimensional penalized signal regression. *Technometrics*, 47:13–22, 2005.

Roger B. Nelsen. *An Introduction to Copulas*. Springer, 2006.

D. Pfeifer, D. Strassburger, and J. Philipps. Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas, 2009.

M.J.D. Powell. Uobyqa: unconstrained optimization by quadratic approximation, 2000.

Mats Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9(2):pp. 65–78, 1982. ISSN 03036898. URL `http://www.jstor.org/stable/4615859`.

Alessio Sancetta and Stephen Satchell. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20 (03):535–562, 2004. URL `http://dx.doi.org/10.1017/S026646660420305X`.

O. Scaillet and J.-D. Fermanian. Nonparametric estimation of copulas for time series. *FAME Research Paper*, 57, 2002. doi: http://dx.doi.org/10.2139/ssrn.372142. URL `http://ssrn.com/abstract=372142`.

Christian Schellhase. *Density and Copula Estimation using Penalized Spline Smoothing*. PhD thesis, University Bielefeld, 2012.

Eugene F. Schuster and Gavin G. Gregory. On the nonconsistency of maximum likelihood nonparametric density estimators. In *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, pages 295–298. Springer US, 1981. ISBN 978-0-387-90633-1.

Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, 1980.

X. Shen, Y. Zhu, and L. Song. Linear B-spline copulas with applications to nonparametric estimation of copulas. *Computational Statistics and Data Analysis*, 52:3806–3819, 2008.

B. W. Silverman. *Density estimation for statistics and data analysis / B.W. Silverman*. Chapman and Hall, London, 1986.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.

Padhraic Smyth. Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing*, 9:63–72, 2000.

Jonathan A. Tawn. Bivariate extreme value theory: Models and estimation. *Biometrika*, 75(3):397–415, 1988.

G. Weiß and M. Scheffer. Smooth Nonparametric Bernstein Vine Copulas. *ArXiv e-prints*, 2012.

Walter Zucchini. Applied smoothing techniques, 2003.