Nonparametric Predictive Inference for Multivariate Data using Copulas

TAGHREED ALMASOUD

Nonparametric Predictive Inference for Multivariate Data using Copulas

Taghreed A. Almasoud

A Thesis presented for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
University of Durham
England

June 2025

Dedicated to

My lovely family

Nonparametric Predictive Inference for Multivariate Data using Copulas

Taghreed A. Almasoud

Submitted for the degree of Doctor of Philosophy
June 2025

Abstract

Modelling dependence among random quantities is a core aspect of multivariate data analysis. Copulas provide a flexible and powerful approach to capturing the dependence structure between random quantities. Several dependence models have been proposed in the literature, including classical copulas, vine copulas, and fully nested Archimedean copulas (FNAC). In parallel, several statistical methods have been developed within the imprecise probability framework, including nonparametric predictive inference (NPI). NPI is based on minimal modelling assumptions and quantifies uncertainty using lower and upper probabilities. Recently, NPI has been applied to bivariate data using both parametric and nonparametric copulas.

This thesis contributes to the use of NPI for multivariate data by presenting different approaches for prediction. The focus is on copulas for modelling dependence, as they provide high flexibility for modelling complex dependency patterns. A generalization is proposed for the method that combines NPI with bivariate data, using a parametric copula with a single parameter to model dependence. The approach is further extended by introducing a fully nonparametric version that uses a nonparametric copula. A novel method for combining NPI with vine copulas is also presented, motivated by the vine copulas ability to capture several dependence structures in a model. In addition, a new method integrating NPI with FNAC is developed, where FNAC is a promising model for capturing different dependencies within a model using Archimedean copulas. The proposed methods are illustrated using examples from the literature. Simulation studies are conducted to evaluate the predictive performance of the proposed methods and to

compare the methods, highlighting their strengths and differences. The results indicate that the methods with either vine copulas or FNAC perform well compared to other methods.

Declaration

The work in this thesis is based on research carried out in the Department of Mathematical Sciences at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification, and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

My deep gratitude goes to Allah, who helped me in accomplishing this thesis.

I would like to express my sincere thanks to my supervisors, Prof. Frank Coolen and Prof. Tahani Coolen-Maturi, for their efforts, support, and guidance through every stage of this journey. Their guidance has greatly shaped my academic path. I truly appreciate the time and effort they invested, which has made this accomplishment possible.

I would also like to extend my warm thanks to my beloved parents, Abdulrahman and Azza Almasoud, for their unconditional love and for being my source of strength and confidence. Their belief in my abilities has given me the confidence to overcome obstacles and reach this milestone. They were always there when I needed them. My sincere thanks extend to my siblings Hazim, Dareen, Ramiz and Shihab for their encouragement and constant support. I will forever owe them more than words can express.

I must thank my husband, Mohammed Alghamdi, for being the shoulder to lean on throughout these years. His support and patience mean a lot to me. His presence gave me strength during the most challenging moments. This journey would not have been the same without him.

I'm also profoundly thankful to my lovely children, Battar, Lateen and Alin, for being my constant source of love, strength, and joy. This journey would not have been the same without their hugs, laughter, and unconditional love.

I sincerely thank the Saudi Cultural Bureau and Jeddah University for their generous funding and continued support. I am especially grateful to Prof. Adnan Alhumaidan, Director of Jeddah University, for his support during my academic journey. I also extend my appreciation to the University of Durham for granting me the opportunity to pursue my studies and for providing a supportive academic environment.

Special thanks to my colleagues for their friendship, valuable discussions, and support. Their collaboration has truly inspired and motivated me throughout this journey. Finally, my heartfelt appreciation goes to everyone who has supported me on this journey, whether with kind words, prayers, or encouragement.

Contents

Chapter 1

Introduction

1.1 Overview

Many multivariate models are introduced to model dependence among variables, which plays an important role in statistical models whenever there are more than two random quantities [9]. The challenge of dependence in multivariate data lies in analyzing the dependence structure, exploring its characteristics and effectively modelling the relationships between variables [19].

In some cases, it is important to measure the dependence between variables, such as the correlation coefficient, which is a fundamental measure. This is a simple measure and insufficient for explaining the dependence structure, especially in multivariate cases. In addition, it assumes a linear relationship between two variables. This means it cannot capture dependencies that are not linear [37]. Several proper measures of modelling dependence can capture complex dependence in high dimensions, including copula, vine copulas and fully nested Archimedean copulas [60].

Copulas are commonly used in modelling dependence between variables. It first appeared by Sklar in [91, 96] based on Fréchet's works in [40]. In general, a copula is a multivariate distribution function with one-dimensional margins that are uniform [60]. It has the feature for modelling the dependence structure and the marginal distributions separately [60, 84]. The Elliptical and Archimedean copula families are well established in the literature, offering flexible approaches to capturing dependence [60, 84]. Many copulas have been developed to capture different types of dependencies and provide a clear

relation between the copula parameters and dependence measures [42, 83].

Joe [60] explores the extension of copulas to the multivariate setting in detail. These models face limitations, as they do not permit negative dependence as in the bivariate case [60]. Copulas are widely applied across various fields, including finance [17, 34, 102], risk management [36, 64] and insurance [20, 61, 93]. A comprehensive overview of classical copula models and their important properties are well described in Chapter 2.

An alternative approach to model dependence is vine copulas, which have proven to be flexible and useful in high dimensions. Vine copulas or pair-copulas constructions (PCC) were first introduced by Joe [60]. It was presented as a decomposition that contains unconditional and conditional bivariate copulas [14, 15]. Modelling dependence using a vine copula can be modelled easily and captures different dependence structures in a model. This is due to the vine construction, which consists of several bivariate copulas, each of which can be of any copula type [3]. Comprehensive overviews of vine copula statistical inference are available by Czado [29], Kurowicka and Joe [67], Dissmann *et al.* [35] and Aas *et al.* [3]. Vine copulas have been widely applied in various fields including finance [1, 30, 85], insurance [94, 101] and risk management [8, 18].

Another model for capturing dependence is the fully nested Archimedean copula (FNAC), which was first introduced by Joe [60]. This model is constructed by nesting two or more bivariate copulas in a model. These bivariate copulas in the FNAC model construction have to be from the same Archimedean family [99]. There is a limitation that FNAC can be modelled by nesting different types of copulas from the Archimedean family to cover various degrees of dependencies [75]. Although FNAC can be constructed by nesting different types of copulas from the Archimedean family to capture various degrees of dependence, certain restrictions must be considered. These restrictions on the combinations of copula types that can be nested to model dependence have been discussed in the literature, Hofert [52], McNeil [74] and Okhrin *et al.* [86]. There are many applications using FNAC model, including finance [55, 90], drought analysis [12, 31, 73] and risk management [36].

Modelling dependence captures the relationships between variables, enhancing the ability to make statistical inferences and produce accurate predictions. Nonparametric predictive inference (NPI) is a frequent statistical framework that requires few assump-

tions [10, 11]. NPI is based on Hill's assumption $A_{(n)}$, which gives a direct conditional probability for one or more future observable random quantities conditional on observed values of related random quantities [49, 50, 51]. Although Hill's assumption $A_{(n)}$ is not sufficient to derive precise probabilities, it provides strong probability bounds for events of interest, including future observations. NPI is a framework of statistical theory and methods that use $A_{(n)}$-based lower and upper probabilities and also considers several variations of $A_{(n)}$ which are suitable for different inferences [10, 11, 26].

NPI was developed recently to address a number of applications in the literature of statistics [24]. NPI has been presented for different data types, including multinomial data [25] and bivariate data [5, 27, 77]. Mainly, two methods using bivariate data have been developed based on the NPI approach. Coolen-Maturi *et al.* [27] and Muhammad *et al.* [77] introduced NPI for bivariate data with parametric or nonparametric copulas. Their performance was evaluated using simulation studies. Muhammad *et al.* used the NPI with copulas for combining bivariate diagnostic tests [78] and survival analysis [80]. A further applications using these methods with smoothed bootstrap methods was introduced by Al Luhayb *et al.* [5, 6] and in spread option pricing model by He [48]. Several studies applied case-based approaches to wildfire prediction using NPI combined with parametric copulas by Muhammad *et al.* [79], Roslin *et al.* [88, 89].

1.2 Outline of the thesis

This thesis presents four methods to extend NPI to multivariate data by combining it with different types of copulas to model the dependence structure. The focus is on parametric copulas with one parameter, nonparametric copulas, vine copulas, and FNAC. The proposed methods are evaluated through simulation studies. The thesis is organized as follows: Chapter 2 presents a brief overview and background of the theoretical notion of copulas, vine copulas and fully nested Archimedean copulas. This chapter also presents an overview of NPI and Hill's assumption $A_{(n)}$, as well as methods for combining NPI with copulas in the bivariate case. Chapter 3 extends the existing methods of combining NPI with bivariate parametric copula and nonparametric copulas that was introduced by Muhammad [27] to the trivariate case. The performance of the two methods investigated

via simulations. A generalization of these two methods will be presented. Chapter 4 introduces a method for combining NPI with vine copulas in three dimensions. Different vine copula structures are presented to illustrate the method, and its performance is evaluated. Chapter 5 presents the method of combining NPI with fully nested Archimedean copulas. The performance of this method is evaluated and a generalization to arbitrary dimensions is provided. A comparison study of the proposed methods is conducted via simulations. Chapter 6 summarizes the key results of this thesis, provides concluding remarks and discusses related topics for future research.

Chapter 2

Preliminaries

2.1 Introduction

This chapter provides an overview of the fundamental concepts from the literature that are used in this thesis. It begins with an introduction to copula theory, covering both parametric and nonparametric copulas. The construction of vine copulas is then discussed in detail, followed by a comprehensive explanation of the fully nested Archimedean copula. Finally, the chapter presents the basic theory of NPI and the method of combining NPI with bivariate data using copulas, which forms the essential framework for this thesis.

2.2 Copula

A copula is a method used to model the dependence between random variables. Sklar was the first to use the term "copula" [96]. A copula is a Latin word that means link, tie or connection, describing how it connects the marginal distributions. Its importance arises from its ability to describe the dependence structure separately from the marginals. A copula is used to combine two or more univariate marginal distributions into a multivariate distribution. Statistically, a copula (C) is a multivariate cumulative distribution function with uniform marginal distributions on $[0, 1]$ and $C : [0, 1]^d \to [0, 1]$ for $d \geq 2$ [59, 84].

Let $\underline{X} = (X_1, X_2, \ldots, X_d)$ and $\underline{x} = (x_1, x_2, \ldots, x_d)$. The multivariate cumulative distribution function $F(x_1, x_2, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$ of a random vector $(X_1, X_2, \ldots, X_d)$ can be expressed in term of the marginal CDFs F_i for $i = 1, \ldots, d$ and

a copula C.

$$F_{\underline{X}}(\underline{x}) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{2.2.1}$$

The corresponding multivariate density function $f_{\underline{X}}$ is given by

$$f_{\underline{X}}(\underline{x}) = c(F_1(x_1), \ldots, F_d(x_d)) * f_1(x_1) \ldots f_d(x_d) \tag{2.2.2}$$

where $c(.)$ is the density function of the copula and f_i is the marginal density function of X_i [59, 84]. The copula function satisfies the following properties

1. $C(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_d) = 0$
2. $C(1, \ldots, x_i, \ldots, 1) = x$ for all $x_i \in [0, 1]$
3. $\sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+i_2+\cdots+i_d} C(x_1, x_2, \ldots, x_d) \geq 0$

Measuring the dependence between random variables is an important tool. A copula has the ability to model the dependence structure. There is a relationship between dependence and the copula's parameter, which helps characterize various types of dependencies. A rank-based dependence measure such as Kendall's τ [63] can be written in terms of the copula [66, 72]. Assume X and Y are continuous random variables with a copula $C_{X,Y}$ associated with $F_{X,Y}(x, y)$, then Kendall's τ can be expressed as

$$\tau = 4 \int_{[0,1]^2} C_{X,Y}(x, y) dC_{X,Y}(x, y) - 1, \tag{2.2.3}$$

2.3 Families of copulas

There is a huge variety of parametric copula models presented in the literature. Each of these models possesses specific characteristics. This thesis focuses on some of the most popular families of copulas, such as the Gaussian (Normal) copula, Clayton, Gumbel, Frank and Joe copulas [84]. The Gaussian copula is one of the elliptical copulas [84]. The Gaussian copula is implemented by applying a Normal distribution to the univariate marginals and a multivariate Normal distribution to the joint distribution [38]. For symmetric, positive definite matrix with $\text{diag}(R) = (1, 1, \ldots, 1)^T$ the Gaussian copula with parameter $\theta_n \in [-1, 1]$ is given by

$$C(x_1, \ldots, x_d) = \Phi_R(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_d)) \tag{2.3.1}$$

where Φ^{-1} is the inverse cumulative distribution function of standard normal distribution and Φ_R is the multivariate standard normal distribution function, with linear correlation matrix R [84, 104]. Another popular family of copulas is the Archimedean copulas [92], which allows for modelling dependence, including different types of tail dependence [71]. This type of copula has a simple closed-form expression and is constructed using a generator function with a single parameter θ, which is restricted to the interval $(0, \infty)$ [84]. It is given by

$$C(x_1, \ldots, x_d; \theta) = \psi^{[-1]}(\psi(x_1; \theta) + \cdots + \psi(x_d; \theta); \theta) \tag{2.3.2}$$

where ψ is a smooth, strictly decreasing, continuous function known as the generator of copula and $\psi^{[-1]}$ denotes the pseudo inverse of the generator function. There are different types of Archimedean copulas, including Clayton [23], Frank [39], Ali-Mikhail-Haq [7] and Gumbel [103]. More details on Archimedean copulas can be found in [44, 84].

The Clayton copula is a one-parameter copula with positive lower tail dependence, where there is more dependence in the negative tail than in the positive tail [23]. The Clayton copula for $\theta_c > 0$ has the form

$$C_{\theta_c}(x_1, \ldots, x_d) = (\sum_{i}^{d} (x_i)^{-\theta_c})^{-1/\theta_c} - d + 1) \tag{2.3.3}$$

The Gumbel copula models upper tail dependence between variables [103]. The generator function is $\psi(x; \theta_g) = (-\ln(x))^{\theta_g})$ and its inverse is $\psi^{[-1]}(x) = \exp(-x^{1/\theta_g})$ for $\theta_g \geq 1$ leading to

$$C_{\theta_g}(x_1, \ldots, x_d) = \exp -[\sum_{i=1}^{m} (-\ln x_i)^{\theta_g}]^{1/\theta_g} \tag{2.3.4}$$

The Frank copula is a symmetric copula [39], has the generator function $\psi(x; \theta_f) = -\ln(\frac{\exp(-\theta_f x)-1}{\exp(-\theta_f)-1})$ and inverse generator function $\psi^{[-1]}(x) = -\frac{1}{\theta_f}\ln(1+\exp(-x)(\exp(-\theta_f)-1))$ for $\theta > 0$ leading to

$$C_{\theta_f}(x_1, \ldots, x_d) = -\frac{1}{\theta_f}\ln(1 + \frac{\prod_i (e^{\theta_f x_i})}{e^{-\theta_f} - 1}) \tag{2.3.5}$$

The Joe copula models upper tail dependence between variables [59]. The generator function is $\psi(x; \theta_j) = -\ln(1-(1-x)^{\theta_j})$ and its inverse is $\psi^{[-1]}(x) = 1 - [1 - \exp(-x)]^{1/\theta_j}$ for $\theta_j \geq 1$ leading to

$$C_{\theta_j}(x_1, \ldots, x_d) = 1 - (1 - [1 - (1 - x)^{\theta_j}] \ldots [1 - (1 - x_d)^{\theta_j}])^{1/\theta_j}; \theta_j > 0 \tag{2.3.6}$$

A detailed review of the properties of these copulas can be found in [19, 60, 84].

The estimation of copula parameters is widely discussed in the literature [21, 45]. Generally, estimation methods are categorized as parametric, semiparametric or nonparametric. The parametric estimation method, such as maximum likelihood estimation (MLE) or inference functions for margins (IFM), is preferred when the marginal distributions are known. However, the performance of these estimators may be affected if the marginal distributions are misspecified. In practice, margins are often unknown, making it challenging to identify the most suitable copula for modelling the dependence structure. To address this issue, semiparametric estimation methods, such as pseudo maximum likelihood estimation method (PML), can be implemented. These methods estimate the margins nonparametrically and estimate the copula parameterically.

The PML method focuses on choosing the value of the parameter $\hat{\theta}$ that maximizes the log pseudo likelihood function with empirical marginal distributions $\hat{F}_{X_i}$. It is shown to be the most efficient and optimal and its properties are studied by Cherubini *et al.* [22]. The copula parameter is estimated by maximising the copula density as

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \log \left[c \left(\hat{F}_{X_1}(x_{1,i}), \ldots, \hat{F}_{X_d}(x_{d,i}); \theta \right) \right] \tag{2.3.7}$$

Nonparametric estimation methods do not depend on assumptions regarding the marginal distributions or the dependence structure, offering flexibility when the true forms of the marginals or copula are unknown [45, 95]. Nonparametric estimation of copula density can be implemented using kernels [69, 81], using wavelets [43], using splines [62], based on Bernstein polynomials [70] or empirical copulas, as introduced by Deheuvels [32]. In this thesis, estimating the copula using kernel is used because it provides a smooth estimate of the copula density. The kernel method is widely used in nonparametric statistics and only requires choosing a bandwidth, which can be selected using cross-validation or a plug-in estimator.

The kernel density estimator is a way of using kernels as weights to estimate the probability density function of a random variable. The idea works by centering a kernel of probability weight $1/n$ on each observation, which is controlled by a smoothing parameter, the bandwidth b. There are many ways to calculate the bandwidth, including the normal reference rule-of-thumb and least square cross-validation (LCSV). The first method is

given by Silverman [95].

$$b = 1.06\ \hat{\sigma}_j n^{-1/(d+4)} \tag{2.3.8}$$

where d is the number of variables. The LSCV method selects a bandwidth that minimizes the integrated squared error of the estimate from the data [95]. The LSCV criterion for selecting the optimal bandwidth b in kernel density estimation is given by

$$LSCV(\mathbf{b}) = \int_{\mathbb{R}^d} \hat{f}(\mathbf{x};\mathbf{b})^2\, d\mathbf{x} - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(\mathbf{X}_i;\mathbf{b}) \tag{2.3.9}$$

from which follows:

$$\hat{\mathbf{b}}_{LSCV} = \arg\min_{\mathbf{b}\in\mathcal{F}} LSCV(\mathbf{b}). \tag{2.3.10}$$

Given the random quantities $\underline{X}$ and $\underline{x} = (x_1, \ldots, x_d)$. Let $(U_{i,1}, \ldots, U_{i,d}) \sim [0,1]$ is the transformed rank of the random quantities $\underline{X}$ with joint distribution C and corresponding probability density function, $c : [0,1]^d \to \mathbb{R}$. Nonparametric copula distributions based on the kernel density estimator is given by

$$\hat{C}(u_1, \ldots, u_d) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K\left(\frac{u_j - U_{i,j}}{b}\right) \tag{2.3.11}$$

where for all $u_1, \ldots, u_d \in [0,1]$, $K : \mathbb{R}^d \to \mathbb{R}$ is a multivariate kernel function and $b > 0$ is the bandwidth.

2.4 Vine copulas

In the previous section, several copulas are introduced, which will be used in the two-dimensional case in Chapter 4 of this thesis. This section provides another way to model dependence, which is a flexible multivariate model called the vine copulas or pair-copula constructions (PCC) model. Vine copulas are introduced by Joe [58, 59] and discussed by Bedford and Cooke [14, 15, 16]. This construction models the multivariate variables using bivariate copulas. The flexible structure of vine copulas allows for the inclusion of various types of dependencies [28]. The vine copula structure is not unique and it has $\frac{d!}{2}$ possible decompositions in a d-dimension [1, 2]. The structure of vine copulas overcomes the limitations of classical copula models, offering flexible multivariate modelling. Consider a random variables $\boldsymbol{X} = (X_1, X_2, X_3)$ with a joint distribution function $f(x_1, x_2, x_3)$

and univariate marginal densities f_1, f_2, f_3 and F_1, F_2, F_3 are the marginal distribution functions. Then, the vine copula decomposition of a trivariate case takes the form

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1) \tag{2.4.1}$$

where $f_{3|12}(x_3|x_1, x_2)$ is determined by considering the bivariate conditional density $f_{13|2}(x_3, x_1|x_2)$ and it can expressed as follows

$$\begin{aligned} f_{3|12}(x_3|x_1, x_2) &= \frac{f_{13|2}(x_3, x_1|x_2)}{f_{1|2}(x_1|x_2)} \\ &= \frac{c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) f_{1|2}(x_1|x_2) f_{3|2}(x_3|x_2)}{f_{1|2}(x_1|x_2)} \\ &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) f_{3|2}(x_3|x_2) \end{aligned} \tag{2.4.2}$$

The last two terms in the right-hand numerator can be expressed as

$$f_{2|1}(x_2|x_1) = \frac{c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) f_1(x_1)}{f_1(x_1)} = c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) \tag{2.4.3}$$

$$f_{3|2}(x_3|x_2) = \frac{c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3) f_2(x_2)}{f_2(x_2)} = c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3) \tag{2.4.4}$$

Substituting these expressions to Equation (2.3.4) leads to

$$\begin{aligned} f(x_1, x_2, x_3) = {} & c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{23}(F_2(x_2), F_3(x_3)) c_{12}(F_1(x_1), F_2(x_2)) \\ & f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned} \tag{2.4.5}$$

This construction is not unique; for three dimensions, there are three different decompositions, where different decompositions lead to different results as presented in Figure 2.1. By reordering the variables, the structure could be written as

$$\begin{aligned} f(x_1, x_2, x_3) = {} & c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) c_{13}(F_1(x_1), F_3(x_3)) c_{12}(F_1(x_1), F_2(x_2)) \\ & f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned} \tag{2.4.6}$$

or

$$\begin{aligned} f(x_1, x_2, x_3) = {} & c_{12|3}(F_{1|3}(x_1|x_3), F_{2|1}(x_2|x_1)) c_{13}(F_1(x_1), F_3(x_3)) c_{23}(F_2(x_2), F_3(x_3)) \\ & f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned} \tag{2.4.7}$$

All these three structures have three parameters $\theta_1, \theta_2, \theta_3$ and a conditional copula term. For example, in Equation (2.4.5) the conditional copula term $c_{13|2}(x_1, x_3|x_2)$ depends on x_2. To simplify estimation, the dependence on the specific conditioning value is

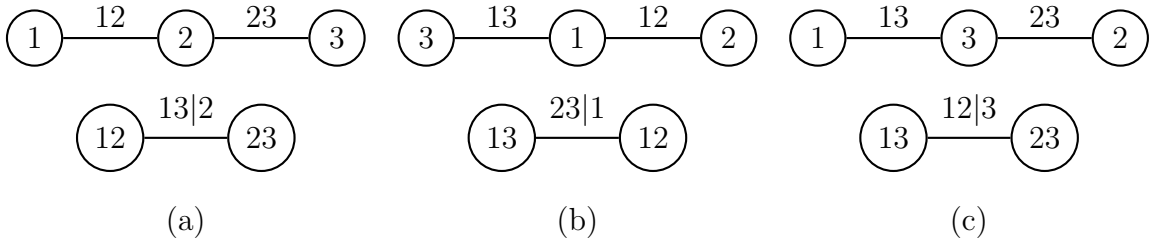


Figure 2.1: Three-dimensional different construction based on reordering variables

ignored [3, 4, 98]. It assumes that the conditional copula is independent of the conditioning variables, except through the conditional distributions [29]. This conditional copula can be obtained by partial differentiation $F_{1|2}(x_1|x_2) = \frac{\partial C_{12}(x_1,x_2;\hat{\theta_1})}{\partial x_2}$ and $F_{3|2}(x_3|x_2) = \frac{\partial C_{23}(x_2,x_3;\hat{\theta_3})}{\partial x_2}$ [59].

2.5 Nested Archimedean copulas (NAC)

Another structure for modelling dependence, in addition to vine copulas and classical copulas, is the Nested Archimedean copula (NAC), also known as the asymmetric or hierarchical copula. NAC was first introduced by Joe [60] and is extensively discussed by Trede and Savu [90] and Embrechts *et al.* [36]. In general, there are two constructions derived under the name of Nested Archimedean copulas: Fully Nested Archimedean copulas (FNAC) and Partially Nested Archimedean copulas (PNAC). In this thesis, FNAC is used in Chapter 5.

A copula C is a fully nested Archimedean copula (FNAC) if it is an Archimedean copula whose arguments can be replaced by other nested Archimedean copulas. FNAC with $d-1$ hierarchies or nesting levels is defined by

$$\begin{aligned} C(x_1,\ldots,x_d) = \psi_{d-1}^{-1}\Big(\psi_{d-1}\circ\psi_{d-2}^{-1}\Big[\cdots\Big(\psi_2\circ\psi_1^{-1}\big[\psi_1(x_1)+\psi_1(x_2)\big] \\ +\psi_2(x_3)\Big)+\cdots+\psi_{d-2}(x_{d-1})\Big]+\psi_{d-1}(x_d)\Big) \end{aligned} \tag{2.5.1}$$

for $(x_1,\ldots,x_d)\in[0,1]^d$, where $\psi_1,\ldots,\psi_{d-1}$ are the generators of Archimedean copulas, $\psi_1^{-1},\ldots,\psi_{d-1}^{-1}$ the inverse generators and $\circ$ is the composite function with parameters $\theta_1\geq\theta_2\geq\cdots\geq\theta_{d-1}$. McNeil [74] proved that FNAC is a copula if all the generators

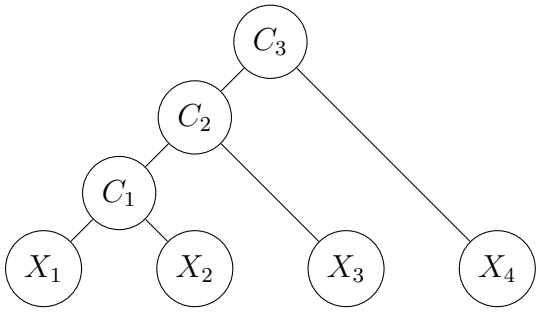


Figure 2.2: Four-dimensional FNAC structure

$\psi_1^{-1}, \psi_2^{-1}, \ldots, \psi_{d-1}^{-1}$ are completely monotone. $\psi_{i+1} \circ \psi_i^{-1}$ must have completely monotone derivatives for all levels i of the nesting [52, 74].

If all the generators are of the same type from the Archimedean family e.g all the generators are Clayton copulas then the FNAC structure is a copula. On the contrary, if the generators are from different types from the Archimedean family then the composition function $\psi_{i+1} \circ \psi_i^{-1}$ may not have a completely monotonic derivative. Thus, the FNAC structure is not a copula. The three-dimensional FNAC expressions of Clayton, Gumbel, Frank and Joe copulas are listed in Table 2.1. The possible choices of which different families could be combined were discussed and investigated by Hofert [52, 53], Hofert and David [54] and McNeil [36].

Equation (2.5.2) and Figure 2.2, represent the FNAC structure in the four-dimensional case. There are three bivariate copulas C_1, C_2, C_3 required to model the dependence for the four-dimensional random variables X_1, X_2, X_3 and X_4 using FNAC. First, X_1 and X_2) are coupled by copula C_1. Then, X_3 is coupled with C_1 by C_2. After that, the random variable X_4 is coupled with C_2 by C_3.

$$\begin{aligned} C(x_1, x_2, x_3, x_4) &= C_3(x_4, C_2(x_3, C_1(x_1, x_2))) \\ &= \psi_3^{-1}\Big(\psi_3(x_4) + \psi_3\big(\psi_2^{-1}\big(\psi_2(x_3) + \psi_2\big(\psi_1^{-1}(\psi_1(x_1) + \psi_1(x_2))\big)\big)\big)\Big) \end{aligned} \tag{2.5.2}$$

FNAC	$C(x_1, x_2, x_3)$	Parameter Range
Clayton	$(x_3^{-\theta_1} + (x_1^{-\theta_2} + x_2^{\theta_2} - 1)^{\frac{\theta_1}{\theta_2}})^{\frac{-1}{\theta_1}}$	$\theta_1 \leq \theta_2 \in [0, \infty)$
Gumbel	$\exp(-([(-\ln x_1)^{\theta_2} + (-\ln x_2)^{\theta_2}]^{\frac{\theta_1}{\theta_2}}) + (-\ln x_3)^{\theta_1})^{\frac{1}{\theta_1}})$	$\theta_1 \leq \theta_2 \in [1, \infty)$
Frank	$-\frac{1}{\theta_1}\ln(1 - (1 - \exp(-\theta_1))^{-1}(1 - [1 - (1 - \exp(-\theta_2))^{-1}(1 - \exp(-x_1\theta_2))$ $(1 - \exp(-x_2\theta_2))]^{\frac{\theta_1}{\theta_2}})(1 - \exp(-x_3\theta_1)))$	$\theta_1 < \theta_2 \in [0, \infty)$
Joe	$1 - (((1 - x_1)^{\theta_2}(1 - (1 - x_2)^{\theta_2}) + (1 - x_2)^{\theta_2})^{\frac{\theta_1}{\theta_2}}(1 - (1 - x_3)^{\theta_1}) + (1 - x_3)^{\theta_1})^{\frac{1}{\theta_1}}$	$\theta_1 \leq \theta_2 \in [1, \infty)$

Table 2.1: The trivariate FNAC

2.6 Nonparametric Predictive Inference (NPI)

Nonparametric predictive inference (NPI) is a statistical method used to make inferences about a future observation based on past data [24]. NPI is built on the assumption $A_{(n)}$ of Hill [50], which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities.

Assume that $X_1, X_2, \ldots, X_n, X_{n+1}$ are continuous and exchangeable real-valued random quantities. Let the ordered observed values of $X_1, X_2, \ldots, X_n$ be denoted by $x_1 < x_2 < \cdots < x_n$. Let $x_0 = -\infty$ and $x_{n+1} = \infty$ [50]. Assume further there are no tied observations, but if there are ties, they can be treated by adding a small value close to zero to break ties [51]. For a future observation X_{n+1}, the assumption $A_{(n)}$ is $P(X_{n+1} \in (x_{i-1}, x_i) = \frac{1}{n+1}$ where $i = 1, 2, \ldots, n+1$.

$A_{(n)}$ does not assume anything else and is a post-data assumption related to exchangeability [50]. Inferences based on $A_{(n)}$ are predictive and nonparametric and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations or if one does not want to use such information. Although $A_{(n)}$ is not sufficient to derive precise probabilities, it provides optimal bounds for probabilities for all events of interest, including future observation [10]. These bounds are lower and upper probabilities in the theory of imprecise probability and are denoted by $\underline{P}(A)$ and $\overline{P}(A)$, respectively. $\underline{P}(A)$ is the lower probability of an event A and can be interpreted as the maximum lower bound for the probability A. $\overline{P}(A)$ is the upper probability of an event A and can be interpreted as the minimum upper bound for the probability of A. In imprecise probability theory, $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ and the lower and upper probabilities are conjugate $\underline{P}(A) = 1 - \overline{P}(A^c)$ where A^c is the complement event of A. These properties

hold for the NPI lower and upper probabilities, as demonstrated by Coolen and Augustin [10].

2.7 Combining NPI with copulas for bivariate data

Coolen-Maturi *et al.* [27, 77] introduced two methods for applying NPI in the bivariate case, taking into account the dependence structure using parametric and nonparametric copulas. The idea can be categorized into two steps: applying NPI to the marginals, then using an assumed copula. The semi-parametric method assumes a parametric copula in the second step, while the nonparametric method assumes a nonparametric copula based on a kernel approach. The notation and concepts provided by Muhammad [76] are used to introduce the two prediction approaches.

Assume there are n bivariate observations $(x_i, y_i), i = 1, \ldots, n$ corresponding to n exchangeable bivariate random quantities with no ties. The observations are ordered such as $x_1 < \cdots < x_i < \cdots < x_n$ and $y_1 < \cdots < y_i < \cdots < y_n$. Let the new future bivariate observation denoted as (X_{n+1}, Y_{n+1}) given the past observations $(x_i, y_i), i = 1, \ldots, n$. Using the assumption $A_{(n)}$ for the matginals gives

$$P(X_{n+1} \in (x_{i-1}, x_i)) = \frac{1}{n+1} \quad \text{and} \quad P(Y_{n+1} \in (y_{j-1}, y_j)) = \frac{1}{n+1} \tag{2.7.3}$$

for $i, j = 1, 2, \ldots, n+1$ where $x_0 = -\infty$, $x_{n+1} = \infty$, $y_0 = -\infty$ and $y_{n+1} = \infty$.

To link the first step to the second one where the dependence structure in the observed data is taken into account. Mohammad [76] introduced a natural transformation of the random quantities X_{n+1} and Y_{n+1} by using a corresponding transformation $\tilde{X}_{n+1}$ and $\tilde{Y}_{n+1}$ where

$$(X_{n+1} \in (x_{i-1}, x_i), Y_{n+1} \in (y_{j-1}, y_j) \Longleftrightarrow (\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1})) \tag{2.7.4}$$

where $i, j = 1, 2, \ldots, n+1$. The $A_{(n)}$ assumption for the marginal using the transformation lead to

$$P(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1})) = P(X_{n+1} \in (x_{i-1}, x_i)) = \frac{1}{n+1} \tag{2.7.5}$$

$$P(\tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}) = P(Y_{n+1} \in (y_{j-1}, y_j)) = \frac{1}{n+1} \tag{2.7.6}$$

This is a transformation from the real space $\mathbb{R}^2$ to $[0,1]^2$, where $[0,1]^2$ is divided into $(n+1)^2$ equal sized squares by the n observed bivariate observations. The uniform marginal distributions have been discretized using this transformation. For the second step a parametric copula is assumed and the copula parameter θ is estimated. The parameter is estimated using transformed data by replacing the observed pair (x_i, y_i), $i = 1, \ldots, n$ by $(\frac{r_i^x}{(n+1)}, \frac{r_i^y}{(n+1)})$ where r_i^x is the rank of the observation x_i among x-observations and r_i^y is the rank of the observation y_i among y-observations. The estimated copula and NPI on the marginal are now combined by defining the probability that the transformed pair $(\tilde{X}_{n+1}, \tilde{Y}_{n+1})$ belonged to a certain square from the $(n+1)^2$ squares into which the space $[0,1]^2$ has been partitioned as follow:

$$h_{ij}(\hat{\theta}) = P_C(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1})|\hat{\theta}) \tag{2.7.7}$$

where $i, j \in 1, \ldots, n+1$ and $P_C(.|\hat{\theta})$ is the assumed copula-based probability with the estimated parameter $\hat{\theta}$. The h_{ij} values satisfies $\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij} = 1$ and each value of h_{ij} is between 0 and 1 and $\sum_{j=1}^{n} h_{ij} = \frac{1}{n+1}$ for all $i \in \{\ 1, \ldots, n+1\}$ and $\sum_{i=1}^{n} h_{ij} = \frac{1}{n+1}$ for all $j \in \{\ 1, \ldots, n+1\}$

In the second method, a nonparametric kernel-based copula replaces the parametric copula. First, NPI is applied to the marginals, followed by the same transformation as before. The kernel-smoothed copula density estimator $\hat{c}$ is defined as

$$\hat{c}(x,y) = \frac{1}{nb_X b_Y} \sum_{i=1}^{n} K\left(\frac{x - F_X(\tilde{X}_i)}{b_X}, \frac{y - F_Y(\tilde{Y}_i)}{b_Y}\right) \tag{2.7.8}$$

where $K : \mathbb{R}^2 \to \mathbb{R}$ is a bivariate kernel function, $b_X, b_Y > 0$ are the bandwidths or smoothing parameters, $F_X(\tilde{X}_i) = \frac{r_i^x}{n+1}$ and $F_Y(\tilde{Y}_i) = \frac{r_i^y}{n+1}$. Now, the NPI approach for the marginals, can be combined with the nonparametric kernel-based copula to take the dependence into account as follows,

$$h_{ij}(\hat{c}) = P_C(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1})|\hat{c}) \tag{2.7.9}$$

where i and $j \in 1, \ldots, n+1$ and $P(.|\hat{c})$ represents the nonparametric kernel-based copula probability with estimated kernel density function $\hat{c}$ as defined in Equation (2.7.8). These

h_{ij} values must satisfy the three conditions as explained in the semi-parametric method. This chapter introduces the theoretical background for the topics discussed in this thesis from the literature. This thesis considers mainly the trivariate case where the novelty lies in the consideration of different copula constructions.

Chapter 3

NPI Combined with Classical Copula

3.1 Introduction

As stated in the previous chapter, Coolen-Maturi *et al.* [27, 77] introduced two methods for predictive inference of a future observation based on bivariate data: a semi-parametric method and a nonparametric method. Both follow a two-step approach: first, applying nonparametric predictive inference (NPI) to the marginals; second, modelling the dependence structure using either a parametric or a nonparametric copula. This chapter presents two extensions of these methods to the multivariate case, maintaining the same structure by first applying NPI to the marginals and then using a multivariate copula to capture dependence.

This chapter is organized as follows. Section 3.2 introduces an extension of the semi-parametric method in the trivariate case followed by a generalization of this method to the multivariate case. Section 3.3 introduces an extension of the nonparametric method in the trivariate case and this is followed by an extension of this method to the multivariate case. The proposed methods are explained through examples in Section 3.4. The predictive performance of the proposed methods is evaluated through simulations in Section 3.5. Section 3.6 presents examples from the literature to illustrate the applications of the proposed methods. Concluding remarks of this chapter are provided in Section 3.7.

3.2 Combining NPI with a parametric copula

A semiparametric method is introduced, combining Nonparametric Predictive Inference (NPI) for the marginals with a parametric trivariate copula. The method extends Muhammad's work by focusing on the trivariate case [76]. The method involves two main steps: first, applying NPI to the marginals; second, assuming a trivariate parametric copula and estimating its parameter to capture the dependence structure.

For the first step, assume that there are n trivariate observations (x_i, y_i, z_i), $i = 1, \ldots, n$, representing the observed values of n exchangeable trivariate random quantities with no ties. The observations of the marginals are ordered and denoted by x_i, y_j and z_k for simplicity, so $x_1 < \cdots < x_i < \cdots < x_n$, $y_1 < \cdots < y_j < \cdots < y_n$ and $z_1 < \cdots < z_k < \cdots < z_n$.

Using Hill's assumption $A_{(n)}$, it is possible to derive a partially specified predictive probability distribution for each X_{n+1}, Y_{n+1} and Z_{n+1} given the observations $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ and $z_1, \ldots, z_n$, respectively, lead to $P(X_{n+1} \in (x_{i-1}, x_i)) = \frac{1}{n+1}$, $P(Y_{n+1} \in (y_{j-1}, y_j)) = \frac{1}{n+1}$ and $P(Z_{n+1} \in (z_{k-1}, z_k)) = \frac{1}{n+1}$ for i, j, $k \in \{ 1, \ldots, n+1\}$, where $x_0 = -\infty, x_{n+1} = \infty$, $y_0 = -\infty, y_{n+1} = \infty$ and $z_0 = -\infty, z_{n+1} = \infty$.

To link the first step with the second step, where the dependence structure in the data is taken into account to provide a partially specified predictive distribution for the trivariate $(X_{n+1}, Y_{n+1}, Z_{n+1})$ by introducing a natural transformation of the three random quantities individually as introduced by Muhammad [76]. Let $\tilde{X}_{n+1}$, $\tilde{Y}_{n+1}$ and $\tilde{Z}_{n+1}$ denote the transformed versions of the random quantities X_{n+1} , Y_{n+1} and Z_{n+1}, respectively, such that

$$(X_{n+1} \in (x_{i-1}, x_i), Y_{n+1} \in (y_{j-1}, y_j, Z_{n+1} \in (z_{k-1}, z_k)) \Longleftrightarrow$$
$$(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1})) \quad (3.2.1)$$

where i, j and $k = 1, \ldots, n+1$. This transformation from the real space $\mathbb{R}^3$ to $[0,1]^3$ is based on n trivariate data, where $[0,1]^3$ is divided into $(n+1)^3$ equal sized blocks. By following these transformations of the marginals, the uniform marginal distribution on $[0,1]$ has been discretized. The $A_{(n)}$ assumption for the marginals after the transformation lead to

$$P(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1})) = P(X_{n+1} \in (x_{i-1}, x_i)) = \frac{1}{n+1} \quad (3.2.2)$$

$$P(\tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}) = P(Y_{n+1} \in (y_{j-1}, y_j)) = \frac{1}{n+1} \tag{3.2.3}$$

and

$$P(\tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1}) = P(Z_{n+1} \in (z_{k-1}, z_k)) = \frac{1}{n+1} \tag{3.2.4}$$

The second step is to use a trivariate parametric copula and estimate the copula parameter. The copula parameter can be estimated using the transformed data where each observed pair (x_i, y_i, z_i), $i = 1, \ldots, n$ is replaced by $(\frac{r_i^x}{(n+1)}, \frac{r_i^y}{(n+1)}, \frac{r_i^z}{(n+1)})$ where r_i^x the rank of the observation x_i among x-observations, r_i^y the rank of the observation y_i among y-observations and r_i^z the rank of the observation z_i among z-observations.

Now, NPI on the marginals can now be combined with the estimated copula by defining the probability for the event that the transformed pair $(\tilde{X}_{n+1}, \tilde{Y}_{n+1}, \tilde{Z}_{n+1})$ belongs to a certain block from $(n+1)^3$ blocks into which the space $[0,1]^3$ has been partitioned,

$$h_{ijk}(\hat{\theta}) = P_C(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1})|\hat{\theta}) \tag{3.2.5}$$

where i, j and $k \in 1, \ldots, n+1$ and $P_C(.|\hat{\theta})$ is the assumed copula-based probability with estimated parameter $\hat{\theta}$ and the cumulative distribution function is

$$H_{ijk}(\hat{c}) = P_C(\tilde{X}_{n+1} \leq \frac{i}{n+1}, \tilde{Y}_{n+1} \leq \frac{j}{n+1}, \tilde{Z}_{n+1} \leq \frac{k}{n+1}|\hat{\theta}) \tag{3.2.6}$$

These values $(n+1)^3$ of $h_{ijk}(\hat{\theta})$ that sum to one provide a fully discretized probability distribution for the transformed future observations. This distribution can be used for making inferences about the actual future observation or any event of interest. The probabilities satisfies:

1. $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\theta}) = 1$
2. $\sum_{j=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\theta}) = \frac{1}{n+1}$, for $i \in \{1, 2, \ldots, n+1\}$, $\sum_{i=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\theta}) = \frac{1}{n+1}$, for $j \in \{1, 2, \ldots, n+1\}$ and $\sum_{i=1}^{n} \sum_{j=1}^{n} h_{ijk}(\hat{\theta}) = \frac{1}{n+1}$, for $k \in \{1, 2, \ldots, n+1\}$
3. $h_{ijk}(\hat{\theta}) \geq 0$, for i, j,$k \in \{1, \ldots, n+1\}$

Using Equation (3.2.5), one can make an inference about an event involving the next future trivariate observation $(X_{n+1}, Y_{n+1}, Z_{n+1})$. Assume that $E(X_{n+1}, Y_{n+1}, Z_{n+1})$

is the event of interest and the corresponding lower and the upper probabilities are $\overline{P}(E(X_{n+1}, Y_{n+1}, Z_{n+1}))$ and $\underline{P}(E(X_{n+1}, Y_{n+1}, Z_{n+1}))$, respectively. For this event to be true, the observed data (x_i, y_i, z_i), for $i = 1, \ldots, n$, partitioned $\mathbb{R}^3$ into $(n+1)^3$ blocks $B_{ijk} = (x_{i-1}, x_i) \odot (y_{j-1}, y_j) \odot (z_{k-1}, z_k)$, for $i, j, k = 1, \ldots, n+1$. By defining the event of interest as follows

$$E(x, y, z) = \begin{cases} 1 & \text{if } E(X_{n+1}, Y_{n+1}, Z_{n+1}) \text{ is true for } X_{n+1} = x,\ Y_{n+1} = y \text{ and } Z_{n+1} = z \\ 0 & \text{else} \end{cases}$$

Let $\overline{E}_{ijk} = \max_{(x,y,z) \in B_{ijk}} E(x, y, z)$, so $\overline{E}_{ijk} = 1$ if there is at least one $(x, y, z) \in B_{ijk}$ for which $E(x, y, z) = 1$, else $\overline{E}_{ijk} = 0$. By defining $\underline{E}_{ijk} = \min_{(x,y,z) \in B_{ijk}} E(x, y, z)$, so $\underline{E}_{ijk} = 1$ if $E(x, y, z) = 1$ for all $(x, y, z) \in B_{ijk}$, else $\underline{E}_{ijk} = 0$. The lower and upper probabilities for the event $E(X_{n+1}, Y_{n+1}, Z_{n+1})$ leads to

$$\underline{P}(E(X_{n+1}, Y_{n+1}, Z_{n+1})) = \sum_{i,j,k} \underline{E}_{ijk} h_{ijk}(\hat{\theta}) \tag{3.2.7}$$

$$\overline{P}(E(X_{n+1}, Y_{n+1}, Z_{n+1})) = \sum_{i,j,k} \overline{E}_{ijk} h_{ijk}(\hat{\theta}) \tag{3.2.8}$$

For example, if the event of interest is that the sum of the next observations X_{n+1}, Y_{n+1} and Z_{n+1} is greater than a value t, expressed as $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$. The lower probability for the event that the sum of the next observations will exceed a particular value t; $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$ is

$$\underline{P}(T_{n+1} > t) = \sum_{(i,j,k) \in L_t} h_{ijk}(\hat{\theta}) \tag{3.2.9}$$

where $L_t = \{\ (i, j, k) : x_{i-1} + y_{j-1} + z_{k-1} > t, 1 \leq i \leq n+1, 1 \leq j \leq n+1, 1 \leq k \leq n+1\}$ and the upper probabilities takes the form

$$\overline{P}(T_{n+1} > t) = \sum_{(i,j,k) \in U_t} h_{ijk}(\hat{\theta}) \tag{3.2.10}$$

where $U_t = \{\ (i, j, k) : x_i + y_j + z_k > t, 1 \leq i \leq n+1, 1 \leq j \leq n+1, 1 \leq k \leq n+1\}$. The lower and upper probabilities in Equations (3.2.9) and (3.2.10), which can also be interpreted as survival functions for the future observation T_{n+1}, will be denoted by $\underline{S}(t)$, $\overline{S}(t)$, respectively.

High Correlation		
X	Y	Z
-0.831	0.181	-0.134
1.197	0.220	0.512
0.407	0.315	0.202
0.042	-0.394	-0.756

No Correlation		
X	Y	Z
-0.445	-0.466	-0.104
1.408	-1.316	0.245
1.205	-0.441	0.423
0.965	1.008	0.050

Table 3.1: Simulated data from a trivariate Gaussian distribution with different correlation structures.

Example 3.2.1 Two three-dimensional visualizations of the probabilities $h_{ijk}(\hat{\theta})$ under different dependence structures are shown in Figures 3.1 and 3.2. Each figure is based on a dataset of size $n = 4$, simulated from a trivariate Gaussian distribution with mean vector zero. Two cases are considered: a no correlation case where the covariance matrix is the identity matrix and a high correlation case where all off-diagonal entries of the covariance matrix are set to 0.9 as shown in Table 3.1.

The parameter is estimated using the pseudo maximum likelihood method, assuming a trivariate Frank copula. Using a classical copula with a single dependence parameter to model the entire trivariate distribution implies that each pair of variables shares the same level of dependence. The relationship between the parameter values and their associated Kendall τ values demonstrates that the dependence structure is governed by this single parameter. As θ increases, the dependence between the variables becomes stronger. For the first dataset, which is highly correlated, the estimated parameter is 8.41 with a corresponding Kendall τ of 0.62. For the second dataset, the estimated parameter is $\hat{\theta} = 2.68$ with a corresponding Kendall τ value of 0.27, indicating a weaker positive dependence. This illustrates how the copula's dependence parameter affects the dependence structure as measured by Kendall τ. The estimated parameter is used to calculate the probabilities $h_{ijk}(\hat{\theta})$ as given by Equation (3.2.5). Figures 3.1 and 3.2 present the marginals of the probabilities $h_{ijk}(\hat{\theta})$. Each figure presents three sides: the right side $h_{.jk} = \sum_i h_{ijk}$, the left side $h_{i.k} = \sum_j h_{ijk}$ and the bottom side $h_{ij.} = \sum_k h_{ijk}$.

On each side, there are large values of the h_{ijk} when $i = j = k$, particularly when the data is highly correlated compared to when the data has no correlation. There is

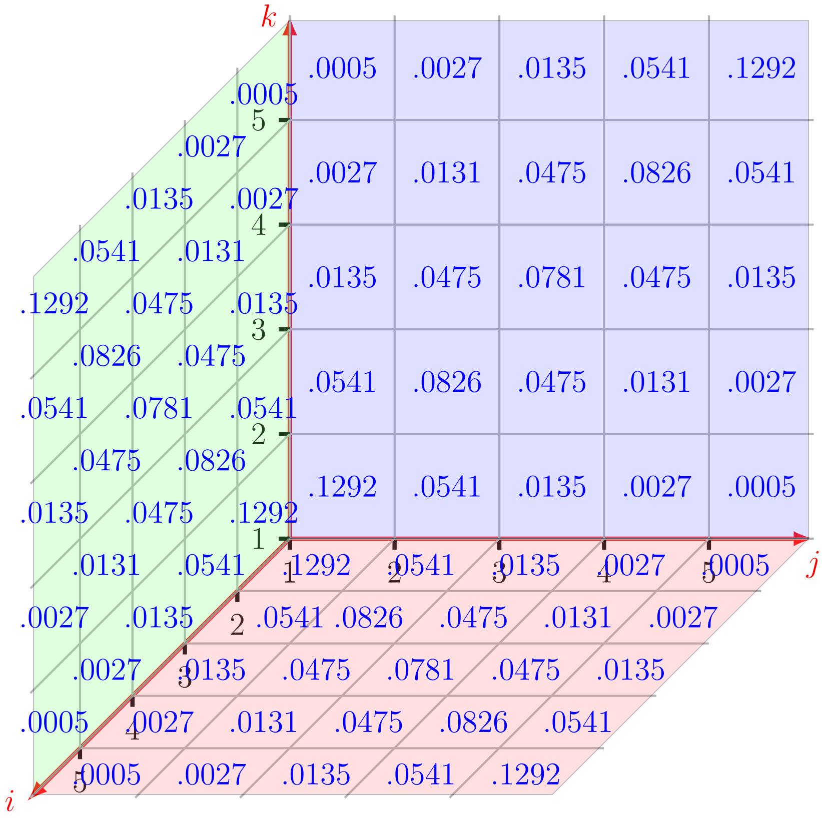


Figure 3.1: The probabilities $h_{.jk}, h_{i.k}$ and $h_{ij.}$, for the high correlation case.

also symmetry around these large values on each side, with similar results observed on all three sides. This effect is due to the copula having one dependence parameter that controls the dependence structure between the three variables. The figures show positive dependence on each side, but the case of high correlation shows larger values of h_{ijk} when $i = j = k$ than when there is no correlation.

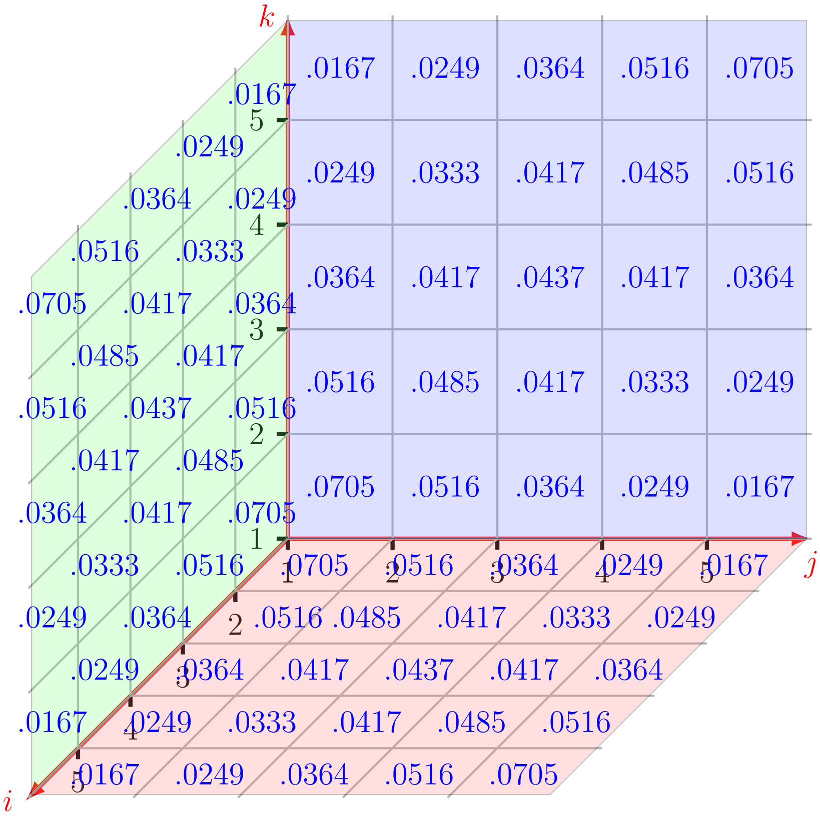


Figure 3.2: The probabilities $h_{.jk}, h_{i.k}$ and $h_{ij.}$, for the no correlation case.

Equivalently, the method of combining NPI with multivariate data can be extended to more than three dimensions by following the same two main steps above, namely, applying NPI for the marginals in the first step and assuming a parametric copula in the second step and estimate the parameter to take the dependence structure into account as follows. Assume that there are n observations of d multivariate random variables $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_d}$ where $\mathbf{X_i} = (X_{1,i}, X_{2,i}, \ldots, X_{d,i})$, $i = 1, \ldots, n$. We are interested in making inferences involving one future multivariate observation, denoted by $(X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1})$.

Using Hill's assumption $A_{(n)}$, it is possible to derive a partially specified predictive probability distribution for each of $X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1}$ given their observations $\mathbf{x_i} = \mathbf{x_1}$, $\mathbf{x_2}, \ldots$, $\mathbf{x_d}$, respectively, where $\mathbf{x_i} = (x_{1,i}, \ldots, x_{d,i})$. These are as follows

$P(X_{1,n+1} \in (x_{1,i_1-1}, x_{1,i_1})) = \frac{1}{n+1}$, $P(X_{2,n+1} \in (x_{2,i_2-1}, x_{2,i_2})) = \frac{1}{n+1}$ and $P(X_{d,n+1} \in (x_{d,i_d-1}, x_{d,i_d})) = \frac{1}{n+1}$ for $i_1, i_2, \ldots, i_d = 1, \ldots, n+1$, where $x_{1,0}, x_{2,0}, \ldots, x_{d,0} = -\infty$ and $x_{1,n+1}, x_{2,n+1}, \ldots, x_{d,n+1} = \infty$ for simplicity.

The two steps can be linked by introducing a natural transformation of the random

quantities individually. Let $\tilde{X}_{1,n+1}, \tilde{X}_{2,n+1}, \ldots, \tilde{X}_{d,n+1}$ be the transformed versions of the random quantities $X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1}$, respectively, such that

$$(X_{1,n+1} \in (x_{1,i_1-1}, x_{1,i_1}), X_{2,n+1} \in (x_{2,i_2-1}, x_{2,i_2}), \ldots, X_{d,n+1} \in (x_{d,i_d-1}, x_{d,i_d})) \Longleftrightarrow$$
$$(\tilde{X}_{1,n+1} \in (\frac{i_1-1}{n+1}, \frac{i_1}{n+1}), \tilde{X}_{2,n+1} \in (\frac{i_2-1}{n+1}, \frac{i_2}{n+1}), \ldots, \tilde{X}_{d,n+1} \in (\frac{i_d-1}{n+1}, \frac{i_d}{n+1})) \tag{3.2.11}$$

for $i_1, i_2, \ldots, i_d \in 1, \ldots, n+1$. The assumption $A_{(n)}$ of the transformations leads to

$$P(\tilde{X}_{1,n+1} \in (\frac{i_1-1}{n+1}, \frac{i_1}{n+1})) = P(X_{1,n+1} \in (x_{1,i_1-1}, x_{1,i_1})) = \frac{1}{n+1} \tag{3.2.12}$$

$$P(\tilde{X}_{2n+1} \in (\frac{i_2-1}{n+1}, \frac{i_2}{n+1})) = P(X_{2,n+1} \in (x_{2,i_2-1}, x_{2,i_2})) = \frac{1}{n+1} \tag{3.2.13}$$

$$P(\tilde{X}_{d,n+1} \in (\frac{i_d-1}{n+1}, \frac{i_d}{n+1})) = P(X_{d,n+1} \in (x_{d,i_d-1}, x_{d,i_d})) = \frac{1}{n+1} \tag{3.2.14}$$

This transformation from the space $\mathbb{R}^d$ to $[0,1]^d$ is based on n trivariate data, where $[0,1]^d$ is divided into $(n+1)^d$ equal sized blocks. By following these transformations of the marginals, the uniform marginal distribution on $[0,1]$ has been discretized. Following these transformations of the marginals, the uniform marginal distributions is discretized on $[0,1]$, which is fully correspond to copulas, as any copula will result in the same discretized uniform marginal distributions.

For the second step when assuming a parametric copula and estimate the parameter, where the observed pairs are replaced by $(\frac{r_i^{x_1}}{n+1}, \ldots, \frac{r_i^{x_d}}{n+1})$ where $r_i^{x_j}$ the rank of the observation x_i among n-x_j observations. NPI on the marginals is now combined with the estimated copula to provide a partially specified predictive distribution for one future multivariate observation and each $(n+1)^d$ blocks is assigned a specific probability as

$$h_{i_1 i_2 \ldots i_d}(\hat{\theta}) = P_C(\tilde{X}_{1,n+1} \in (\frac{i_1-1}{n+1}, \frac{i_1}{n+1}), \tilde{X}_{2,n+1} \in (\frac{i_2-1}{n+1}, \frac{i_2}{n+1}), \ldots,$$
$$\tilde{X}_{d,n+1} \in (\frac{i_d-1}{n+1}, \frac{i_d}{n+1}) | \hat{\theta}) \tag{3.2.15}$$

where $i_1, i_2, \ldots, i_d \in 1, \ldots, n+1$. $P_C(.|\hat{\theta})$ is the assumed copula-based probability and $\hat{\theta}$ is the estimated parameter value and the corresponding cumulative distribution function,

$$H_{i_1 i_2 \ldots i_d}(\hat{\theta}) = P_C(\tilde{X}_{1,n+1} \leq \frac{i_1}{n+1}, \tilde{X}_{2,n+1} \leq \frac{i_2}{n+1}, \ldots, \tilde{X}_{d,n+1} \leq \frac{i_d}{n+1} | \hat{\theta}) \tag{3.2.16}$$

These $(n+1)^d$ values of $h_{i_1 i_2 \ldots i_d}(\hat{\theta})$ provide the fully discretized probability distribution for the transformed future observations, which can be used for statistical inference on the future observation or any event of interest involving the future observation. These $h_{i_1 i_2 \ldots i_d}(\hat{\theta})$ probabilities satisfy the following conditions

1. $\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_d=1}^{n} h_{i_1 i_2 \ldots i_d} = 1$
2. $\sum_{i_2=1}^{n} \cdots \sum_{i_d=1}^{n} h_{i_1 i_2 \ldots i_d} = \frac{1}{n+1}$, for all $i_1 \in \{\ 1, 2, \ldots, n+1\}$ this summation condition is repeated for each marginal by fixing a different index and summing over the others.
3. $h_{i_1 i_2 \ldots i_d} \geq 0$, for all $i_1, i_2, \ldots, i_d \in \{\ 1, \ldots, n+1\}$

3.3 Combining NPI with a nonparametric copula

Combining NPI with a nonparametric bivariate copula was first introduced by Muhammad in [76]. This section presents an extension of Muhammad's work by applying NPI to the marginals with a nonparametric trivariate copula. In general, the idea is similar to that presented in Section 3.2, which consists of two main steps: applying NPI for the marginals for the first step and assuming a nonparametric trivariate copula in the second step. Let $(X_{n+1}, Y_{n+1}, Z_{n+1})$ be a future trivariate observation and $\tilde{X}_{n+1}$, $\tilde{Y}_{n+1}$ and $\tilde{Z}_{n+1}$ be the transformation versions of the random quantities X_{n+1}, Y_{n+1} and Z_{n+1}, respectively, following from the natural transformations related to the marginal $A_{(n)}$ assumptions as presented in Section 3.2. For the second step, a kernel-based copula is assumed, with an estimated probability density function defined as:

$$\hat{c}(x, y, z) = \frac{1}{n b_X b_Y b_Z} \sum_{i=1}^{n} K\left(\frac{x - F_X(\tilde{X}_i)}{b_X}, \frac{y - F_Y(\tilde{Y}_i)}{b_Y}, \frac{z - F_Z(\tilde{Z}_i)}{b_Z}\right) \tag{3.3.1}$$

where $K : \mathbb{R}^3 \to \mathbb{R}$ is a trivariate kernel function, $b_X, b_Y, b_Z > 0$ are the bandwidths or smoothing parameters, $F_X(\tilde{X}_i) = \frac{r_i^x}{n+1}$, $F_Y(\tilde{Y}_i) = \frac{r_i^y}{n+1}$ and $F_Z(\tilde{Z}_i) = \frac{r_i^z}{n+1}$. Now, the NPI approach for the marginals can be combined with the nonparametric kernel-based copula Equation (3.3.1) to take the dependence into account as follows

$$h_{ijk}(\hat{c}) = P_C(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1}) | \hat{c}) \tag{3.3.2}$$

where i, j and $k \in \{ 1, \ldots, n+1\}$, $P_C(.|\hat{c})$ is the nonparametric kernel-based copula probability and $\hat{c}$ is the estimated kernel density function. The cumulative distribution function takes the form

$$H_{ijk}(\hat{c}) = P_C\left(\tilde{X}_{n+1} \leq \frac{i}{n+1}, \tilde{Y}_{n+1} \leq \frac{j}{n+1}, \tilde{Z}_{n+1} \leq \frac{k}{n+1} \,|\, \hat{c}\right) \tag{3.3.3}$$

Note that the $h_{ijk}(\hat{c})$ values must satisfy the three conditions presented after Equation (3.2.5) in Section 3.2.

Similarly, the method of combining NPI with nonparametric copula can be extended to more than three dimensions and is similar to the method discussed in Section 3.2, by following the same two main steps: applying NPI for the marginals in the first step and assuming a nonparametric copula in the second step. Then, combining these steps together to define the probabilities $h_{i_1 i_2 \ldots i_d}$ as follows. Let $(X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1})$ be a future multivariate observation and let $\tilde{X}_{1,n+1}, \tilde{X}_{2,n+1}, \ldots, \tilde{X}_{d,n+1}$ be the transformed versions of the random quantities $X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1}$ respectively, following from the natural transformations related to the marginal $A_{(n)}$ assumptions. For the second step when assuming a kernel-based copula, Equation (2.3.10), the estimated probability density function defined as

$$\hat{c}(\mathbf{x}) = \frac{1}{n b_{X_1} \ldots b_{X_d}} \sum_{i=1}^{n} K\left(\frac{x_1 - F_{X_1}(\tilde{X}_{1,i})}{b_{X_1}}, \frac{x_2 - F_{X_2}(\tilde{X}_{2,i})}{b_{X_2}}, \ldots, \frac{x_n - F_{X_n}(\tilde{X}_{d,i})}{b_{X_d}}\right) \tag{3.3.4}$$

where $K : \mathbb{R}^d \to \mathbb{R}$ is a trivariate kernel function, $\mathbf{b} = b_{X_1}, b_{X_2}, \ldots, b_{X_d} > 0$ are the bandwidths, $F_{X_1}(\tilde{X}_{1i}) = \frac{r_i^{x_1}}{n+1}$, $F_{X_2}(\tilde{X}_{2i}) = \frac{r_i^{x_2}}{n+1}$,..., $F_{X_d}(\tilde{X}_{di}) = \frac{r_i^{x_d}}{n+1}$. Now, the NPI approach for the marginals, can be combined with the nonparametric kernel-based copula to take the dependence into account as follows

$$\begin{aligned} &h_{i_1 i_2 \ldots i_n}(\hat{c}) = \\ &P_C(\tilde{X}_{1,n+1} \in (\frac{i_1 - 1}{n+1}, \frac{i_1}{n+1}), \tilde{X}_{2,n+1} \in (\frac{i_2 - 1}{n+1}, \frac{i_2}{n+1}), \ldots, \tilde{X}_{d,n+1} \in (\frac{i_d - 1}{n+1}, \frac{i_d}{n+1}) | \hat{c}) \end{aligned} \tag{3.3.5}$$

where $i_1, i_2, \ldots, i_d \in 1, \ldots, n+1$. Further, the corresponding cumulative distribution function is

$$H_{i_1 i_2 \ldots i_d}(\hat{c}) = P_C(\tilde{X}_{1,n+1} \leq \frac{i_1}{n+1}, \tilde{X}_{2,n+1} \leq \frac{i_2}{n+1}, \ldots, \tilde{X}_{d,n+1} \leq \frac{i_d}{n+1}) | \hat{c}) \tag{3.3.6}$$

The probabilities $h_{i_1 i_2 \ldots i_d}(\hat{c})$ must satisfy the conditions as presented in Section 3.2 after Equation (3.2.13).

3.4 Examples

This section illustrates the proposed methods—combining NPI with a parametric copula as in Section 3.2 and with a nonparametric copula as in Section 3.3 using trivariate Gaussian datasets with zero mean vectors. Three covariance matrices representing different correlation structures are considered, with sample sizes $n = 10, 25, 50$. These covariance matrices are as follows:

$$\begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.15 & 0.15 \\ 0.15 & 1 & 0.15 \\ 0.15 & 0.15 & 1 \end{pmatrix}$$

The dependence strength is defined as High (H), Moderate (M), and Low (L) corresponding to assumed correlations of 0.9, 0.5 and 0.15, respectively. First four parametric copula types are used: Clayton, Gumbel, Frank and Joe, as presented in Section 2.3. The pseudo maximum likelihood method, commonly used to estimate copula parameters, is implemented via the R package *copula* [100].

Following the method presented in Section 3.2, which assumes a trivariate copula to take the dependence structure into account between the variables, the estimated parameters for the four assumed trivariate copulas are presented in Table 3.2. Table 3.2 shows that the Gumbel copula consistently has the highest Kendall's τ value of 0.84 across all the copulas considered, regardless of the correlation level. When the sample size is 50, both the Frank and Gumbel copulas yield the same Kendall's τ value. It is clear that the parameter estimates tend to increase with higher correlation in the generated data, as expected.

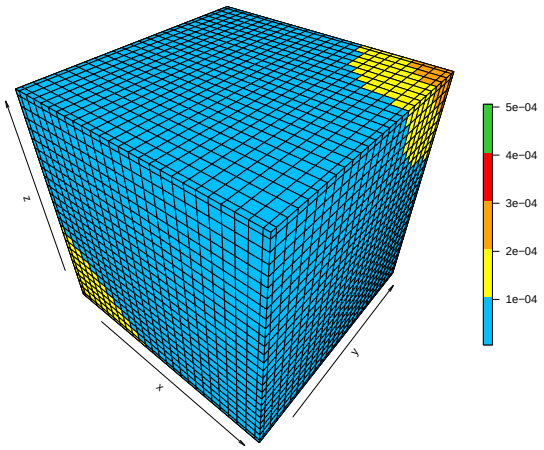
With the estimated parameters and the assumed trivariate copulas, the probabilities $h_{ijk}(\hat{\theta})$ are obtained according to Equation (3.2.5). Figure 3.3 presents a three-dimensional plot of the probabilities $h_{ijk}(\hat{\theta})$ for Frank trivariate copulas, considering different correlation levels and sample sizes. Generally, it can be seen from these figures that the probabilities $h_{ijk}(\hat{\theta})$ are similar but not identical in each block. Some of these figures

τ	n	Clayton		Gumbel		Frank		Joe	
		τ	$\hat{\theta}_c$	τ	$\hat{\theta}_g$	τ	$\hat{\theta}_f$	τ	$\hat{\theta}_j$
	10	0.77	6.23	0.84	6.15	0.79	16.81	0.81	9.10
H	25	0.61	3.15	0.77	4.28	0.76	14.54	0.72	6.04
	50	0.66	3.95	0.73	3.77	0.73	12.75	0.66	4.71
	10	0.38	1.20	0.46	1.85	0.41	4.37	0.41	2.28
M	25	0.36	1.13	0.49	1.95	0.47	5.16	0.44	2.43
	50	0.34	1.01	0.38	1.61	0.38	3.94	0.31	1.81
	10	0.23	0.60	0.30	1.43	0.22	2.07	0.28	1.68
L	25	0.22	0.57	0.24	1.31	0.23	2.17	0.20	1.44
	50	0.13	0.29	0.13	1.15	0.13	1.15	0.10	1.20

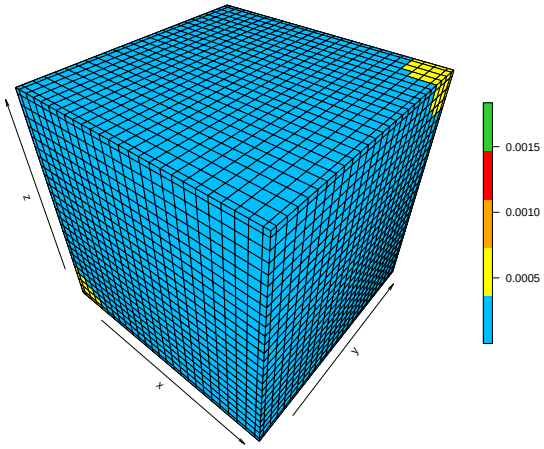
Table 3.2: Estimated parameters and corresponding Kendall's τ values from simulated data with varying sample sizes, correlation levels, and copula types.

show little difference in the probabilities $h_{ijk}(\hat{\theta})$, which is particularly apparent in the top right panel for smaller sample sizes.

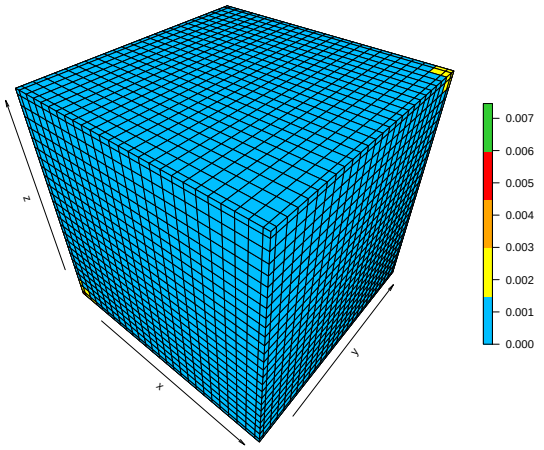
As explained in Example 3.2.1, the marginals of the probabilities $h_{ijk}(\hat{\theta})$ show a positive relationship pattern when the generated data is highly correlated. So for all these copulas, highly correlated data leads to h_{ijk} probabilities having larger values when i, j and k are close to each other compared to the situation when the data is moderately correlated. When the data is weakly correlated, the probabilities $h_{ijk}(\hat{\theta})$ will still show a positive pattern when i, j and k are closer, but this pattern is weaker compared to when the data is more correlated. Similar results are also achieved when assuming either Clayton, Gumbel or Joe copulas. The detailed results are presented in Appendix A, Figures A.1-A.6.



(a) Low correlated



(b) Moderate correlated



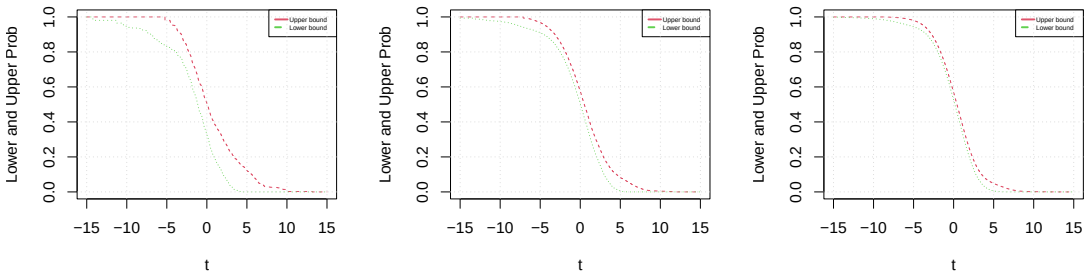
(c) High correlated

Figure 3.3: The h_{ijk} probabilities obtained from simulated data $n = 25$ using Frank copulas with different correlation levels.

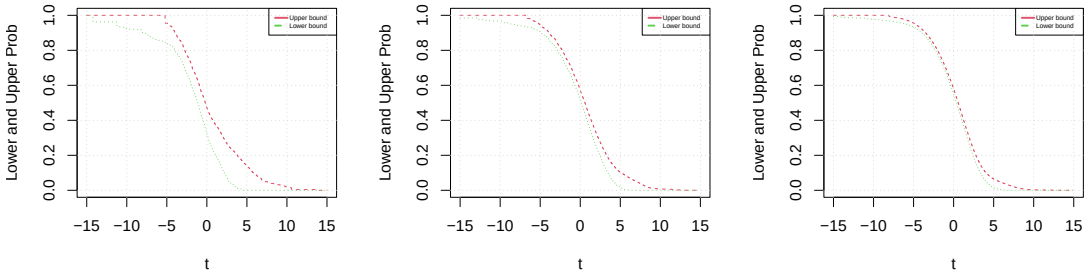
Assume that the event of interest is $T_{n+1} = X_{n+1}+Y_{n+1}+Z_{n+1} > t$. The NPI lower and upper probabilities from Equations (3.2.9) and (3.2.10) are shown in Figures 3.4 and 3.5 for the Clayton and Gumbel copulas, respectively. These figures show the results across different correlation levels and sample sizes. The NPI lower and upper probabilities for selected values of t are presented in Table 3.3. The imprecision, which is the difference between the upper and lower probabilities, is larger with a positive correlation than when there is a weak correlation.

This is because the event $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$ is closely linked to the probabilities $h_{ijk}(\hat{\theta})$, which is fundamental for inference. In positively correlated data, large values of $h_{ijk}(\hat{\theta})$ tend to occur when i, j, and k are are close to each other. Calculating the lower and upper probabilities, as in Equations (3.2.9) and (3.2.10), tends to include more $h_{ijk}(\hat{\theta})$ values and with the event $T_{n+1} > t$, theses values are included in the lower and upper probabilities for most values of t. As a result, imprecision remains small across most t values, as shown in Figures 3.4 and 3.5. Similar patterns are observed for Frank and Joe copulas

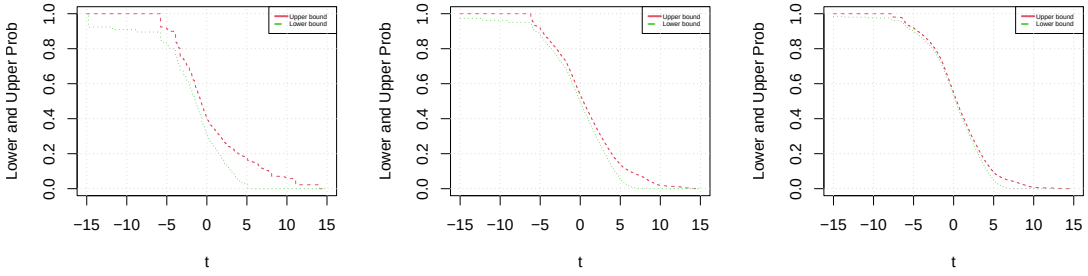
For all copulas, high positive correlation results in consistent imprecision across t values, while low correlation causes greater imprecision in the tails and less at the center. This pattern occurs at similar t values across copula types. Imprecision at the center decreases as correlation increases. The results also show that sample size affects the NPI lower and upper probabilities. Overall, imprecision remains small with large datasets, regardless of the copula type.



(a) Low,$n = 10$ (b) Low,$n = 25$ (c) Low,$n = 50$

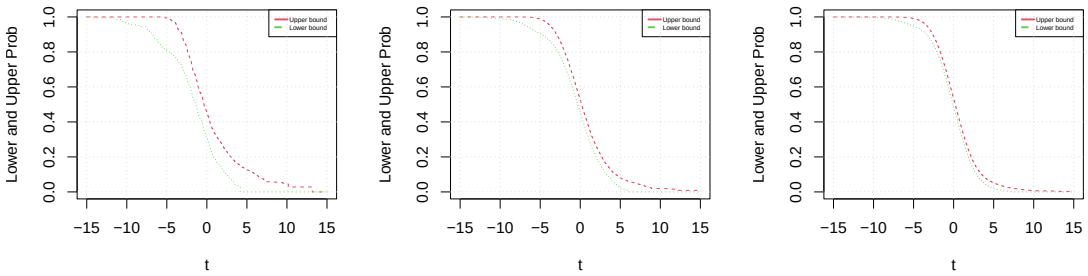


(d) Moderate, $n = 10$ (e) Moderate, $n = 25$ (f) Moderate, $n = 50$

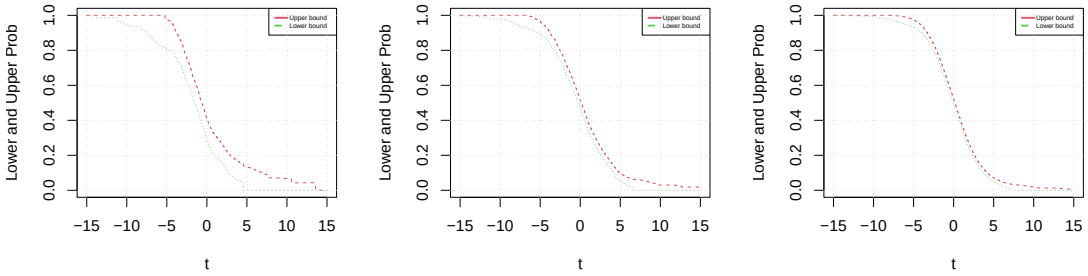


(g) High, $n = 10$ (h) High, $n = 25$ (i) High, $n = 50$

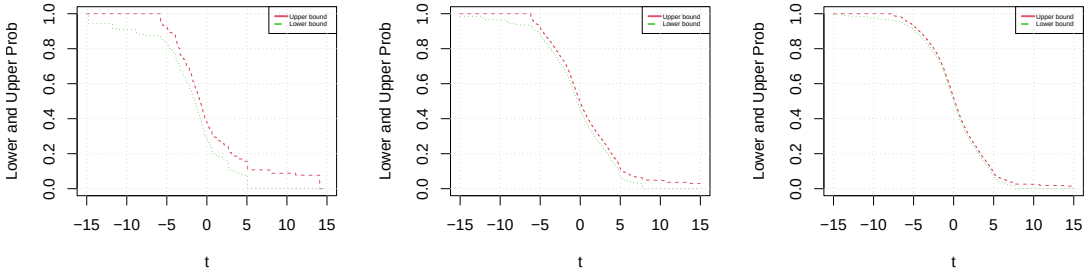
Figure 3.4: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and Clayton copula.



(a) Low, $n = 10$ (b) Low, $n = 25$ (c) Low, $n = 50$



(d) Moderate, $n = 10$ (e) Moderate, $n = 25$ (f) Moderate, $n = 50$



(g) High, $n = 10$ (h) High, $n = 25$ (i) High, $n = 50$

Figure 3.5: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and Gumbel copula.

τ	n	t	Clayton		Gumbel		Frank		Joe	
			$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
		-6.00	0.8951	1.0000	0.8752	1.0000	0.8579	1.0000	0.8359	1.0000
	10	-3.00	0.6721	0.7564	0.6671	0.7604	0.6655	0.7572	0.6584	0.7527
		0.00	0.3120	0.4058	0.2881	0.3809	0.2957	0.3871	0.2873	0.3789
H		-6.00	0.9266	0.9625	0.9234	0.9707	0.9120	0.9800	0.9030	0.9925
	25	-3.00	0.7625	0.7996	0.7502	0.7919	0.7448	0.7844	0.7337	0.7818
		0.00	0.4961	0.5363	0.4524	0.4932	0.4570	0.4956	0.4449	0.4843
		-6.00	0.9305	0.9489	0.9377	0.9614	0.9300	0.9665	0.9372	0.9866
	50	-3.00	0.8193	0.8393	0.8109	0.8337	0.8025	0.8241	0.7941	0.8257
		0.00	0.5279	0.5476	0.4985	0.5194	0.5043	0.5240	0.4837	0.5037
		-6.00	0.8573	1.0000	0.8316	1.0000	0.8242	1.0000	0.8286	1.0000
	10	-3.00	0.7274	0.8314	0.6868	0.8263	0.6833	0.8192	0.6592	0.8428
		0.00	0.3221	0.4726	0.2917	0.4097	0.3145	0.4370	0.2789	0.3908
M		-6.00	0.9280	0.9737	0.9151	0.9904	0.9082	0.9945	0.9191	0.9983
	25	-3.00	0.8122	0.8542	0.7869	0.8457	0.7745	0.8333	0.7763	0.8621
		0.00	0.5142	0.5707	0.4667	0.5157	0.4855	0.5306	0.4391	0.4876
		-6.00	0.9511	0.9745	0.9487	0.9898	0.9453	0.9914	0.9582	0.9968
	50	-3.00	0.8505	0.8720	0.8423	0.8789	0.8269	0.8640	0.8515	0.9029
		0.00	0.5500	0.5793	0.4929	0.5221	0.5116	0.5375	0.4636	0.4961
		-6.00	0.8593	1.0000	0.8510	1.0000	0.8546	1.0000	0.8600	1.0000
	10	-3.00	0.7386	0.8733	0.7000	0.8776	0.6985	0.8843	0.6831	0.8947
		0.00	0.3281	0.5072	0.3105	0.4546	0.3178	0.4822	0.2947	0.4347
L		-6.00	0.9262	0.9856	0.9292	0.9975	0.9260	0.9981	0.9399	0.9992
	25	-3.00	0.8332	0.8888	0.8165	0.9023	0.8077	0.8941	0.8192	0.9201
		0.00	0.5040	0.5748	0.4548	0.5257	0.4726	0.5399	0.4330	0.5101
		-6.00	0.9567	0.9901	0.9640	0.9972	0.9633	0.9974	0.9688	0.9984
	50	-3.00	0.8795	0.9153	0.8788	0.9292	0.8736	0.9249	0.8849	0.9393
		0.00	0.5249	0.5675	0.4868	0.5318	0.4974	0.5406	0.4766	0.5252

Table 3.3: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and copula types.

To illustrate the method of combining NPI with a nonparametric trivariate copula. Two bandwidth methods are used: the normal reference rule-of-thumb and the LSCV methods, which are available in the *np* package in R [47] and presented in Section 2.3. Following the proposed method presented in Section 3.3, where NPI is applied for the marginals and the bandwidth values are selected to obtain the probabilities $h_{ijk}(\hat{c})$ as defined in Equation (3.3.2).

The bandwidth values are shown in Table 3.4, which demonstrates that using the normal reference rule-of-thumb method gives a constant value regardless of the correlation strength of the three random quantities. This is because it is a fixed type, due to the normal reference rule-of-thumb bandwidth equation, as presented in Section 2.3, Equation (2.3.8). This bandwidth value decreases as the sample size increases, while the LSCV method gives the smallest bandwidth values compared to the normal reference rule-of-thumb method. This method depends on minimizing the integrated squared error. In general, the bandwidth values using the LSCV method decrease as the correlation level and the sample size increase, as shown in Table 3.4. Using these bandwidth values and an assumed correlation of the generated data leads to different probabilities $h_{ijk}(\hat{c})$, as seen in Figure 3.6.

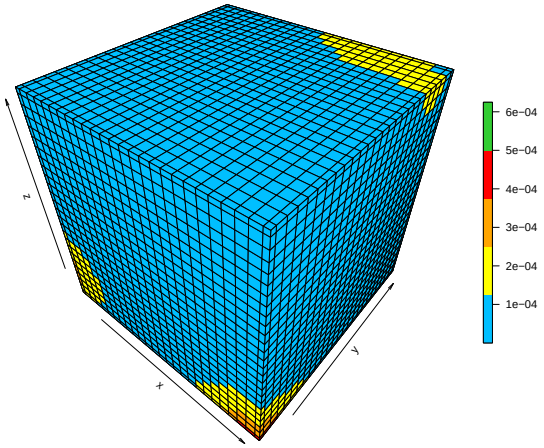
Figure 3.6 presents the probabilities $h_{ijk}(\hat{c})$ using the normal reference rule-of-thumb. Applying the normal reference rule-of-thumb with a small sample size shows similar probabilities, except that the left corner has higher probabilities. As the correlation increases, the probabilities become higher in both the top right and bottom left corners. Using the LSCV method shows that some probabilities are higher in certain blocks. Increasing the correlation causes the probabilities $h_{ijk}(\hat{c})$ to become similar, except in the top right and bottom left corners, where they are slightly higher. For high positive correlation, the bandwidth decreases and the probabilities $h_{ijk}(\hat{c})$ include larger values when i, j and k are closer. In contrast, when the data is weakly correlated, the probabilities $h_{ijk}(\hat{c})$ are highly scattered. The remaining results of probabilities $h_{ijk}(\hat{c})$ are presented in Appendix A, Figures A.7-A.9.

τ	n	Normal Reference	LSCV		
		$\boldsymbol{b}$	b_X	b_Y	b_Z
	10	0.203	0.103	0.063	0.086
H	25	0.154	0.037	0.052	0.061
	50	0.123	0.030	0.047	0.059
	10	0.203	0.104	0.040	0.251
M	25	0.123	0.100	0.039	0.062
	50	0.123	0.100	0.039	0.062
	10	0.203	0.223	0.001	0.232
L	25	0.154	0.161	0.091	0.084
	50	0.123	0.084	0.077	0.102

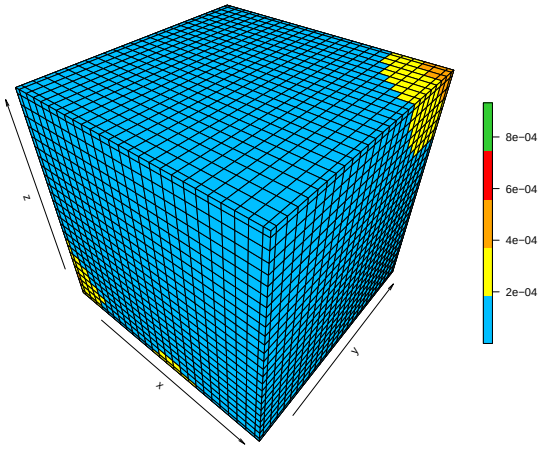
Table 3.4: The bandwidth values from simulated data using different sample sizes, correlation levels, and bandwidth types, $\boldsymbol{b} = b_X = b_Y = b_Z$

Then, the NPI lower and upper probabilities for the event of interest are presented in Figures 3.7 and 3.8 and Table 3.5. This table shows how the results vary based on the sample size, the correlation between variables, and the chosen bandwidth method for selected values of t. Generally, Table 3.5 It also indicates that increasing the sample size results in less imprecision. Additionally, the imprecision decreases when the generated data are more highly correlated compared to when they are weakly correlated.

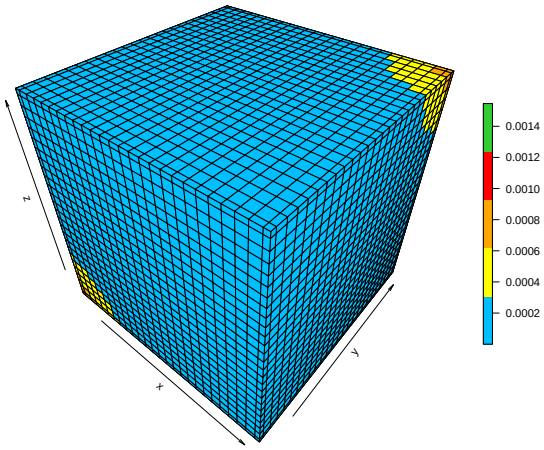
For the event of interest $T_{n+1} > t$, the NPI lower and upper probabilities are typically include several additional probabilities of $h_{ijk}(\hat{c})$. With a high positive correlation, these additional probabilities $h_{ijk}(\hat{c})$ include a few larger values for most values of t, compared to a weak correlation.



(a) Low correlated

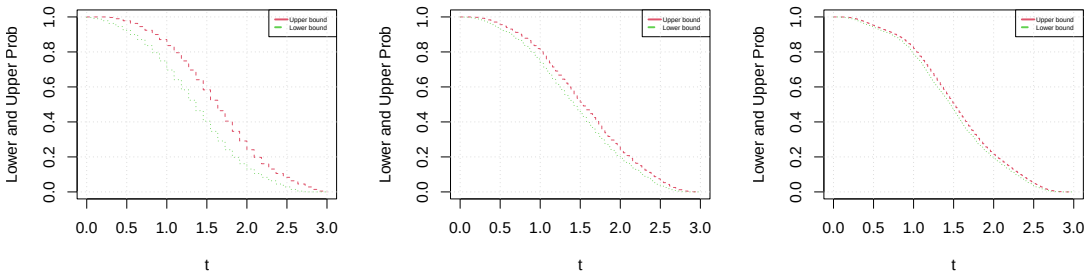


(b) Moderate correlated

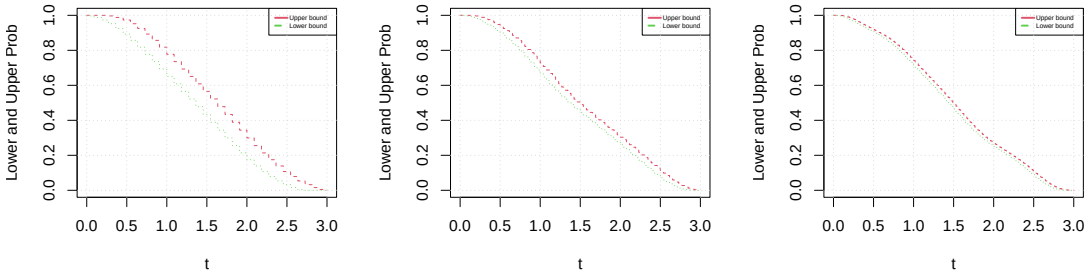


(c) High correlated

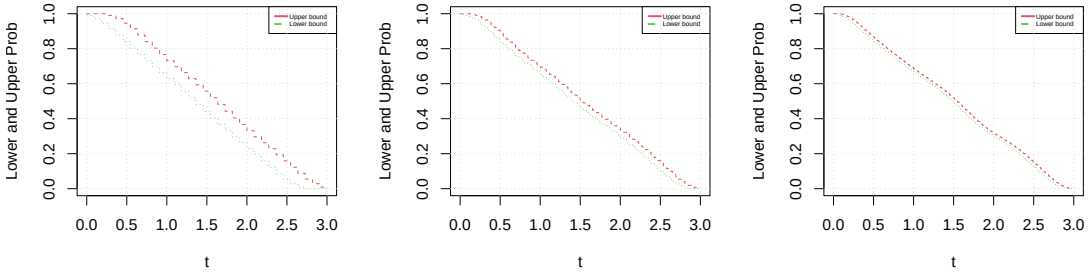
Figure 3.6: The h_{ijk} probabilities, obtained from simulated data $n = 25$ using different correlations and normal reference rule-of-thumb.



(a) Low, $n = 10$ (b) Low, $n = 25$ (c) Low, $n = 50$

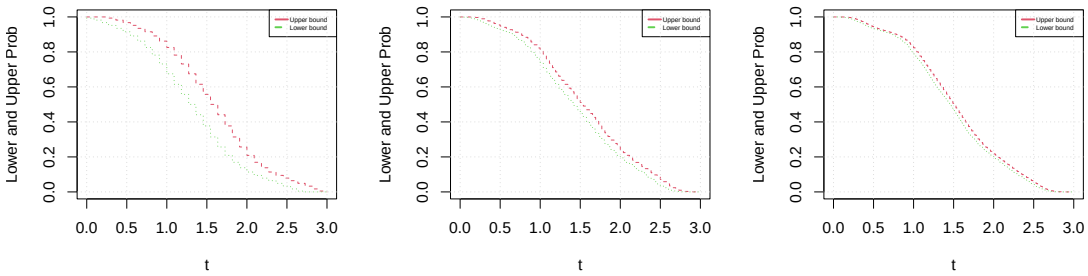


(d) Moderate, $n = 10$ (e) Moderate, $n = 25$ (f) Moderate, $n = 50$

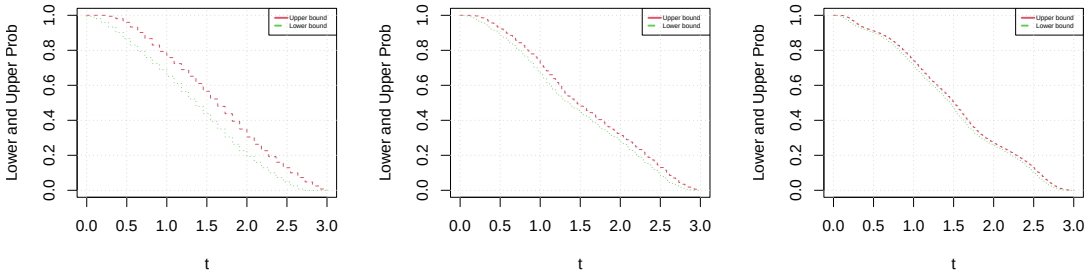


(g) High, $n = 10$ (h) High, $n = 25$ (i) High, $n = 50$

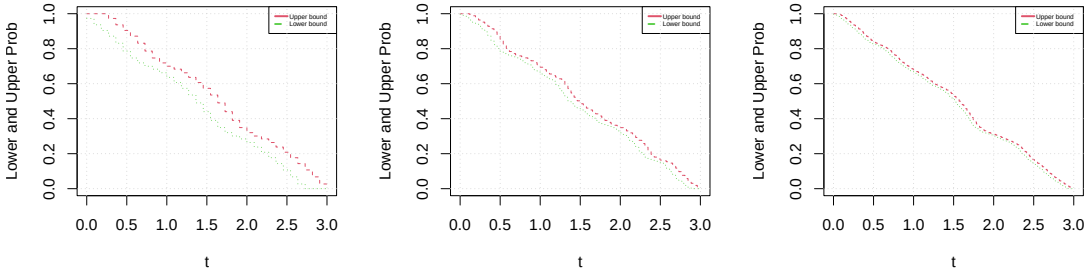
Figure 3.7: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and normal reference rule-of-thumb bandwidth.



(a) Low, $n = 10$ (b) Low, $n = 25$ (c) Low, $n = 50$



(d) Moderate, $n = 10$ (e) Moderate, $n = 25$ (f) Moderate, $n = 50$



(g) High, $n = 10$ (h) High, $n = 25$ (i) High, $n = 50$

Figure 3.8: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and LSCV bandwidth.

τ	n	t	Normal Reference		LSCV	
			$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
		1.00	0.6290	0.7314	0.6363	0.7000
	10	1.50	0.4425	0.5574	0.4401	0.5736
		2.00	0.2285	0.3312	0.2639	0.3201
H		1.00	0.6538	0.6943	0.6554	0.6942
	25	1.50	0.4645	0.5076	0.4518	0.4896
		2.00	0.2959	0.3353	0.3190	0.3488
		1.00	0.6663	0.6853	0.6642	0.6790
	50	1.50	0.4945	0.5191	0.5093	0.5358
		2.00	0.3000	0.3168	0.3009	0.3111
		1.00	0.6508	0.7772	0.6519	0.7584
	10	1.50	0.4335	0.5653	0.4377	0.5656
		2.00	0.1748	0.2992	0.1924	0.3046
M		1.00	0.6682	0.7286	0.6613	0.7253
	25	1.50	0.4477	0.4923	0.4457	0.4829
		2.00	0.2611	0.3037	0.2763	0.3135
		1.00	0.7132	0.7416	0.7094	0.7390
	50	1.50	0.4702	0.5009	0.4720	0.5058
		2.00	0.2514	0.2695	0.2556	0.2700
		1.00	0.6966	0.8363	0.6836	0.8303
	10	1.50	0.4044	0.5833	0.4013	0.5713
		2.00	0.1307	0.2410	0.1208	0.2272
L		1.00	0.7377	0.8015	0.7357	0.8038
	25	1.50	0.4516	0.5176	0.4524	0.5160
		2.00	0.1926	0.2411	0.1951	0.2414
		1.00	0.7872	0.8194	0.7887	0.8235
	50	1.50	0.4681	0.5079	0.4662	0.5063
		2.00	0.1921	0.2146	0.1970	0.2176

Table 3.5: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and bandwidth.

3.5 Predictive performance

This section presents the results of a simulation study conducted to evaluate the predictive performance of the methods proposed in Sections 3.2 and 3.3. For each method, $N = 100$ datasets of size $n+1$ are generated. For each generated dataset, the first n observations are used for the proposed methods, and the last observation is used to evaluate the predictive performance. Sample sizes of $n = 10$ and $n = 25$ are used to avoid the long computation times associated with larger datasets

In this study, the focus is on the sum of the next observations $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1}$. Let (x_i^j, y_i^j, z_i^j) denote the ith observation in the jth simulated sample, where $i = 1, 2, \ldots, n$ refers to the observation within a sample, and $j = 1, 2, \ldots, N$ indexes the simulated samples. Let (x_f^j, y_f^j, z_f^j) be the extra simulated pair that will be used for the evaluation and let the corresponding sum denoted by $t_f^j = x_f^j + y_f^j + z_f^j$. Recall from the lower and upper probabilities for the event of interest $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1}$, Equations (3.2.9) and (3.2.10) as presented in Section 3.2 and denote that $\underline{S}(t) = \underline{P}(T_{n+1} > t)$ and $\overline{S}(t) = \overline{P}(T_{n+1} > t)$. The inverse values of these lower and upper survival functions of T_{n+1}, for a value $q \in (0, 1)$, are defined as

$$\underline{t}_q^j = \inf_{t \in \mathbb{R}} \{ \underline{S}(t) \leq q \} \tag{3.5.7}$$

$$\overline{t}_q^j = \inf_{t \in \mathbb{R}} \{ \overline{S}(t) \leq q \} \tag{3.5.8}$$

where $\underline{t}_q \leq \overline{t}_q$ holds. If the two inequalities listed below hold true, then the proposed method works effectively.

$$p_1 = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(t_f^j > t_q^j) \leq q \tag{3.5.9}$$

$$p_2 = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(t_f^j > t_q^j) \geq q \tag{3.5.10}$$

The values $q = 0.25, 0.50, 0.75$ are chosen for performance evaluation, though different quantiles can also be used. Quantiles provide a good indicator of the performance of the proposed method. The simulation is conducted by varying degrees of dependence, using various values of Kendall's (τ) and a variety of sample sizes to determine whether there is an influence on the performance.

The datasets are generated from assumed parametric dependence models, setting the parameter values corresponding to the selected τ values. For comparison, the four commonly used copulas: Clayton, Gumbel, Frank and Joe, are used in this study. The estimated parameters are obtained using the pseudo maximum likelihood estimation method, which is presented in Section 2.3. This method outperforms other estimation techniques, particularly when the marginal distribution is unknown, which is common in many real-world applications, as noted in the literature. Moreover, it is preferred due to its faster computational performance compared to alternative methods.

Predictive performance: parametric copula

In this study, the focus is on the performance evaluation of the method presented in Section 3.2. Four common one-parameter trivariate copulas are used: Clayton, Gumbel, Frank and Joe copulas, which are presented in Section 2.3. In the simulation, Kendall's τ values of $0.25, 0.5$, and 0.75 are used to represent low (L), moderate (M), and high (H) dependence, respectively.

The results are presented in Tables 3.6-3.9, which present the p_1 and p_2 values calculated using Equations (3.5.9) and (3.5.10). For optimal performance, the q value must satisfy $p_1 \leq q \leq p_2$. Tables 3.6-3.9 confirm good performance for all the assumed copulas and the results adhere to the condition $p_1 \leq q \leq p_2$, except for $n = 25$ using the Clayton copula, as highlighted in Table 3.6. One possible explanation is that larger sample sizes reduce imprecision, which may result in some q values falling outside the predicted intervals $[p_1, p_2]$. The tables show that imprecision is relatively large when the sample size is $n = 10$, and it tends to decrease as the sample size increases to $n = 25$. The results indicate that the parameters are well estimated, as the true and estimated values are close to each other, particularly for the sample size of $n = 25$. Thus, the dependence structures are well described, with the corresponding Kendall's tau values of the true and estimated values being close to each other.

τ	θ_C	q	$n = 10$			$n = 25$		
			$\hat{\theta}_C$	p_1	p_2	$\hat{\theta}_C$	p_1	p_2
L	0.6667	0.25	1.1152	0.18	0.31	0.8540	0.22	0.28
		0.50		0.42	0.57		0.48	0.53
		0.75		0.70	0.81		0.73	0.79
M	2	0.25	2.6230	0.17	0.31	2.1949	0.24	0.30
		0.5		0.43	0.55		**0.51**	0.56
		0.75		0.69	0.77		0.75	0.78
H	6	0.25	7.2905	0.22	0.34	5.7748	0.24	0.29
		0.5		0.47	0.57		0.48	0.52
		0.75		0.70	0.78		0.73	0.77

Table 3.6: Simulation study I, Predictive performance, Clayton copula.

In addition, the tables clearly show that imprecision decreases as the correlation among variables increases, regardless of the copula type. This can be attributed to two main factors related to the event $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$, which is explained through the probabilities h_{ijk}, fundamental for inference. For high positive correlation, and with the event defined as the sum, the lower and upper probabilities in Equations (3.2.9) and (3.2.10) tend to include additional $h_{ijk}(\hat{\theta})$. These additional terms generally have large values for most t under positive correlation. Conversely, for weak correlation, the h_{ijk} probabilities where i, j, k are close are not as large as in the high positive correlation case.

τ	θ_G	q	$n = 10$			$n = 25$		
			$\hat{\theta}_G$	p_1	p_2	$\hat{\theta}_G$	p_1	p_2
L	1.33335	0.25	1.6048	0.17	0.28	1.4453	0.24	0.28
		0.5		0.41	0.56		0.47	0.53
		0.75		0.65	0.80		0.70	0.76
M	2	0.25	2.4191	0.20	0.30	2.1749	0.25	0.28
		0.5		0.42	0.52		0.48	0.53
		0.75		0.65	0.79		0.71	0.76
H	4	0.25	4.5301	0.22	0.30	4.0539	0.25	0.28
		0.5		0.43	0.53		0.48	0.53
		0.75		0.65	0.77		0.71	0.76

Table 3.7: Simulation study I, Predictive performance, Gumbel copula.

τ	θ_F	q	$n = 10$			$n = 25$		
			$\hat{\theta}_F$	p_1	p_2	$\hat{\theta}_F$	p_1	p_2
L	2.37193	0.25	3.0163	0.19	0.32	2.5720	0.22	0.26
		0.5		0.42	0.55		0.46	0.52
		0.75		0.67	0.79		0.71	0.77
M	5.736283	0.25	6.4797	0.21	0.31	6.0112	0.24	0.28
		0.5		0.44	0.53		0.48	0.53
		0.75		0.65	0.77		0.74	0.78
H	14.138501	0.25	13.7274	0.21	0.31	13.5046	0.22	0.26
		0.5		0.44	0.52		0.47	0.52
		0.75		0.67	0.78		0.72	0.77

Table 3.8: Simulation study I, Predictive performance, Frank copula.

τ	θ_J	q	$n = 10$			$n = 25$		
			$\hat{\theta}_J$	p_1	p_2	$\hat{\theta}_J$	p_1	p_2
L	1.596108	0.25	2.0293	0.23	0.31	1.7708	0.22	0.27
		0.5		0.43	0.58		0.48	0.54
		0.75		0.65	0.82		0.71	0.78
M	2.856257	0.25	2.4191	0.20	0.30	3.0676	0.22	0.26
		0.5		0.42	0.52		0.48	0.53
		0.75		0.65	0.79		0.74	0.78
H	6.782365	0.25	7.6760	0.20	0.30	6.5317	0.22	0.26
		0.5		0.46	0.58		0.49	0.52
		0.75		0.68	0.79		0.72	0.76

Table 3.9: Simulation study I, Predictive performance, Joe copula.

Predictive performance: nonparametric copula

This study focuses on evaluating the method presented in Section 3.3 using the approach discussed in Section 3.5, with particular attention to bandwidth selection using the LSCV method, as the normal reference rule-of-thumb produces similar results. By using different quantiles to investigate the method's performance, one can choose any value of q. In this study, the chosen values are $q = 0.25, 0.50, 0.75$, which offer valuable insight into the performance of the proposed method, as tested using Equations (3.5.9) and (3.5.10). Table 3.14 presents the average bandwidth values b_x, b_y and b_z over 100 runs using the LSCV bandwidth method. It also shows the corresponding copula types with the assumed Kendall's τ values used for simulation. In general, the bandwidth values decrease as the level of dependence increases, regardless of the copula type, and they are smaller for $n = 25$. Simulations based on the Clayton copula with low correlation show larger bandwidth values compared to other copulas, due to the Clayton copula's characteristic lower tail dependence.

Tables 3.10-3.13 present the predictive performance results using the nonparametric copula with the LSCV bandwidth method. Table 3.10 shows the predictive performance of the proposed method and the generated data is from the trivariate Clayton copula. It

shows that in most cases, q is not included between p_1 and p_2, regardless of the sample size or the strength of dependence. In the table, bold font numbers indicate these values. Simulating data from a one-parameter trivariate Clayton copula exhibits lower tail dependence. Choosing bandwidth using the LSCV method leads the probabilities $h_{ijk}(\hat{c})$ to be strongly spread out. As the event of interest is on the sum, the lower and upper probabilities tend to include extra values of the probabilities $h_{ijk}(\hat{c})$.

Table 3.11 shows the predictive performance when the generated data is from the Gumbel copula. The method generally performs well, with the exception of cases where $n = 25$. This pattern is consistent with expectations, as imprecision decreases with increasing sample size. Tables 3.12 and 3.13 present the predictive performance results when the generated data are from the Frank copula and Joe copula, respectively. As the dependence increases, there are fewer cases where the q value is not between p_1 and p_2 compared to Table 3.10. The case when q is not in the interval $[p_1, p_2]$ occurs when the strength of dependence increases. This is due to the chosen copula type for simulation with a specific level of dependence and the bandwidth selection method that impacts the probabilities $h_{ijk}(\hat{c})$. In addition, this is reflected in the results when considering the event $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$ that can be explained through the probabilities $h_{ijk}(\hat{c})$. The imprecision is larger when the correlation is weak compared to when it is strong.

Overall, the simulation studies indicate that the proposed method performs well when combined with parametric copulas. In contrast, its performance is less effective with nonparametric copulas. This suggests the need for further investigation using alternative nonparametric copula approaches, to enable a more comprehensive comparison with the results obtained in this chapter.

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	**0.27**	0.43	**0.26**	0.31
	0.5	**0.51**	0.64	0.47	0.53
	0.75	0.73	0.83	0.65	**0.71**
M	0.25	**0.29**	0.39	**0.26**	0.31
	0.5	0.49	0.58	0.49	0.53
	0.75	0.69	0.78	0.67	**0.71**
H	0.25	0.19	0.25	**0.29**	0.34
	0.5	0.32	**0.41**	0.37	**0.42**
	0.75	0.58	**0.69**	0.63	**0.67**

Table 3.10: Predictive performance, Clayton copula

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.19	0.34	0.21	**0.24**
	0.5	0.39	0.53	0.39	**0.47**
	0.75	0.63	0.77	0.66	**0.70**
M	0.25	0.21	0.27	0.21	**0.24**
	0.5	0.38	0.53	0.40	**0.46**
	0.75	0.67	0.77	0.68	**0.71**
H	0.25	0.12	**0.23**	0.19	**0.21**
	0.5	0.40	0.56	0.44	**0.48**
	0.75	0.66	0.79	0.70	**0.74**

Table 3.11: Predictive performance, Gumbel copula

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.18	0.29	0.23	0.29
	0.5	0.45	0.56	0.47	0.52
	0.75	0.65	0.76	0.74	0.82
M	0.25	**0.26**	0.33	**0.28**	0.31
	0.5	0.45	0.60	**0.54**	0.56
	0.75	0.72	0.82	0.71	0.77
H	0.25	**0.28**	0.34	0.18	**0.20**
	0.5	0.49	0.53	0.50	0.53
	0.75	0.73	0.75	0.73	0.78

Table 3.12: Predictive performance, Frank copula

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.21	0.34	0.24	0.29
	0.5	0.43	0.56	0.49	0.53
	0.75	0.61	0.75	0.74	0.80
M	0.25	**0.27**	0.34	**0.27**	0.31
	0.5	0.46	0.54	**0.55**	0.58
	0.75	0.65	**0.74**	0.70	0.76
H	0.25	0.21	0.34	**0.28**	0.32
	0.5	0.43	0.56	**0.53**	0.57
	0.75	0.61	0.75	0.71	0.79

Table 3.13: Predictive performance, Joe copula

copula type	τ	θ	$n = 10$			$n = 25$		
			b_X	b_Y	b_Z	b_X	b_Y	b_Z
Clayton	L	0.6667	0.248	0.239	0.243	0.128	0.129	0.136
	M	2	0.193	0.202	0.192	0.113	0.109	0.113
	H	6	0.105	0.122	0.137	0.070	0.070	0.070
Gumbel	L	1.3333	0.208	0.183	0.191	0.119	0.111	0.111
	M	2	0.170	0.152	0.147	0.096	0.084	0.098
	H	4	0.111	0.091	0.091	0.055	0.053	0.067
Frank	L	2.3719	0.213	0.206	0.203	0.122	0.119	0.128
	M	5.736	0.165	0.146	0.177	0.099	0.106	0.105
	H	14.1385	0.073	0.061	0.066	0.073	0.061	0.066
Joe	L	1.5961	0.175	0.182	0.196	0.106	0.104	0.109
	M	2.8562	0.092	0.082	0.094	0.080	0.080	0.084
	H	6.7823	0.175	0.182	0.196	0.048	0.050	0.053

Table 3.14: Bandwidth values from simulated data across different sample sizes, correlation levels, and copula types.

3.6 Applications

This section uses two datasets from the literature to demonstrate the methods introduced in Sections 3.2 and 3.3. Both datasets relate to events of interest for a single future observation. The first dataset is applied to the trivariate case, and the second to the four-variate case, with both methods illustrated using these dimensions.

3.6.1 Typical Meteorological Year Data (TMY)

A Typical Meteorological Year (TMY) dataset is provided in the *pvlib* R package [56]. This dataset includes several weather measurements, with this example focusing on the apparent temperature (AT), also known as the "feels like" temperature, which is commonly used in meteorology.It provides an approximation of the heat balance experienced by the human body.

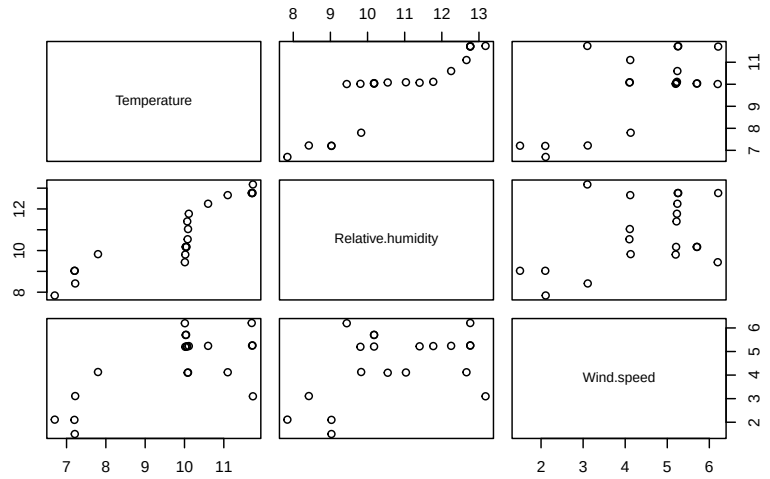


Figure 3.9: Pairwise scatterplots of variables in the TMY dataset.

Temperature	Water vapour pressure	Wind speed	Apparent temperature
10.01	9.440016	6.20	4.785
10.02	9.807809	5.20	5.617
10.03	10.175600	5.70	5.398
10.04	10.175610	5.71	5.401
10.05	10.175620	5.21	5.761
10.08	10.543390	4.10	6.689
10.09	11.033790	4.11	6.854
10.07	11.401580	5.22	6.179
10.11	11.769370	5.23	6.333
10.60	12.250260	5.24	6.975
11.71	12.765140	6.21	7.575
11.72	12.765143	5.25	8.257
11.73	12.765141	5.26	8.260
11.74	13.176920	3.10	9.918
11.10	12.664120	4.12	8.395
7.80	9.828025	4.13	4.152
7.20	9.027794	2.10	4.709
7.21	9.027794	1.50	5.139
7.22	8.419179	3.11	3.821
6.70	7.841308	2.11	3.811

Table 3.15: Dataset of dry bulb temperature, water vapour pressure and wind speed with their corresponding apparent temperature (AT) values.

This measure depends on air temperature (X), relative humidity (Y) and wind speed (Z). The apparent temperature is defined as [97]:

$$AT = T_a + 0.3e - 0.70ws - 4 \tag{3.6.11}$$

where T_a is the dry bulb temperature in degrees Celsius, e is the water vapour pressure in hectopascals and ws is the wind speed in meters per second. The water vapour pressure can be calculated from the temperature and relative humidity in case the vapour pressure is not given, which is defined as

$$e = \frac{rh}{100} 6.105 \exp(17.27 T_a/(237.7 + T_a))$$

where rh is the relative humidity.

A subset of the TMY dataset is used, with a small value of about 0.01 is added to the duplicated observations to avoid tied observations. The data are presented in Table 3.15. For this dataset, there is a high positive correlation of about 0.90 between dry bulb temperature and water vapour pressure, and about 0.70 between water vapour pressure and wind speed. A moderate positive correlation of 0.46 between dry bulb temperature and wind speed, as shown in Figure 3.9.

Assume that the event of interest is that the next apparent temperature value is greater than t, that is $AT_{n+1} > t$, where is given $AT_{n+1} = X_{n+1}, +0.33Y_{n+1} - 0.79Z_{n+1} - 4$, where X_{n+1}, Y_{n+1} and Z_{n+1} refer to the next values of temperature, relative humidity, and wind speed, respectively.

Combining NPI with a parametric copula

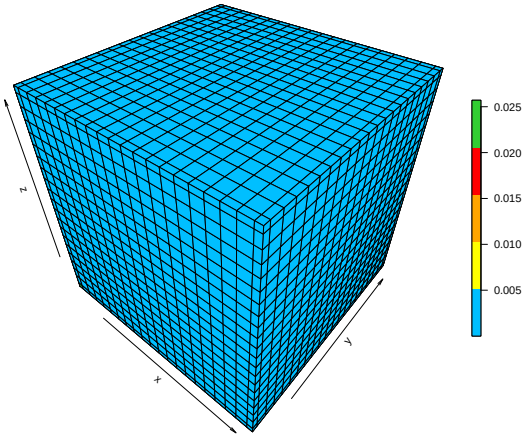
By applying the proposed method presented in Section 3.2, where assuming a trivariate parametric copula and using the pseudo maximum likelihood method to estimate the parameters. Thus, the estimated parameters and corresponding Kendall's (τ) values are: $\hat{\theta}_C = 1.78$ with $\tau = 0.47$ for the Clayton copula, $\hat{\theta}_G = 1.87$ with $\tau = 0.46$ for the Gumbel copula, $\hat{\theta}_F = 5.50$ with $\tau = 0.50$ for the Frank copula and $\hat{\theta}_J = 2.04$ with $\tau = 0.36$ for the Joe copula.

	Clayton		Gumbel		Frank		Joe	
t	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
-6.3570	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-4.3570	0.9996	1.0000	0.9975	1.0000	0.9978	1.0000	0.9975	1.0000
-2.3570	0.9923	1.0000	0.9811	1.0000	0.9817	1.0000	0.9748	1.0000
-0.3570	0.9548	0.9996	0.9589	0.9998	0.9587	0.9996	0.9570	0.9997
1.6430	0.9509	0.9989	0.9449	0.9983	0.9449	0.9984	0.9416	0.9974
3.6430	0.8821	0.9345	0.8540	0.9093	0.8625	0.9110	0.8369	0.8874
5.6430	0.6631	0.6759	0.6781	0.6987	0.6845	0.6915	0.6633	0.6909
7.6430	0.1640	0.2145	0.1970	0.2289	0.1782	0.2230	0.2230	0.2532
9.6430	0.0012	0.0649	0.0031	0.0572	0.0019	0.0634	0.0101	0.0600
11.6430	0.0000	0.0381	0.0000	0.0201	0.0000	0.0361	0.0001	0.0187
13.6430	0.0000	0.0049	0.0000	0.0049	0.0000	0.0067	0.0000	0.0049
15.6430	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

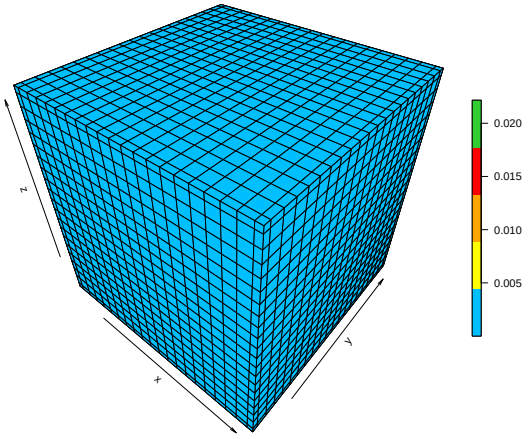
Table 3.16: The NPI lower and upper probabilities of the event $AT_{n+1} > t$ using different types of copula at selected values of t.

Figure 3.10 shows the probabilities h_{ijk} with different trivariate copulas. This figure illustrates how these probabilities are affected by the estimated parameter and the type of copula. As explained in Section 3.2, for a positive high correlation, the probabilities $h_{ijk}(\hat{\theta})$ include large values when i, j and k are close to each other. Choosing different copula types leads to different results in the probabilities $h_{ijk}(\hat{\theta})$ and as a result, the NPI lower and upper probabilities are different.

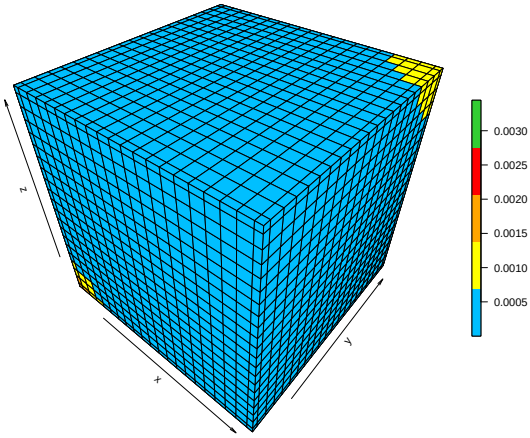
The lower and the upper probabilities for the event $AT_{n+1} > t$ are presented in Figure 3.11 presented in Table 3.16 for selected values of t. The figure illustrates that the imprecision, defined as the difference between the corresponding upper and lower probabilities, remains relatively consistent across the values of t. This is due to the impact discussed in Section 3.2, particularly in relation to the high positive correlation between the variables and considering the event of interest $AT_{n+1} > t$.



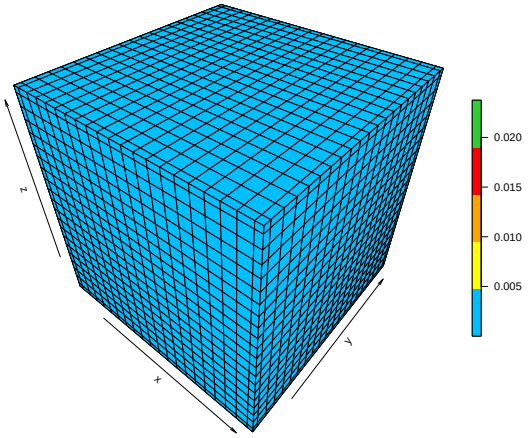
(a) Clayton copula



(b) Gumbel copula

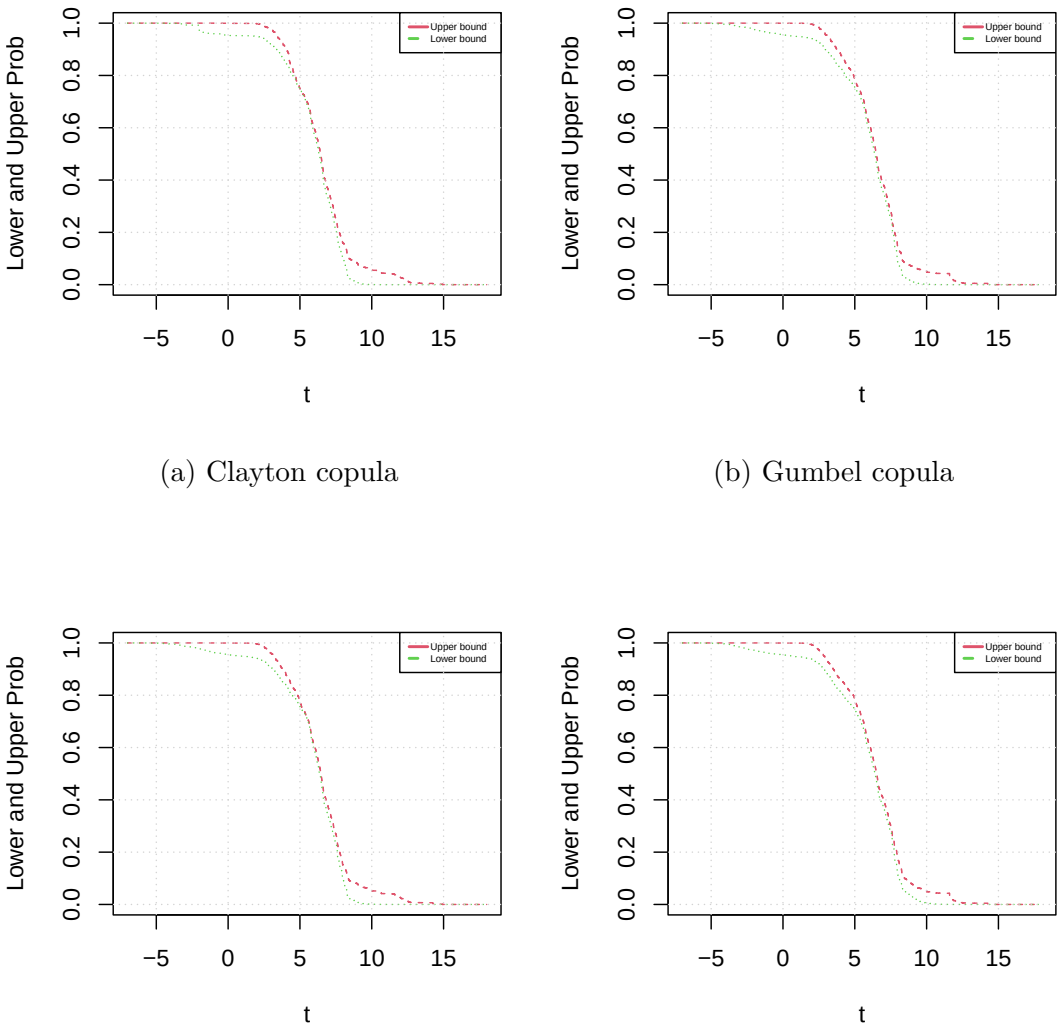


(c) Frank copula



(d) Joe copula

Figure 3.10: The h_{ijk} probabilities, Example 3.6.1.



(a) Clayton copula

(b) Gumbel copula

(c) Frank copula

(d) Joe copula

Figure 3.11: The NPI lower and upper probabilities of the event $AT_{n+1} > t$ using different types of copula.

Combining NPI with a nonparametric copula

This example implements the proposed method introduced in Section 3.3, assuming a nonparametric copula, and applies two types of bandwidth selection: the normal reference rule-of-thumb and LSCV.

The bandwidths of X, Y and Z remain the same when using the normal reference

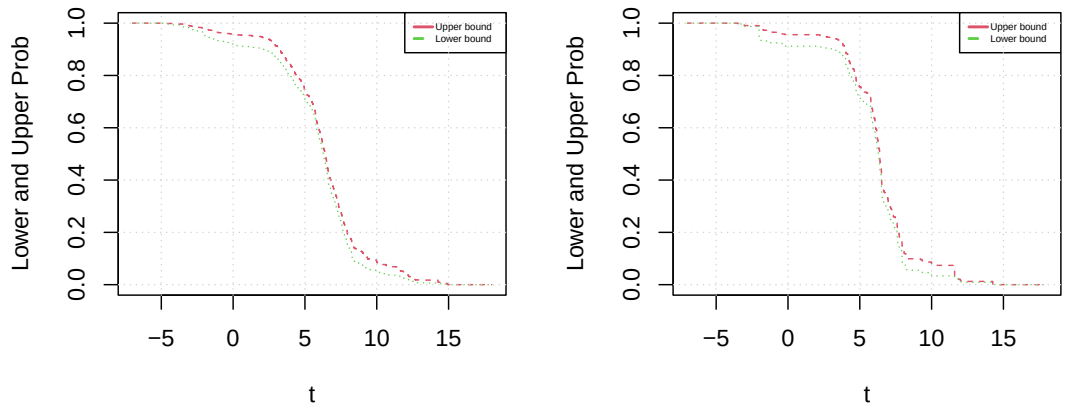
Normal Reference			LSCV	
t	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
-6.3570	1.0000	1.0000	1.0000	1.0000
-4.3570	0.9970	0.9987	1.0000	1.0000
-2.3570	0.9688	0.9834	0.9913	0.9920
-0.3570	0.9259	0.9602	0.9132	0.9565
1.6430	0.9051	0.9493	0.9122	0.9555
3.6430	0.8246	0.8657	0.8995	0.9429
5.6430	0.6479	0.6768	0.6881	0.7321
7.6430	0.2058	0.2452	0.1585	0.2018
9.6430	0.0570	0.0967	0.0433	0.0857
11.6430	0.0322	0.0535	0.0165	0.0172
13.6430	0.0068	0.0174	0.0079	0.0087
15.6430	0.0000	0.0001	0.0000	0.0000

Table 3.17: The NPI lower and upper probabilities of the event that $AT_{n+1} > t$ using different types of bandwidths at selected values of t.

rule-of-thumb, at approximately 0.160. For the LSCV method, the bandwidth values are 0.029, 0.029 and 0.036 for X, Y and Z, respectively. These bandwidth selection methods are provided by the *np* R package [47].

The NPI lower and upper probabilities for the event $AT_{n+1} > t$ correspond to the bandwidth methods and are presented in Figure 3.12 and, for selected values of t in Table 3.17. The figure shows little difference; indicating that all bandwidth selections similar lower and upper probabilities.

It is clear that noticeable differences are observed in the results from Figure 3.12 between the values 0 and 5, whereas the remaining values show little variation. This occurs due to the chosen bandwidth, where the LSCV focuses on selecting the bandwidth values based on minimizing the integrated mean squared error. As a result, this affects the probabilities $h_{ijk}(\hat{c})$, which is essential for making inferences in the proposed method. Consequently, the NPI lower and upper probabilities for the event $AT_{n+1} > t$ are affected.



(a) Normal reference (b) LSCV

Figure 3.12: The NPI lower and upper probabilities of the event $AT_{n+1} > t$ using different types of bandwidths.

3.6.2 Weekly return data

A dataset of 50 weekly return observations on each of ten stocks was presented by Jobson [57]. The portfolios were constructed using stocks from the Toronto Stock Exchange, beginning in 1982. This example involves four selected stocks, with 35 observations per stock, showing high positive correlation, as shown in Figure 3.13.

Let the portfolio consist of four stocks denoted by X, Y,Z and W. The portfolio return formula is defined as $PR = w_1X + w_2Y + w_3Z + w_4W$, where w_i are the weights with equally weighted 0.25. Assume that one is interested in the next portfolio return PR_{n+1} that exceed a value t, i.e $PR_{n+1} > t$

Combining NPI with a parametric copula

The generalized method from Section 3.2 is applied to the case $d = 4$, where NPI is combined with a four-variate parametric copula. Table 3.18 and Figure 3.14 present the NPI lower and upper probabilities of the event of interest $PR_{n+1} > t$. The results are calculated using the same parametric copulas but in the four-variate case and the parameters are estimated using the pseudo maximum likelihood estimation method. As shown

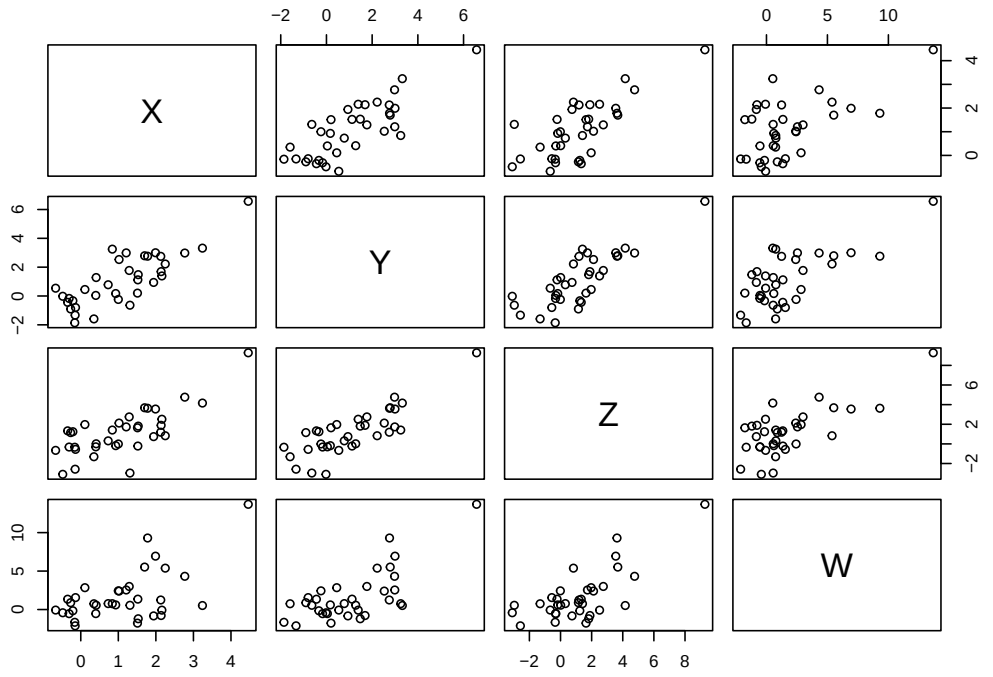
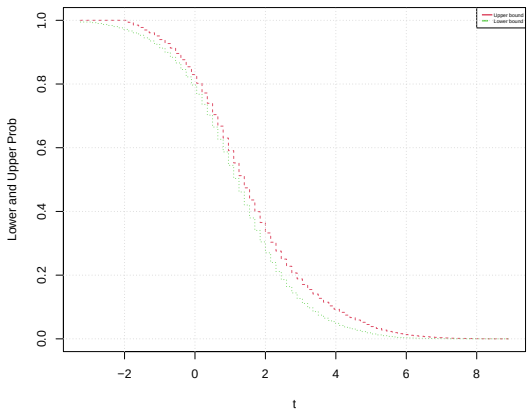
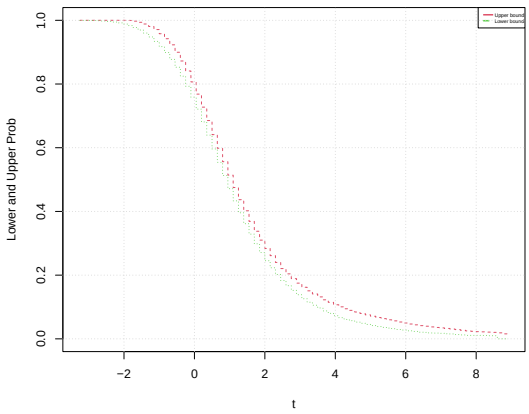


Figure 3.13: Pairwise scatterplots of variables in the weekly return dataset.

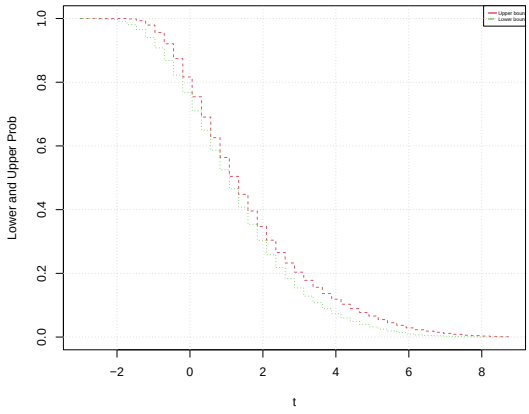
in Figure 3.14, the results appear visually similar across the parametric copulas. Table 3.18 provides a clearer comparison at specific values t. The assumed parametric copulas lead to different probabilities $h_{ijkl}(\hat{\theta})$. Considering an event involving the sum can be interpreted through the probabilities $h_{ijkl}(\hat{\theta})$. For a high positive correlation, the probabilities $h_{ijkl}(\hat{\theta})$ include large values when i, j, k, l are close to each other. Consequently, the lower and upper probabilities are affected. This is due to the effect discussed in Section 3.2, specifically the high positive correlation between random quantities, along with the interest in the sum of these quantities.



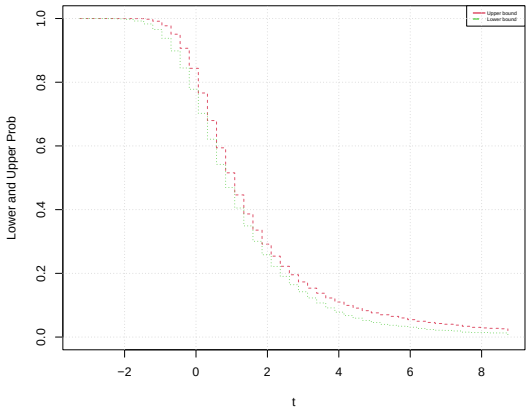
(a) Clayton copula



(b) Gumbel copula



(c) Frank copula



(d) Joe copula

Figure 3.14: The NPI lower and upper probabilities of the event $PR_{n+1} > t$ using different types of copula.

	Clayton		Gumbel		Frank		Joe	
t	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
-3.25	0.9946	1.0000	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000
-2.75	0.9889	1.0000	0.9985	1.0000	0.9997	1.0000	0.9999	1.0000
-2.25	0.9775	1.0000	0.9928	1.0000	0.9966	1.0000	0.9990	1.0000
-1.75	0.9603	0.9850	0.9773	0.9973	0.9826	0.9991	0.9929	0.9998
-1.25	0.9320	0.9578	0.9434	0.9778	0.9440	0.9819	0.9689	0.9928
-0.75	0.8892	0.9177	0.8847	0.9297	0.8760	0.9280	0.9069	0.9570
-0.25	0.8224	0.8541	0.7931	0.8409	0.7805	0.8295	0.7948	0.8601
0.25	0.7247	0.7608	0.6667	0.7129	0.6647	0.7065	0.6433	0.7034
0.75	0.5998	0.6426	0.5260	0.5696	0.5432	0.5820	0.4917	0.5395
1.25	0.4610	0.5124	0.3957	0.4369	0.4267	0.4665	0.3679	0.4065
1.75	0.3280	0.3876	0.2888	0.3281	0.3217	0.3646	0.2738	0.3077
2.25	0.2203	0.2844	0.2087	0.2466	0.2342	0.2809	0.2038	0.2356

Table 3.18: The NPI lower and upper probabilities of the event $PR_{n+1} > t$ using different types of copula at selected values of t.

Combining NPI with a nonparametric copula

Applying the proposed generalization method introduced in Section 3.3 to the four-variate case. The bandwidths are selected using the normal reference rule-of-thumb and the LSCV. The bandwidth values using the LSCV have the smallest values of 0.070, 0.075, 0.002 and 0.070, compared to those obtained by selecting the bandwidth using the normal reference rule-of-thumb, which gives identical results of 0.138.

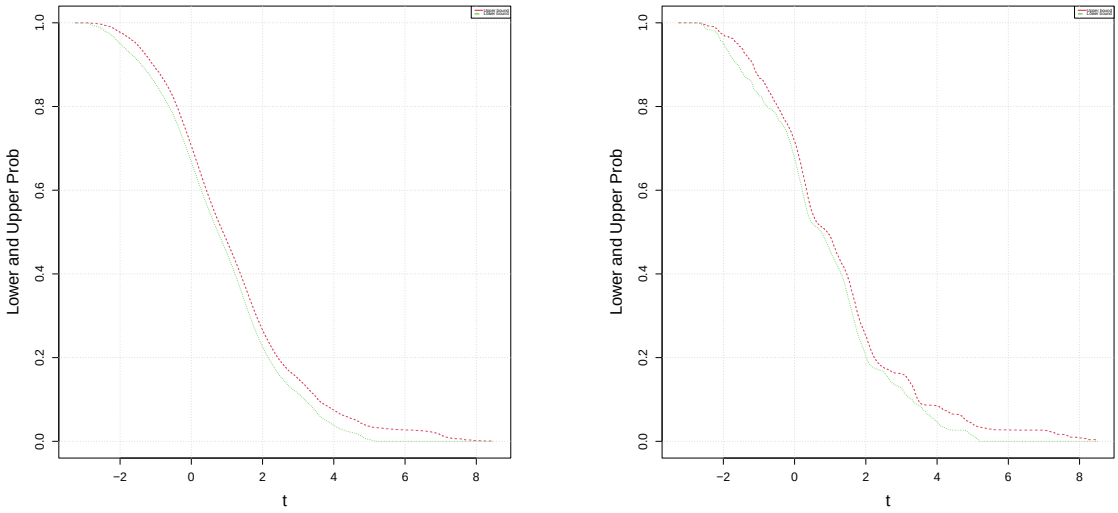
Table 3.19 and Figure 3.15 present the resulting NPI lower and upper probabilities of the event of interest $PR_{n+1} > t$. Given a strong positive correlation in the dataset, the probabilities $h_{ijkl}(\hat{c})$ indicate large values when i, j, k, l are close to each other. Therefore, calculating the lower and upper probabilities for an event of interest tend to include additional $h_{ijk}(\hat{\theta})$. These additional terms generally have large values for most t under positive correlation. Considering the event of interest $PR_{n+1} > t$, these additional probabilities $h_{ijkl}(\hat{c})$ often contain a few larger values for most values of t leading to small imprecision.

The TMY and portfolio returns examples illustrate the proposed methods in Section

	Normal Reference		LSCV	
t	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
-3.25	0.9995	1.0000	1.0000	1.0000
-2.75	0.9939	0.9987	0.9997	0.9999
-2.25	0.9713	0.9895	0.9801	0.9895
-1.75	0.9299	0.9642	0.9150	0.9629
-1.25	0.8838	0.9201	0.8637	0.9104
-0.75	0.8192	0.8606	0.7981	0.8414
-0.25	0.7257	0.7656	0.7452	0.7638
0.25	0.6070	0.6427	0.5858	0.6336
0.75	0.4990	0.5273	0.5000	0.5131
1.25	0.3950	0.4279	0.4119	0.4373
2.25	0.1864	0.2248	0.2608	0.3110

Table 3.19: The NPI lower and upper probabilities of the event that $PR_{n+1} > t$ using different types of bandwidths at selected values of t.

3.2 and 3.3. For the TMY example, where the event of interest is $AT_{n+1} > t$, the results show that the proposed method performs well using either a parametric copula or a kernel-based nonparametric copula. For the portfolio returns example, the method applies both parametric and nonparametric copulas in the four-variate case. As discussed in this example the proposed method performs well in that multivariate setting, also demonstrating the ability to be applied in any dimension.



(a) Normal reference

(b) LSCV

Figure 3.15: The NPI lower and upper probabilities of the event that $PR_{n+1} > t$ using different types of bandwidths.

3.7 Concluding remarks

This chapter extends the two methods for combining NPI with bivariate copulas, originally introduced by Coolen-Maturi *et al.* [27, 77], to the trivariate case. Both the semiparametric and nonparametric approaches are presented, using copulas to model the dependence structure. The focus is on predictive inference for a single future trivariate observation. A generalization to higher dimensions ($d > 3$) is developed following a similar structure. The methods are illustrated through examples, and their predictive performance is assessed via simulations in the trivariate case. Additionally, applications to four-variate data from the literature demonstrate their use in higher dimensions.

For the first method, a parametric copula with one dependence parameter is used to describe the dependence, meaning that each pair of variables shares the same dependence level. The performance of this method is evaluated through simulation studies. Throughout this work, the focus is on assuming a trivariate parametric copula with one parameter. It might be of interest to study classical copulas with multiple parameters and this is left for future work. One could explore copulas that differ in their representation of the dependence structure that contain multiple parameters, such as vine copulas and fully nested Archimedean copulas (FNAC). These types of dependence models can capture a range of dependency structures, and their use will be examined in Chapters 4 and 5.

For the second method, which assumes a nonparametric copula—specifically a kernel-based copula—the focus is on bandwidth selection, including types such as the normal reference rule-of-thumb and least squares cross-validation (LSCV). The method's performance is evaluated through simulations, revealing poor results regardless of sample size or dependence level between variables. Given its unsatisfactory performance compared to the first method, further investigation is needed, including exploring alternative nonparametric copula approaches. This is left for future work. Additionally, both methods become increasingly time-consuming and computationally demanding as dataset size and dimensionality grow.

Chapter 4

NPI Combined with Vine Copula

4.1 Introduction

This chapter introduces a method of combining nonparametric predictive inference (NPI) with vine copulas for predictive inference. A vine copula is a type of dependence model that captures different dependencies among variables and provides a flexible framework for constructing multivariate dependence structures using bivariate copulas. There are no restrictions on the choice of bivariate copula families within vine copulas, and each bivariate copula in the vine structure can belong to any copula family. This flexibility enables vine copulas to represent a wide range of dependence structures within a model. The proposed method focuses on predictive inference in the trivariate case.

This chapter is organized as follows: In Section 4.2, the method of NPI combined with trivariate vine copulas is introduced. The effectiveness of this newly proposed methodology is illustrated in Section 4.3. The performance of this method is investigated in Section 4.4. An example from the literature illustrating the real-world application of the proposed method is presented in Section 4.5. Some concluding remarks are included in Section 4.6.

4.2 Combining NPI with a parametric vine copula

This section presents a method for combining NPI with trivariate vine copula. The bivariate h_{ij} presented in Section 2.7 is used to construct vine copulas for the NPI multivariate

approach as

$$h_{ijk}^{xyz} = h_{ij|k}^{xy|z} \times h_{ik}^{xz} \times h_{jk}^{yz} \times (n+1)^3 \tag{4.2.1}$$

where h_{ik}^{xz} and h_{jk}^{yz} follow the same formulation, based on the work by Coolen-Maturi *et al.*, [27] presented in Section 2.7. Assume that there are n trivariate observations (x_i, y_i, z_i), $i = 1, \ldots, n$, which are the observed values of n exchangeable trivariate random quantities with no ties. The observations of the marginals are ordered and denoted by x_i, y_j and z_k for simplicity, so $x_1 < \cdots < x_i < \cdots < x_n$, $y_1 < \cdots < y_j < \cdots < y_n$ and $z_1 < \cdots < z_k < \cdots < z_n$.

By applying the assumption $A_{(n)}$ for the marginals and using the natural transformations with an assumed bivariate copulas as

$$(X_{n+1} \in (x_{i-1}, x_i), Y_{n+1} \in (y_{j-1}, y_j)) \Longleftrightarrow (\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1})) \tag{4.2.2}$$

$$(X_{n+1} \in (x_{i-1}, x_i), Z_{n+1} \in (z_{k-1}, z_k)) \Longleftrightarrow (\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1})) \tag{4.2.3}$$

where $i, j, k \in 1, ..., n+1$. Assuming a bivaraite parametric copulas after applying NPI for the marginals by using transformed data such as $(\frac{r_i^x}{(n+1)}, \frac{r_i^y}{(n+1)})$ and $(\frac{r_i^x}{(n+1)}, \frac{r_i^z}{(n+1)})$ instead of (x_i, y_i) and (x_i, z_i), respectively, to coincide to the transformation method for the marginals, where r_i^x is the rank of the observation x_i among the $x-$observations, r_i^y is the rank of the observation y_i among the $y-$observations and r_i^z is the rank of the observation z_i among the $z-$observations. Therefore, the probabilities h_{ik}^{xz} and h_{jk}^{yz} are defined as

$$h_{ik}(\hat{\theta_1}) = P(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1}) | \hat{\theta_1}) \tag{4.2.4}$$

$$h_{ij}(\hat{\theta_2}) = P(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}) | \hat{\theta_2}) \tag{4.2.5}$$

The estimated parameter of h_{ik}^{xz} is $\hat{\theta}_1$ and the estimated parameter of h_{ij}^{xy} is $\hat{\theta}_2$. The probability $h_{jk|i}^{yz|x}$ is similar to h_{ik}^{xz} and h_{ij}^{xy} and the difference lies in introducing the observations under a simplifying assumption, which neglects conditioning on $X \in I_i^x$. This assumption, commonly used in the literature, is explicitly adopted in this work and discussed

in Section 2.4. Therefore, under the simplifying assumption for modeling the conditional copula, we first use the estimated parameters $\hat{\theta}_1$ and $\hat{\theta}_2$ to define the pseudo-observations l and m. Let the conditional variables $Y|X$ and $Z|X$ be denoted by L and M, respectively. The pseudo-observations are then defined as follows, in order to obtain $h_{ij|k}^{xy|z}$:

$$l = F_{Y|X=x}(y|x;\hat{\theta}_1) = \frac{\partial C_{xy}(x,y;\hat{\theta}_1)}{\partial x} \tag{4.2.6}$$

$$m = F_{Z|X=x}(z|x;\hat{\theta}_2) = \frac{\partial C_{xz}(x,z;\hat{\theta}_2)}{\partial x} \tag{4.2.7}$$

where $\hat{\theta}_1$ is the estimated parameter of the bivariate copula of (X,Z) and $\hat{\theta}_2$ is the estimated parameter of the bivariate copula of (X,Y). The conditional copula under the simplifying assumption is defined as

$$c_{yz|x}(y,z|x) = c_{yz}(F_{Y|X}(y|x), F_{Z|X}(z|x)) \quad \text{for} \quad y,z \in [0,1] \tag{4.2.8}$$

where $F_{Z|X}(z|x)$ and $F_{Y|X}(y|x)$ are the conditional distribution functions of Y given $X = x$ and Z given $X = x$, respectively. Let l_q and m_ν denote the ordered observations so, $l_1 < \cdots < l_q < \cdots < l_n$ and $m_1 < \cdots < m_\nu < \cdots < m_n$. By using the same natural transformations associated with the marginal $A_{(n)}$ assumptions, as outlined in Section 2.7 we have

$$\begin{aligned}(L_{n+1} \in (l_{q-1}, l_q), M_{n+1} \in (m_{\nu-1}, m_\nu)) \Longleftrightarrow \\ (\tilde{L}_{n+1} \in (\frac{q-1}{n+1}, \frac{q}{n+1}), \tilde{M}_{n+1} \in (\frac{\nu-1}{n+1}, \frac{\nu}{n+1}))\end{aligned} \tag{4.2.9}$$

where $\tilde{L}_{n+1}$ and $\tilde{M}_{n+1}$ are the transformation of the two random quantities L_{n+1} and M_{n+1}. $q, \nu = 1, 2, \ldots, n+1$, where $q_0 = -\infty, q_{n+1} = \infty$ and $\nu_0 = -\infty, \nu_{n+1} = \infty$. The parameter is estimated by assuming a bivariate parametric copula such as $(\frac{r_i^l}{(n+1)}, \frac{r_i^m}{(n+1)})$ instead of (l_i, m_i). Where r_i^l is the rank of the observation l_q among the l- observations and r_i^m is the rank of the observation m_ν among the m- observations. The probabilities h_{ij}^{lm} can be defined by combining NPI on the marginal with the estimated parameter $\hat{\theta}_3$ as follows:

$$h_{q\nu}^{lm}(\hat{\theta}_3) = P(\tilde{L}_{n+1} \in (\frac{q-1}{n+1}, \frac{q}{n+1}), \tilde{M}_{n+1} \in (\frac{\nu-1}{n+1}, \frac{\nu}{n+1})|\hat{\theta}_3) \tag{4.2.10}$$

Therefore, using h_{ik}^{xz},h_{jk}^{yz} and $h_{ij|k}^{xy|z}$ to build the construction as in Equation (4.2.1)

$$\begin{aligned} h_{ijk}(\hat{\boldsymbol{\theta}}) = P(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1})|\hat{\theta}_1) \\ \times P(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1})|\hat{\theta}_2) \\ \times P(\tilde{L}_{n+1} \in (\frac{q-1}{n+1}, \frac{q}{n+1}), \tilde{M}_{n+1} \in (\frac{\nu-1}{n+1}, \frac{\nu}{n+1})|\hat{\theta}_3) \times (n+1)^3 \end{aligned} \quad (4.2.11)$$

where $P(.|\hat{\boldsymbol{\theta}})$ is the copula-based probability with estimated parameters $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ are the estimated parameters of the bivariate copulas. Equation (4.2.11) should meet the conditions as presented in Section 3.2, where

1. $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\boldsymbol{\theta}}) = 1$
2. $\sum_{j=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\boldsymbol{\theta}}) = \frac{1}{n+1}$, for $i \in \{1, 2, \ldots, n+1\}$, $\sum_{i=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\boldsymbol{\theta}}) = \frac{1}{n+1}$, for $j \in \{1, 2, \ldots, n+1\}$ and $\sum_{i=1}^{n} \sum_{j=1}^{n} h_{ijk}(\hat{\boldsymbol{\theta}}) = \frac{1}{n+1}$, for $k \in \{1, 2, \ldots, n+1\}$
3. $h_{ijk}(\hat{\boldsymbol{\theta}}) \geq 0$, for i, j,$k \in \{1, \ldots, n+1\}$

Equation (4.2.11) will be used to make inferences about an event of interest, as explained in Section 3.2, using Equations (3.2.9) and (3.2.10).

Example 4.2.1 This example illustrates the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ as given in Equation (4.2.11). A dataset of 9 observations was generated from a trivariate Gaussian distribution with mean vector zero and a variance-covariance matrix equal to the identity matrix.

Assume that each bivariate copula follows a Frank copula and their parameters are estimated using the pseudo maximum likelihood estimation method. The estimated parameters of each bivariate copula, along with their corresponding Kendall's τ values are presented in Table 4.1. This table shows that the first two bivariate copulas, (X, Y) and (X, Z), exhibit dependence values of approximately 0.44 and 0.79, respectively, while the third copula indicates a weaker dependence of about 0.19. This result aligns with the vine structure where the first two pairs show stronger dependence than the third, which reflects a noticeably weaker association.

After estimating the parameters, the probabilities h_{ijk} are calculated as described in Equation (4.2.11). While the vine copula model offers greater flexibility in capturing a range of dependencies compared to a classical trivariate copula with a single parameter,

Pairs	τ	$\hat{\theta}$
x, y	0.44	4.77
x, z	0.79	17.50
y, z	0.19	1.76

Table 4.1: Estimated parameters and corresponding Kendall's τ values for the simulated data using Frank copulas.

this flexibility introduces a notable challenge. The computed probabilities h_{ijk} show that the marginals over X, Y and Z are not equal to $1/(n+1)$ and the total sum of h_{ijk} is not exactly equal to one, as is clear from Table 4.2 (left side).

This issue arises when discretizing each bivariate copula in the vine structure to obtain the probabilities as in Equation (4.2.11) which leads to inconsistent results. Since the discretization process is applied independently to each bivariate copula, the marginal distributions derived from different pair-copulas may not match exactly. As a result, the overall structure may no longer satisfy the defining properties of a copula, most notably, the requirement that all marginal distributions be uniform on the interval $[0, 1]$. In this sense, the resulting discretized vine is no longer a copula. This issue can be solved using the iterated proportional fitting (IPF), which was first introduced by Deming and Stephan [33] and investigated with discrete copula by Geenens [41]. This algorithm focuses on reaching the desired margins and preserving the dependence structure of the joint probabilities. This algorithm consists of normalizing the rows and columns of the probabilities h_{ijk} alternately to derive uniform marginals. The probabilities h_{ijk} after applying the IPF satisfy their conditions, as presented in this example. The IPF algorithm is straightforward and can be generalized in any dimension and it is available in the R package *mipfp* [13]. Modifying the probabilities h_{ijk} using the IPF algorithm shows all the conditions satisfied as in Table 4.2 (right side). Thus, these probabilities h_{ijk}, with their sufficient conditions, now become suitable for investigation and visualization.

	$h_{.jk}$	$h_{i.k}$	$h_{ij.}$		$h_{.jk}$	$h_{i.k}$	$h_{ij.}$
1	0.1394	0.1407	0.1398	1	0.1	0.1	0.0999
2	0.1254	0.1248	0.1252	2	0.1	0.1	0.1
3	0.1136	0.1131	0.1135	3	0.1	0.1	0.1
4	0.1057	0.1055	0.1057	4	0.1	0.1	0.1
5	0.1019	0.1018	0.1018	5	0.1	0.1	0.1001
6	0.1019	0.1018	0.1018	6	0.1	0.1	0.1001
7	0.1057	0.1055	0.1057	7	0.1	0.1	0.1
8	0.1136	0.1131	0.1135	8	0.1	0.1	0.1
9	0.1254	0.1248	0.1252	9	0.1	0.1	0.1
10	0.1394	0.1407	0.1398	10	0.1	0.1	0.0999

Table 4.2: The marginals of each variable before (left) and after (right) applying the IPF

Example 4.2.2 Three three-dimensional visualizations of the probabilities $h_{ijk}(\hat{\theta})$ under different dependence structures are shown in Figures 4.1-4.3. Each figure is based on a dataset of size $n = 4$, simulated from a trivariate Gaussian distribution with mean vector zero. Three cases are considered: a no correlation case where the covariance matrix is the identity matrix, a high correlation case where all off-diagonal entries of the covariance matrix are set to 0.9 and a negative high correlation case where all off-diagonal entries of the covariance matrix are set to -0.9 as given in Table 4.3.

Assume that the bivariate copulas for the pairs $(X, Y), (X, Z)$ and $(Y, Z|X)$, which capture the dependence structure between each pair of variables, are Gaussian copulas, representing a symmetrical dependence structure. The pseudo maximum likelihood estimation method is used to estimate the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and their corresponding τ values, as given in Table 4.4. The results illustrate that the dependence for the first two pairs $(X, Y), (X, Z)$ is stronger than in the third pair $(Y, Z|X)$ which show how these dependencies are affected when the assumed correlation of the generated data changes.

Since the main focus is to visualize the $h_{ijk}(\boldsymbol{\theta})$ probabilities, the probabilities are calculated using the estimated parameters $\boldsymbol{\theta}$ in Equation (4.2.11). Figures 4.1-4.3 illustrate different cases when varying the assumed correlation of the generated data. Each figure

Positive Correlation			No Correlation			Negative Correlation		
X	Y	Z	X	Y	Z	X	Y	Z
-0.831	0.181	-0.134	0.687	1.066	0.537	-0.071	0.452	-0.278
1.197	0.220	0.512	1.906	1.063	1.370	0.698	-0.464	-0.040
0.407	0.315	0.202	0.528	0.403	1.168	-0.205	-0.377	0.677
0.042	-0.394	-0.756	0.793	-1.214	-1.006	1.727	-1.039	-0.752

Table 4.3: Simulated data from a trivariate Gaussian distribution with different correlation structures.

Pairs	Positive correlation		No correlation		Negative correlation	
	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$
(x, y)	0.61	0.81	-0.32	-0.48	-0.73	-0.91
(x, z)	0.73	0.91	0.45	0.65	-0.77	-0.93
$(y, z\|x)$	0.50	0.70	0.19	0.30	-0.42	-0.61

Table 4.4: Estimated parameters and corresponding Kendall's τ values from simulated data, correlation levels, and Gaussian vine copula.

presents three sides: the right side, where $h_{.jk} = \sum_i h_{ijk}$, the left side, where $h_{i.k} = \sum_j h_{ijk}$ and the bottom side, where $h_{ij.} = \sum_k h_{ijk}$. The case where the data is highly correlated is presented in Figure 4.1. This figure shows a positive relation with large values of $h_{ijk}(\hat{\boldsymbol{\theta}})$ when i, j, k are close to each other. When there is no correlation in the generated data, the results appear scattered, as shown in Figure 4.2. For the negative case, Figure 4.3 displays large values with a negative relation on each side of $h_{ij.}$ and $h_{i.k}$. This is because the generated data with negative correlation affects the results on the marginals $h_{ij.}$ and $h_{i.k}$ for the pairs (X, Y) and (X, Z), whereas the third pair $(Y, Z|X)$ shows less effect. Also, the chosen pairs in this order are not unique and changing the order leads to different h_{ijk} results, but the relation between variables remains the same.

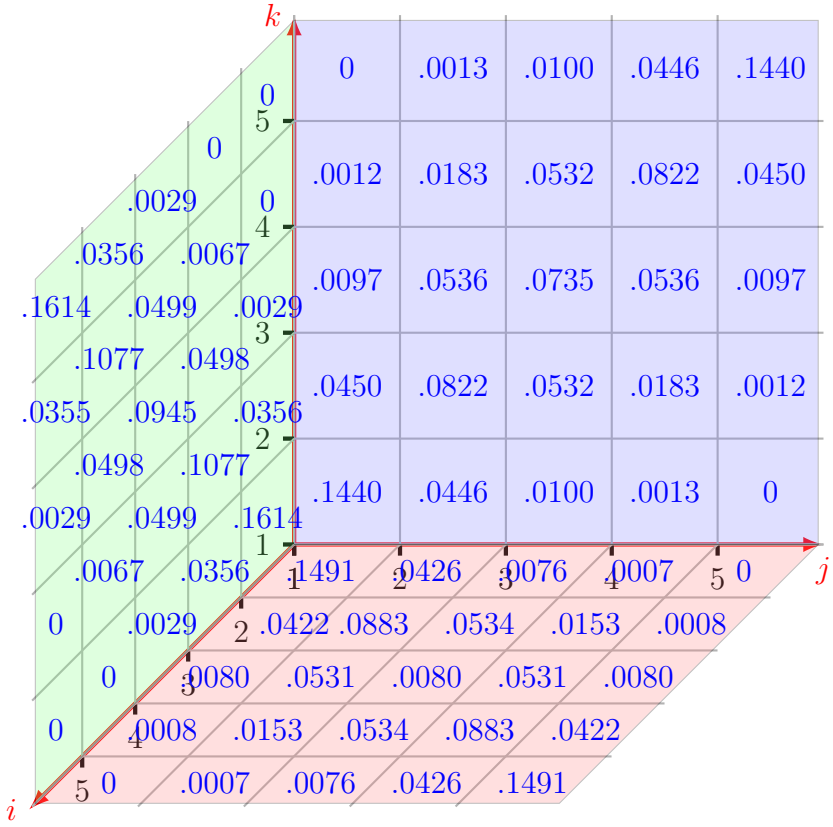


Figure 4.1: The probabilities $h_{.jk}, h_{i.k}$ and $h_{ij.}$, for the high correlation case.

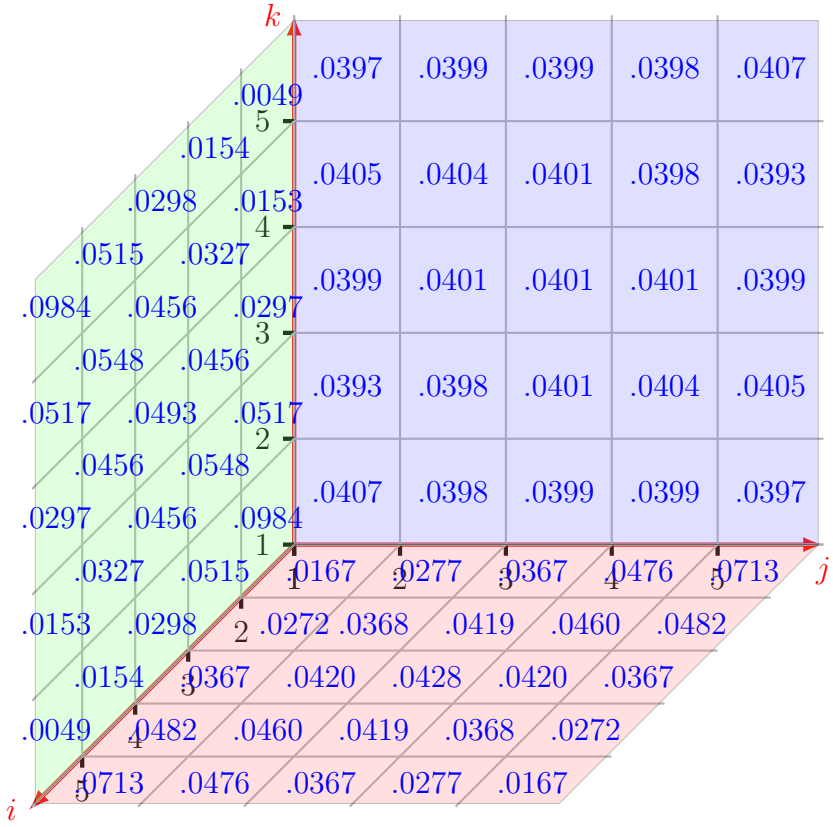


Figure 4.2: The probabilities $h_{.jk}, h_{i.k}$ and $h_{ij.}$, for the no correlation case.

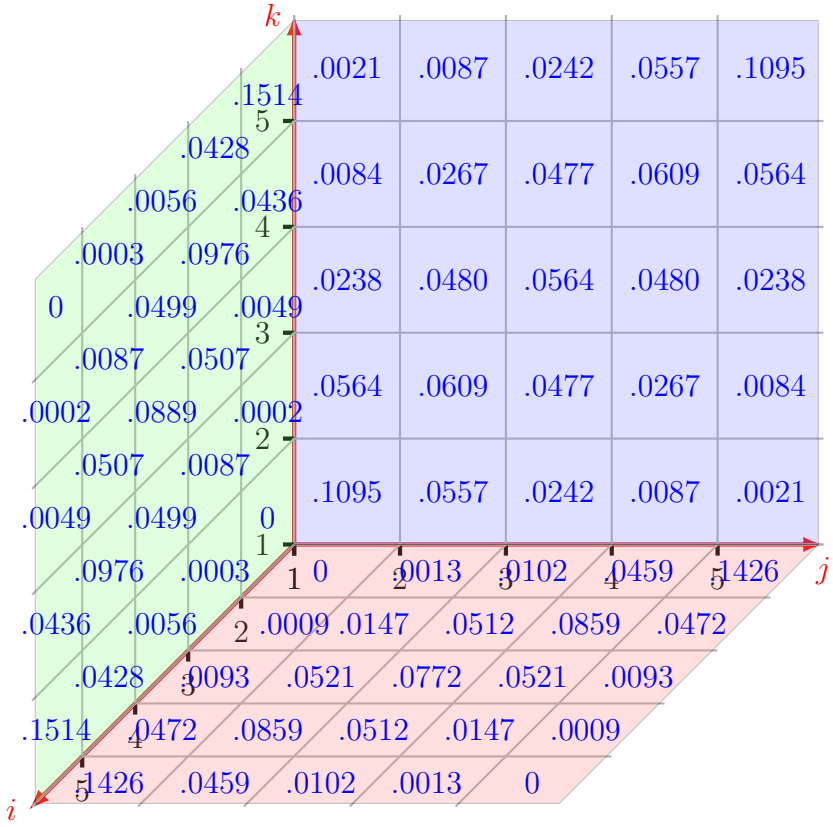


Figure 4.3: The probabilities $h_{.jk}, h_{i.k}$ and $h_{ij.}$, for the negative high correlation case.

4.3 Example

This section illustrates the proposed methods combining NPI with vine copulas as in Section 4.2 using trivariate Gaussian datasets with zero mean vectors. Three covariance matrices representing different correlation structures are considered, with sample sizes $n = 10, 25, 50$. These covariance matrices are as follows:

$$\begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The assumed correlations between the random variables in each sample are $0.9, 0.5, 0$ denoted as High (H), Moderate (M) and Low (L) for convenience. All samples are simulated using the R package *mvtnorm* [46]. The pseudo maximum likelihood estimation method is applied to estimate the bivariate copula parameters, assuming the Gaussian copula as the parametric form. These estimations can be performed using the *VineCopula* package in R [82].

In the first scenario (Case I), the pair-copula construction is based on the selected pairs (X, Y), (X, Z), and the conditional pair $(Y, Z|X)$. Following the proposed method introduced in Section 4.2, the corresponding probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ are computed accordingly. The second scenario considers (Case II) a pair-copula construction based on the following selected pairs (X, Y), (Y, Z) and the conditional pair $(X, Z|Y)$. The third scenario (Case III) involves the pairs (Y, Z), (X, Z), and the conditional pair $(X, Y|Z)$. In the fourth scenario (Case IV), a pair-copula construction is based on the chosen pairs: (X, Y), (Y, Z) and (X, Z), where the third pair does not depend on a conditioned variable. The first three scenarios are based on selecting specific orders of variable pairs and the effect on the NPI lower and upper probabilities is studied. Changing the selected pairs leads to different probabilities h_{ijk} and hence different NPI lower and upper probabilities for the event of interest. The first three scenarios are based on selecting specific orders of variable pairs and the effect on the NPI lower and upper probabilities is studied. Changing the selected pairs leads to different probabilities h_{ijk} and hence different NPI lower and upper probabilities for the event of interest.

Case I		Case II		Case III		Case IV	
τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$
0.61	0.81	0.36	0.53	0.41	0.27	0.36	0.53
0.73	0.91	0.36	0.54	0.36	0.54	0.27	0.41
0.50	0.70	0.12	0.19	0.26	0.40	0.36	0.54

Table 4.5: Estimated parameters and corresponding Kendall's τ values from simulated data with varying sample sizes, moderate correlation, and copula types.

Case I		Case II		Case III		Case IV	
τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$
0.70	0.89	0.70	0.89	0.69	0.89	0.70	0.89
0.69	0.89	0.72	0.90	0.72	0.90	0.69	0.89
0.35	0.53	0.27	0.41	0.32	0.48	0.72	0.90

Table 4.6: Estimated parameters and corresponding Kendall's τ values from simulated data with varying sample sizes, high correlation, and copula types.

Given the simulated data, the parameters of the assumed parametric copula (Gaussian copula) are estimated using the pseudo maximum likelihood method to account for the dependence structure. Using the estimated parameters $\hat{\theta}$, the probabilities $h_{ijk}(\hat{\theta})$ are computed for each scenario. The estimated parameters for these four scenarios with $n = 50$ and strong or moderate assumed correlation are shown in Tables 4.5 and 4.6.

These results show that if the generated data are highly correlated, this implies that the estimated parameters and their corresponding Kendall τ values are high. This is due to the relationship between the copula parameters and the strength of dependence between the variables. Thus, when the data are highly correlated, the estimated parameter of the copula reflects this high degree of dependence, leading to a higher value for the copula parameter. This is clear in the first two rows in each case, Tables 4.5 and 4.6.

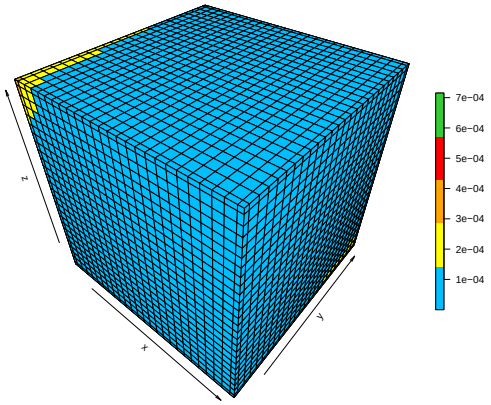
A noticeable difference appears on the last row in all cases except Case IV of Tables 4.5 and 4.6, where the estimated parameter of conditional pairs presents corresponding Kendall τ values that are usually weaker than the first two pairs. This is convenient

with the vine copula structure, where the first two pairs have stronger dependence than the third pair. The difference in Case IV is that the third pair is used without applying the simplifying assumption. Therefore, the estimated parameter value of the third pair is similar to the estimated parameter values of the first two pairs. When the data is highly correlated, the estimated copula parameter shows this high level of dependence, resulting in a higher parameter value. Similar results are obtained when the simulation is conducted for different sample sizes and correlation values reported in Appendix B, Table B.1.

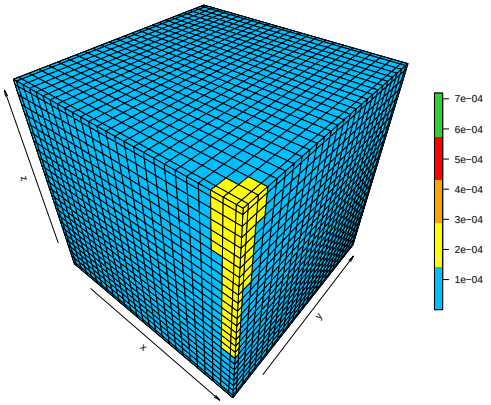
The probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ are determined after the parameters are estimated. Figure 4.4 presents three-dimensional plots of $h_{ijk}(\hat{\boldsymbol{\theta}})$ when $n = 50$ for the four scenarios. From Figure 4.4, it seems these probabilities are similar but not exactly the same in most cases. For positive, highly correlated data, the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ tend to have large probabilities when i, j, k close to each other, compared to when the data are moderately correlated for all scenarios. When the data are very weak or there is no correlation, then the marginals are highly scattered. Similar results with different correlations and varying sample sizes are visualized in Appendix B.

The probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ are used to obtain the NPI lower and upper probabilities for the event of interest, $X_{n+1} + Y_{n+1} + Z_{n+1} > t$. The NPI lower and upper probabilities are obtained by using Equations (3.2.9) and (3.2.10) in Section 3.2 and they are presented in Table 4.7 and Figure 4.5. Table 4.7 represents the NPI lower and upper probabilities at $t = 0$ for all four scenarios obtained from generated data with sample sizes $n = 10$, $n = 25$ and $n = 50$ and with the assumed correlations. Data are generated from a trivariate normal distribution with a mean vector of zero, which is symmetric around zero, one of the properties of the normal distribution. With the event of interest $X_{n+1} + Y_{n+1} + Z_{n+1} > t$, at $t = 0$, the expected value of the sum is zero, and the NPI lower and upper probabilities tend to include 0.5.

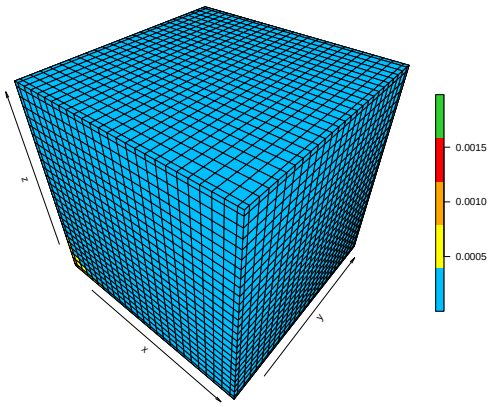
Case I **Case II**



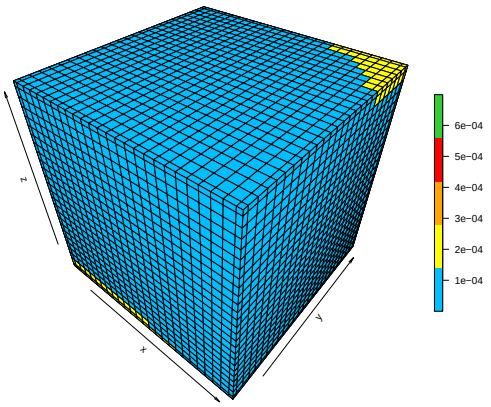
(a) Low correlated



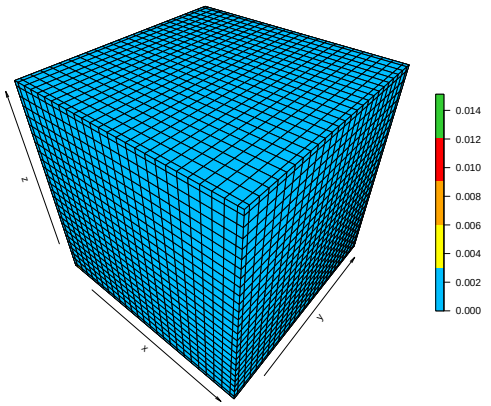
(d) Low correlated



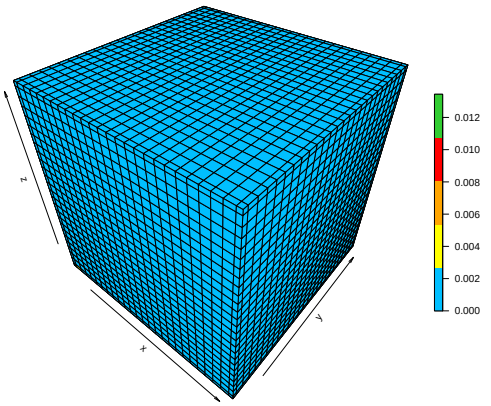
(b) Moderate correlated



(e) Moderate correlated



(c) High correlated

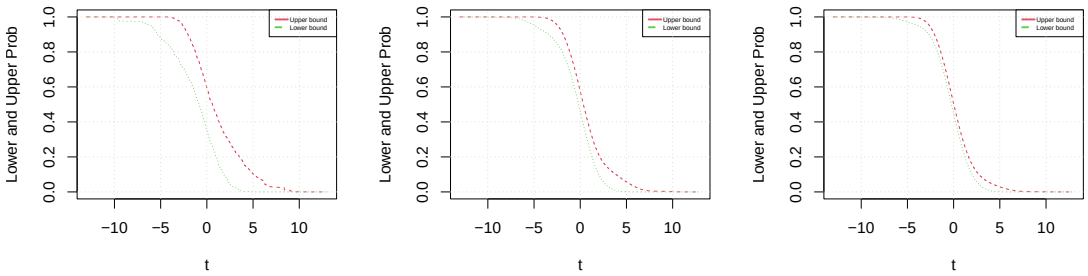


(f) High correlated

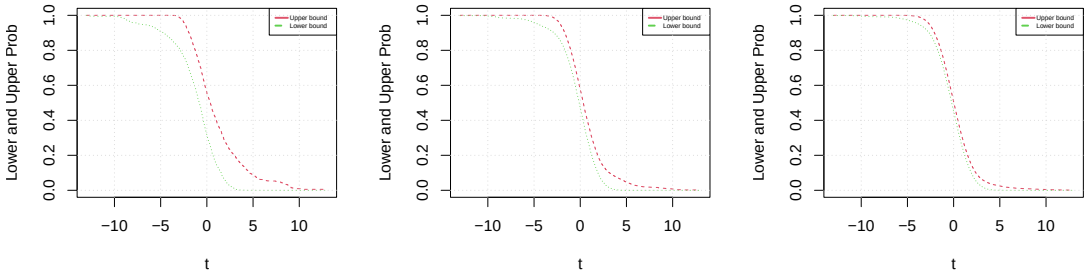
Figure 4.4: The h_{ijk} probabilities obtained from simulated data $n = 25$ using Gaussian vine copula with different correlation levels.

For all scenarios, the value 0.5 is included between the lower and upper probabilities at $t = 0$ when $n = 10$. Increasing the sample size to $n = 25$ shows the value $0.5 \in [\underline{P}, \overline{P}]$ except when the correlated data is high for Cases I, II and IV. For $n = 50$, where the generated data is moderately correlated, Cases I, II and IV show the NPI probabilities are greater than 0.5 and this is due to the randomness in the data. Also, Case III, the only case, shows $0.5 \in [\underline{P}, \overline{P}]$ regardless of the sample size or the correlation strength. This is due to the selection of a specific pair for the vine copula construction. As expected, imprecision decreases as the sample size increases, consistent with the concept of imprecision. As n increases, the imprecision of the lower and upper probabilities decreases. Also, the imprecision decreases when the correlation strength increases. This can be explained through the $h_{ijk}(\hat{\boldsymbol{\theta}})$, that is for highly positively correlated data the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ tend to include additional large probabilities. As the event of interest is on the sum, the NPI lower and upper probabilities tends to include several additional probabilities. With a high positive correlation, these additional probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ include a few larger values for most values of t, compared to a weak correlation.

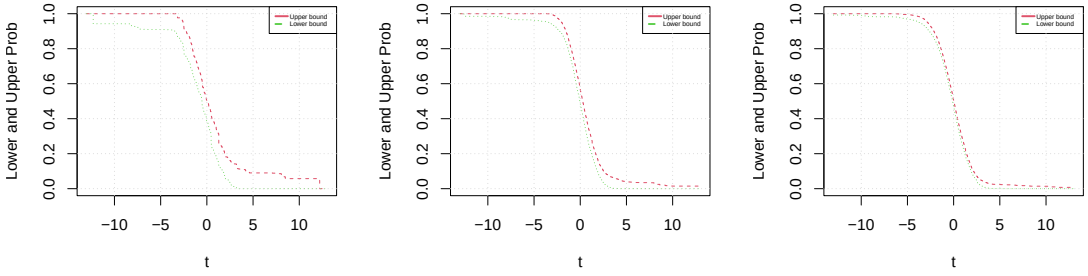
Combining NPI with a vine copula offers flexibility for modelling multivariate data. The flexibility arises when selecting the pairs for modelling dependence and selecting the copula type for each bivariate copula. Vine copula structure also reduces complexity when the dimension is increased, making it more adaptable than the classical copulas, which rely on a single parameter regardless of the dimension. Each pairwise dependence in the vine copulas can be any type of copula to capture the dependence structure.



(a) Not correlated, $n = 10$ (b) Not correlated, $n = 25$ (c) Not correlated, $n = 50$



(d) Moderate, $n = 10$ (e) Moderate, $n = 25$ (f) Moderate, $n = 50$



(g) High, $n = 10$ (h) High, $n = 25$ (i) High, $n = 50$

Figure 4.5: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, Case (III).

τ	Case	$n = 10$		$n = 25$		$n = 50$	
		$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
H	I	0.4098	0.5080	0.5091	0.5518	0.4964	0.5176
	II	0.4103	0.5082	0.5086	0.5516	0.4961	0.5174
	III	0.3857	0.5033	0.4839	0.5662	0.4674	0.5047
	IV	0.4082	0.5000	0.5041	0.5428	0.4958	0.5156
M	I	0.3675	0.5328	0.4838	0.5456	0.4636	0.4916
	II	0.3534	0.5398	0.4856	0.5627	0.4646	0.4935
	III	0.3137	0.5557	0.4709	0.5745	0.4501	0.5052
	IV	0.3607	0.5419	0.4850	0.5428	0.4683	0.4925
L	I	0.2754	0.6281	0.4356	0.5875	0.4413	0.5014
	II	0.2461	0.6385	0.4464	0.5988	0.4439	0.5045
	III	0.3500	0.5954	0.4588	0.5741	0.4447	0.5031
	IV	0.2769	0.6276	0.4449	0.5770	0.4437	0.5003

Table 4.7: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations

4.4 Predictive performance

This section presents simulation results evaluating the predictive performance of the method proposed in Section 4.2. The approach is similar to that in Section 3.5, but vine copulas are used here. For each of $N = 100$ datasets of size $n + 1$, the first n observations are used to apply the method, and the last observation evaluates predictive performance. To reduce computation time, simulations are performed for sample sizes $n = 10$ and $n = 25$.

The trivariate Gaussian distribution with mean vector zero. Three covariance matrices representing different correlation structures are considered (corresponding to High (H), Moderate (M) and Low (L) dependence, respectively) is used for the simulation, as follows:

$$\begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Given that the vine copula structure comprises bivariate copula dependence models, each of the four commonly used copulas, namely, Clayton, Gumbel, Frank, and Joe, is applied to all pairs of variables. The parameters are estimated using the pseudo-maximum likelihood method, as described in Section 2.3.

Table 4.12 represents the average of the 100 parameter estimates along with the corresponding Kendall τ values. The results indicate that the parameters are well estimated, as the true and estimated values are close to each other, especially for $n = 25$. The dependence structures are well described, as the corresponding Kendall's τ values for the true and estimated values are also close to each other. As described at the beginning of this chapter, vine copulas decompose multivariate dependencies into pairwise copulas. $\hat{\theta_3}$ corresponds to the dependence between the third pair $(Y, Z|X)$ of variables in the vine structure, in contrast to $\hat{\theta_1}$ and $\hat{\theta_2}$, which capture the strongest dependence between the first two pairs (X, Y) and (X, Z). As a result, the dependence captured by $\hat{\theta_3}$ is typically weaker, leading to a lower estimate compared to $\hat{\theta_1}$ and $\hat{\theta_2}$, which is consistent with the vine structure.

As explained in Section 3.5, the inverse values of the lower and upper survival functions of T_{n+1} for a value $q \in (0, 1)$ as defined in Equations (3.5.7) and (3.5.8) which yield the two inequalities p_1 and p_2 that are presented in Equations (3.5.9) and (3.5.10) for testing the performance. By using different quantiles to assess the method's performance, any value of q can be selected. Quantiles offer a good indicator for evaluating the method's effectiveness. Thus, for a good performance $p_1 < q < p_2$ must hold.

The results of the predictive performance are presented in Tables 4.8–4.11, which highlight the strong performance of the proposed method. Table 4.8 presents the results when the bivariate copulas in the vine structure are all Clayton copulas. The results indicate that when the sample size is $n = 10$, q typically lies within the interval $[p_1, p_2]$. For $n = 25$, the imprecision decreases, and q tends to fall outside this interval. Tables 4.8-4.11 represent the results when assuming the pair copulas in the vine structure are

all Gumbel, all Frank or all Joe copulas, respectively. These tables show that in a few cases, the q values fall outside the range $[p_1, p_2]$. At $n = 10$, there are just two values in each table that are not within the range, mainly due to randomness in the data for these copulas when the dependence level increases. However, when $n = 25$, the imprecision reduces, which leads to q falling outside $[p_1, p_2]$.

These tables clearly show that as the correlation between variables increases, imprecision decreases, regardless of the copula type.

This can be attributed to two main factors related to the event $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$, which is explained through the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$, fundamental for inference. For high positive correlation, and with the event defined as the sum, the lower and upper probabilities in Equations (3.2.9) and (3.2.10) tend to include additional $h_{ijk}(\hat{\boldsymbol{\theta}})$. These additional terms generally have large values for most t under positive correlation. Conversely, for weak correlation, the h_{ijk} probabilities where i, j, k are close are not as large as in the high positive correlation case. Also, when assuming no correlation, the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ become more scattered.

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.02	0.30	0.13	**0.23**
	0.50	0.32	0.57	0.41	0.51
	0.75	0.71	0.93	0.70	0.82
M	0.25	0.10	0.28	0.15	**0.22**
	0.50	0.36	0.54	0.40	**0.45**
	0.75	0.74	0.88	0.68	**0.74**
H	0.25	0.10	0.28	0.16	**0.22**
	0.50	0.37	0.53	0.41	**0.45**
	0.75	0.74	0.85	0.68	**0.72**

Table 4.8: Predictive performance, Clayton copula

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.07	0.29	0.15	0.25
	0.50	0.39	0.65	0.44	0.54
	0.75	0.77	0.99	0.69	0.85
M	0.25	0.14	0.30	0.19	**0.23**
	0.50	0.45	0.58	0.45	0.53
	0.75	**0.77**	0.89	0.68	0.82
H	0.25	0.16	0.29	0.19	**0.23**
	0.50	0.48	0.59	0.45	0.53
	0.75	**0.76**	0.88	0.68	0.79

Table 4.9: Predictive performance, Gumbel copula

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.06	0.33	0.12	**0.24**
	0.50	0.32	0.62	0.43	0.52
	0.75	0.72	0.99	0.68	0.85
M	0.25	0.10	0.29	0.15	**0.21**
	0.50	0.42	0.57	0.44	0.51
	0.75	**0.78**	0.90	0.71	0.82
H	0.25	0.13	0.28	0.17	**0.21**
	0.50	0.48	0.57	0.44	0.51
	0.75	**0.77**	0.88	0.71	0.81

Table 4.10: Predictive performance, Frank copula

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.08	0.31	0.16	0.25
	0.50	0.41	0.67	0.46	0.55
	0.75	**0.80**	0.99	0.69	0.85
M	0.25	0.17	0.31	0.21	0.27
	0.50	0.49	0.63	0.49	0.55
	0.75	**0.79**	0.93	0.68	0.82
H	0.25	0.16	0.29	0.19	0.25
	0.50	0.50	0.62	0.49	0.54
	0.75	0.71	0.86	0.69	0.83

Table 4.11: Predictive performance, Joe copula

Copula	τ	$n = 10$						$n = 25$					
		$\hat{\theta}_1$	τ_1	$\hat{\theta}_2$	τ_2	$\hat{\theta}_3$	τ_3	$\hat{\theta}_1$	τ_1	$\hat{\theta}_2$	τ_2	$\hat{\theta}_3$	τ_3
Clayton	L	0.58	0.19	0.66	0.20	0.50	0.16	0.31	0.12	0.37	0.14	0.23	0.09
	M	1.79	0.42	1.77	0.40	0.72	0.21	1.22	0.37	1.33	0.38	0.46	0.18
	H	3.23	0.56	3.42	0.56	0.84	0.24	2.36	0.52	2.46	0.54	0.56	0.21
Gumbel	L	1.34	0.20	1.38	0.21	1.31	0.18	1.84	0.14	1.20	0.14	1.14	0.10
	M	2.10	0.46	2.12	0.44	1.46	0.25	1.78	0.42	1.81	0.43	1.30	0.21
	H	3.16	0.62	3.13	0.61	1.62	0.30	2.56	0.59	2.62	0.60	1.38	0.25
Frank	L	1.03	0.10	1.15	0.10	1.12	0.10	0.97	0.10	0.13	0.12	0.71	0.07
	M	5.26	0.42	5.18	0.40	2.98	0.26	4.54	0.41	4.71	0.42	2.37	0.24
	H	9.34	0.59	9.20	0.59	3.92	0.31	8.04	0.59	8.23	0.59	3.03	0.30
Joe	L	1.53	0.19	1.59	0.19	1.45	0.15	1.28	0.12	1.28	0.12	1.19	0.08
	M	2.61	0.41	2.65	0.39	1.62	0.21	2.09	0.35	2.11	0.36	1.39	0.16
	H	4.33	0.56	4.09	0.54	2.07	0.25	3.13	0.51	3.21	0.52	1.49	0.19

Table 4.12: Estimated parameters and corresponding Kendall's τ values from simulated data with varying sample sizes, correlation levels, and vine copula types.

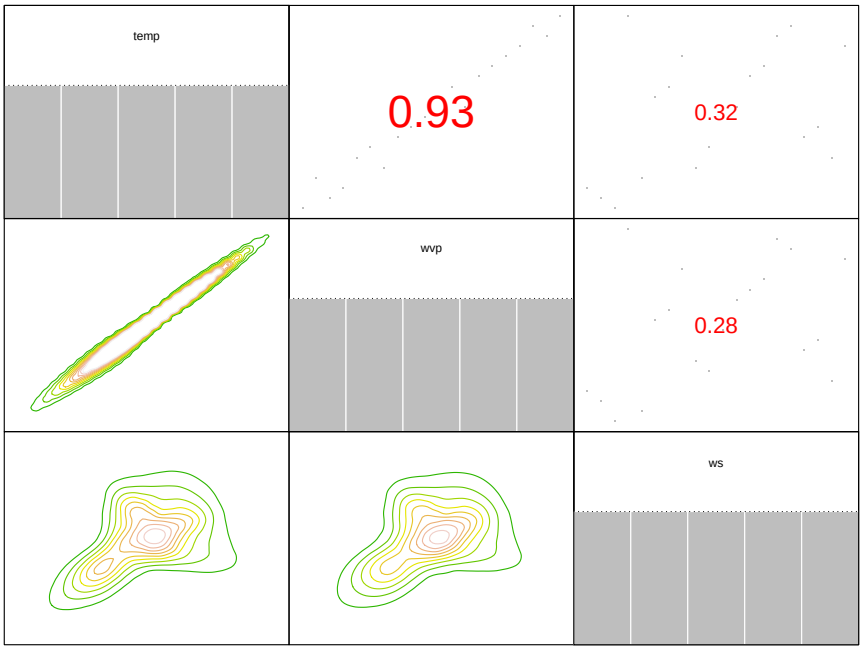


Figure 4.6: The contour bivariate plot (lower panel), the scatter plots and Kendall's tau values (upper panel) of the TMY data

4.5 Application

This section applies the proposed method to the Typical Meteorological Year (TMY) dataset previously used in Example 3.6.1. Before fitting the proposed method, there are three ways to order the pairs in a trivariate vine copula. As mentioned previously, the vine structure in three dimensions is not unique and any order of the pairs yields different results. In vine copulas, it is preferable to select the appropriate pairs of variables. The first pairs in the vine structure capture the strongest dependencies, while later pairs represent conditional dependencies, which are usually weaker than the first two pairs. Placing strongly dependent variables later in the vine model may not capture them effectively. Figure 4.6 presents the correlation between variables, showing that dry bulb temperature and water vapour pressure variables have the highest correlation (X, Y), followed by the variables water vapour pressure and wind speed (Y, Z), and then dry bulb temperature and wind speed (X, Z).

A common method used to select the pairs is to order the variables depending on

the value of Kendall's tau [2]. Once the most appropriate order of variables has been determined, the next step is to select a copula that best represents the data for each pair in order to capture dependence. Several methods can be used to select the best bivariate copula, including contour plots that visualize the dependence structure, as shown in Figure 4.6. This figure illustrates the dependence between variables, helping to identify the most suitable model for each pair based on the observed relationships. Figure 4.6 indicates that the Gumbel copula is most appropriate for modeling the dependence between dry bulb temperature and water vapour. The contour plot reveals an asymmetric structure with strong overall dependence, reflecting a high correlation between the variables. The stronger upper-tail dependence compared to the lower tail aligns with the characteristics of the Gumbel copula. For the other variable pairs, the contour plots show stronger lower-tail dependence and a more dispersed pattern in the upper tail, suggesting the use of the survival Joe copula. This type of copula is a rotated version of Joe copula rotated 180° by considering $c_{180}(x_1, x_2) = x_1 + x_2 - 1 + c(1 - x_1, 1 - x_2)$. For example, a Joe copula that exhibits upper tail dependence results in lower tail dependence when rotated by 180°.

By applying the proposed method presented in Section 4.2, where bivariate copulas are assumed in the vine structure and the pseudo maximum likelihood estimation method is used to estimate the parameters, the estimated parameters for each copula are given in Table 4.13. This table shows that the first two pairs have Kendall's tau values stronger than the third pair and this is due to the vine structure explained in Section 4.2. Figure 4.7 illustrates how the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ are influenced by the selected copulas and the estimated parameters. Three parameters control the dependence, each with a specific dependence value, as shown in Table 4.13. The probabilities with high positive correlation show that the marginals exhibit a positive relationship when i, j, k are close to each other. For very weak or zero correlation, no obvious relationship is observed. These probabilities obtained are fundamental for inference and affect the NPI lower and upper probabilities when considering an event of interest. For positive correlations, the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ tend to include additional large probabilities. Considering the event of interest as the apparent temperature tends, the NPI lower and upper probabilities tends to include several additional probabilities. The NPI lower and upper probabilities of the event $AT_{n+1} > t$ for some selected values t are presented in Table 4.14 and Figure 4.8.

Pairs	Copula Family	τ	$\hat{\theta}$
x, y	Gumbel	0.89	9.12
x, z	Survival Joe	0.38	2.12
$y, z\|x$	Survival Joe	0	1

Table 4.13: Estimated parameters and corresponding Kendall's τ values using vine copula.

t	-6.35	-4.35	-2.35	-0.35	1.64	3.64	5.64	7.64	9.64	11.64	13.64
$\underline{P}$	1.0000	0.9977	0.9845	0.9532	0.9461	0.8505	0.6308	0.2359	0.0174	0.0003	0.0000
$\overline{P}$	1.0000	1.0000	1.0000	0.9985	0.9960	0.8998	0.6350	0.2544	0.0623	0.0433	0.0399

Table 4.14: The NPI lower and upper probabilities of the event $AT_{n+1} > t$ using different vine copula at selected values of t.

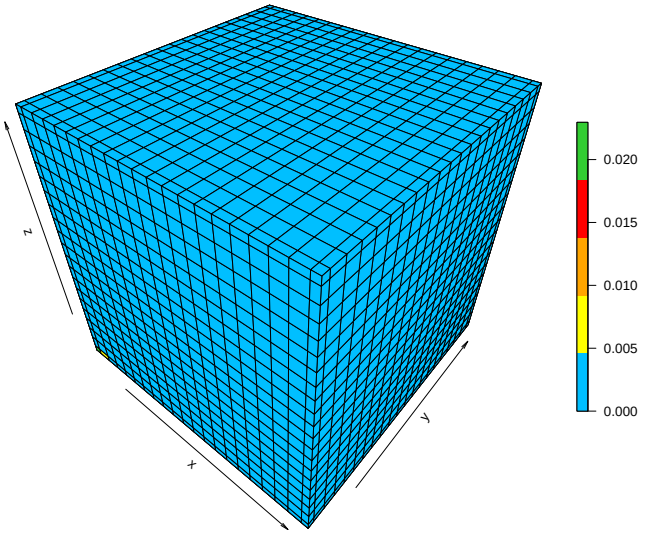


Figure 4.7: The h_{ijk} probabilities.

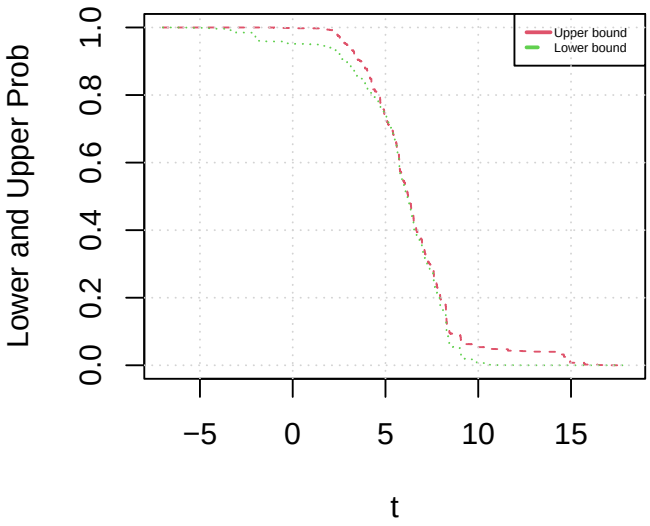


Figure 4.8: The NPI lower and upper probabilities of the event $AT_{n+1} > t$ using vine copula.

4.6 Concluding remarks

This chapter introduced the method of combining NPI with a vine copula. Vine copula is considered to be more flexible for modelling different dependencies than the classical copula, which relies on a single parameter. It can model multiple dependence structures by specifying each bivariate copula with a chosen copula type. Computing the probabilities h_{ijk}, which are fundamental for inference, is the most challenging step in this method. In a trivariate copula model, the challenge arises in capturing the dependencies between the three variables. Modelling the pairwise dependencies in such a vine copula structure can be complex, particularly when there are multiple copulas involved. The iterated proportional fitting (IPF) is often used to fit the copula model, adjusting the marginals while preserving the dependencies among the three variables. This is obvious from the provided example in Section 4.2 showing that the marginals were not identical and this does not satisfy the h_{ijk} conditions. By using the IPF algorithm that makes, the probabilities h_{ijk} are made to satisfy their conditions. These probabilities were investigated and visualized with varying correlation values.

The proposed method is illustrated by considering different scenarios and an example based on data from the literature. The performance is evaluated through simulations. The method performs well in general except when the vine structure contains Clayton copulas with $n = 25$,where the imprecision decreases leading to $q \notin [p_1, p_2]$. Throughout this work, the focus was on using trivariate parametric vine copulas, with the simplifying assumption outlined in Section 2.4. However, extending this approach to higher dimensions is worth exploring, though it introduces additional computational challenges. It may be beneficial to model the vine copula without the simplifying assumption, as this could provide a more accurate representation of dependencies, though at the cost of increased complexity. This remains an area for future work, as further exploration is needed to fully assess its potential and computational implications.

Chapter 5

NPI Combined with FNAC

5.1 Introduction

As the main goal of this thesis is to develop NPI-based methods for multivariate data, this chapter presents a method that combines nonparametric predictive inference (NPI) with fully nested Archimedean copulas (FNAC) for predictive inference. In previous chapters, NPI was combined with dependence models such as classical copulas with one parameter and vine copulas. This chapter extends that work by introducing FNAC as an alternative dependence model within the NPI framework.

This chapter is organized as follows: In Section 5.2, the method of NPI combined with trivariate FNAC is introduced. The effectiveness of this newly proposed method is illustrated in Section 5.3. The performance of this method is evaluated in Section 5.4. Two examples from the literature to illustrate the application of the proposed method to the real world are presented in Section 5.5. A comparison study is conducted to evaluate the performance of all the methods proposed throughout this thesis in Section 5.6. The comparison highlights the strengths and limitations of each approach and provides insight into their applicability. Some concluding remarks are included in Section 5.7.

5.2 Combining NPI with FNAC

This section introduces the method of combining NPI with a parametric trivariate fully nested Archimedean copulas in two stages. The first stage applies NPI for the marginals

and then in the second step, a trivariate parametric FNAC is assumed and the parameters are estimated to take into account the dependence structure.

For the first stage, which applies NPI for the marginals, assume that there are n trivariate observations (x_i, y_i, z_i), $i = 1, \ldots, n$, which are the observed values of n exchangeable trivariate random quantities with no ties. The observations of the marginals are ordered and denoted by x_i, y_j and z_k for simplicity, so $x_1 < \cdots < x_i < \cdots < x_n$, $y_1 < \cdots < y_j < \cdots < y_n$ and $z_1 < \cdots < z_k < \cdots < z_n$.

Using Hill's assumption $A_{(n)}$, it is possible to derive a partially specified predictive probability distribution for X_{n+1}, Y_{n+1} and Z_{n+1} given the observations $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ and $z_1, \ldots, z_n$ that, respectively, lead to $P(X_{n+1} \in (x_{i-1}, x_i)) = \frac{1}{n+1}$, $P(Y_{n+1} \in (y_{j-1}, y_j)) = \frac{1}{n+1}$ and $P(Z_{n+1} \in (z_{k-1}, z_k)) = \frac{1}{n+1}$ for i, j, $k \in \{ 1, \ldots, n+1\}$, where $x_0 = -\infty, x_{n+1} = \infty$, $y_0 = -\infty, y_{n+1} = \infty$ and $z_0 = -\infty, z_{n+1} = \infty$.

To link the first stage with the second stage, where the dependence structure in the data is taken into account to provide a partially specified predictive distribution for the trivariate $(X_{n+1}, Y_{n+1}, Z_{n+1})$ is by introducing a natural transformation of the three random quantities individually as introduced by Muhammad [76]. Let $\tilde{X}_{n+1}$, $\tilde{Y}_{n+1}$ and $\tilde{Z}_{n+1}$ denote the transformed versions of the random quantities X_{n+1} , Y_{n+1} and Z_{n+1} respectively, such that

$$(X_{n+1} \in (x_{i-1}, x_i), Y_{n+1} \in (y_{j-1}, y_j, Z_{n+1} \in (z_{k-1}, z_k)) \Longleftrightarrow$$
$$(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1})) \quad (5.2.1)$$

where i, j and $k \in \{ 1, \ldots, n+1\}$. This transformation from the real space $\mathbb{R}^3$ to $[0,1]^3$ is based on n trivariate data, where $[0,1]^3$ is divided into $(n+1)^3$ equal sized blocks. By following these transformations of the marginals, the uniform marginal distribution on $[0,1]$ has been discretized. The $A_{(n)}$ assumption for the marginals after the transformation lead to

$$P(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1})) = P(X_{n+1} \in (x_{i-1}, x_i)) = \frac{1}{n+1} \quad (5.2.2)$$

$$P(\tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}) = P(Y_{n+1} \in (y_{j-1}, y_j)) = \frac{1}{n+1} \quad (5.2.3)$$

and

$$P(\tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1}) = P(Z_{n+1} \in (z_{k-1}, z_k)) = \frac{1}{n+1} \quad (5.2.4)$$

For the second stage, a trivariate parametric FNAC is assumed with parameters θ_1 and θ_2 to capture the dependence structure and the parameters are estimated. The parameters can be estimated where the observed pair (x_i, y_i), $i = 1, \ldots, n$ is replaced by $(\frac{r_i^x}{n+1}, \frac{r_i^y}{n+1})$, after that assume a bivariate parametric Archimedean copula to couple $(\frac{r_i^x}{n+1}, \frac{r_i^y}{n+1})$ with $\frac{r_i^z}{(n+1)}$ where r_i^x is the rank of the observation x_i among x-observations r_i^y is the rank of the observation y_i among y-observations and r_i^z is the rank of the observation z_i among z-observations.

Now, NPI on the marginals can now be combined with the estimated copulas by defining the probability for the event that the transformed variables $(\tilde{X}_{n+1}, \tilde{Y}_{n+1}, \tilde{Z}_{n+1})$ belongs to a specific block out of the $(n+1)^3$ blocks into which the space $[0,1]^3$ has been partitioned.

$$h_{ijk}(\hat{\theta}_1; \hat{\theta}_2) =$$
$$P_C(\tilde{X}_{n+1} \in (\frac{i-1}{n+1}, \frac{i}{n+1}), \tilde{Y}_{n+1} \in (\frac{j-1}{n+1}, \frac{j}{n+1}), \tilde{Z}_{n+1} \in (\frac{k-1}{n+1}, \frac{k}{n+1}) | \hat{\theta}_1; \hat{\theta}_2) \quad (5.2.5)$$

where i, j and $k \in \{1, \ldots, n+1\}$ and $P_C(.|\hat{\theta}_1; \hat{\theta}_2)$ is the assumed copula-based probability with estimated parameters $\hat{\theta}_1$ and $\hat{\theta}_2$. These values $(n+1)^3$ of $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ that sum up to one provide a fully discretized probability distribution for the transformed future observations. This distribution can be used for making inferences about the actual future observation or any event of interest. The $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ probabilities satisfy

1. $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\theta}_1; \hat{\theta}_2) = 1$
2. $\sum_{j=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\theta}_1; \hat{\theta}_2) = \frac{1}{n+1}$, for $i \in \{1, 2, \ldots, n+1\}$, $\sum_{i=1}^{n} \sum_{k=1}^{n} h_{ijk}(\hat{\theta}_1; \hat{\theta}_2) = \frac{1}{n+1}$, for $j \in \{1, 2, \ldots, n+1\}$ and $\sum_{i=1}^{n} \sum_{j=1}^{n} h_{ijk}(\hat{\theta}_1; \hat{\theta}_2) = \frac{1}{n+1}$, for $k \in \{1, 2, \ldots, n+1\}$
3. $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2) \geq 0$, for i, j,$k \in \{1, \ldots, n+1\}$

The method of combining NPI with multivariate data can be extended beyond the trivariate case by following two main steps: applying NPI for the marginals in the first step and then assuming a parametric FNAC to capture the dependence structure through the estimated parameters as follows. Assume that there are n observations of d multivariate random variables $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_d}$ where $\mathbf{X_i} = (X_{1,i}, X_{2,i}, \ldots, X_{d,i})$, $i = 1, \ldots, n$. We are interested in prediction in event involving these future observations multivariate observation as $(X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1})$.

Using Hill's assumption $A_{(n)}$, it is possible to derive a partially specified predictive probability distribution for each of $X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1}$ given their observations $\mathbf{x_i} = \mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_d}$, respectively, $\mathbf{x_i} = (x_{1,i}, \ldots, x_{d,i})$, these as follows:

$P(X_{1,n+1} \in (x_{1,i_1-1}, x_{1,i_1})) = \frac{1}{n+1}$, $P(X_{2,n+1} \in (x_{2,i_2-1}, x_{2,i_2})) = \frac{1}{n+1}, \ldots$, $P(X_{d,n+1} \in (x_{d,i_d-1}, x_{d,i_d})) = \frac{1}{n+1}$ for $i_1, i_2, \ldots, i_d \in \{ 1, \ldots, n+1\}$, where $x_{1,0} = x_{2,0} = ... = x_{d,0} = -\infty$ and $x_{1,n+1} = x_{2,n+1} = ... = x_{d,n+1} = \infty$ are introduced for notation simplicity.

The two steps can be linked by introducing a natural transformation of the random quantities individually. Let $\tilde{X}_{1,n+1}, \tilde{X}_{2,n+1}, \ldots, \tilde{X}_{d,n+1}$ be the transformed versions of the random quantities $X_{1,n+1}, X_{2,n+1}, \ldots, X_{d,n+1}$, respectively, such that

$$(X_{1,n+1} \in (x_{1,i_1-1}, x_{1,i_1}), X_{2,n+1} \in (x_{2,i_2-1}, x_{2,i_2}), \ldots, X_{d,n+1} \in (x_{d,i_d-1}, x_{d,i_d})) \Longleftrightarrow$$
$$(\tilde{X}_{1,n+1} \in (\frac{i_1-1}{n+1}, \frac{i_1}{n+1}), \tilde{X}_{2,n+1} \in (\frac{i_2-1}{n+1}, \frac{i_2}{n+1}), \ldots, \tilde{X}_{d,n+1} \in (\frac{i_d-1}{n+1}, \frac{i_d}{n+1})) \quad (5.2.6)$$

for $i_1, i_2, \ldots, i_d = 1, \ldots, n+1$. The assumption $A_{(n)}$ of the transformations lead to

$$P(\tilde{X}_{1,n+1} \in (\frac{i_1-1}{n+1}, \frac{i_1}{n+1})) = P(X_{1,n+1} \in (x_{1,i_1-1}, x_{1,i_1})) = \frac{1}{n+1} \quad (5.2.7)$$

$$P(\tilde{X}_{2n+1} \in (\frac{i_2-1}{n+1}, \frac{i_2}{n+1})) = P(X_{2,n+1} \in (x_{2,i_2-1}, x_{2,i_2})) = \frac{1}{n+1} \quad (5.2.8)$$

$$P(\tilde{X}_{d,n+1} \in (\frac{i_d-1}{n+1}, \frac{i_d}{n+1})) = P(X_{d,n+1} \in (x_{d,i_d-1}, x_{d,i_d})) = \frac{1}{n+1} \quad (5.2.9)$$

For the second step when assuming a parametric FNAC with parameters $\boldsymbol{\theta}$ and estimate the parameters is by using the transformed data, where the observed pairs are replaced by $(\frac{r_i^{x_1}}{n+1}, \ldots, \frac{r_i^{x_d}}{n+1})$ wher $r_i^{x_j}$ the rank of the observation x_i among n-x_j observations. NPI on the marginals is now combined with the estimated copula to provide a partially specified predictive distribution for one future multivariate observation and each $(n+1)^d$ blocks is assigned a specific probability as

$$h_{i_1 i_2 \ldots i_d}(\boldsymbol{\theta}) =$$
$$P_C(\tilde{X}_{1,n+1} \in (\frac{i_1-1}{n+1}, \frac{i_1}{n+1}), \tilde{X}_{2,n+1} \in (\frac{i_2-1}{n+1}, \frac{i_2}{n+1}), \ldots, \tilde{X}_{d,n+1} \in (\frac{i_d-1}{n+1}, \frac{i_d}{n+1}) | (\boldsymbol{\theta})) \quad (5.2.10)$$

where $i_1, i_2, \ldots, i_d \in \{ 1, \ldots, n+1\}$. $P_C(.|\boldsymbol{\theta})$ is the assumed copula-based probability and $\boldsymbol{\theta}$ is the estimated parameters values. These $(n+1)^d$ values of $h_{i_1 i_2 \ldots i_d}(\boldsymbol{\theta})$ provide the fully

High Correlation			No Correlation		
X	Y	Z	X	Y	Z
-1.023	-1.038	-0.825	0.687	1.066	0.537
-0.553	-0.327	-0.840	1.906	1.063	1.370
0.796	0.413	0.340	0.528	0.403	1.168
0.087	0.606	0.053	0.793	-1.214	-1.006

Table 5.1: Simulated data from a trivariate Gaussian distribution with different correlation structures.

discretized probability distribution for the transformed future observations, which can be used for statistical inference on the future observation or an event of interest involving the future observation. These $h_{i_1 i_2 \ldots i_d}(\boldsymbol{\theta})$ probabilities satisfy the following conditions:

1. $\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} h_{i_1 i_2 \ldots i_d}(\boldsymbol{\theta}) = 1$
2. $\sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} h_{i_1 i_2 \ldots i_d}(\boldsymbol{\theta}) = \frac{1}{n+1}$, for all $i_1 \in \{\ 1, 2, \ldots, n+1\}$ this summation condition is repeated for each marginal by fixing a different index and summing over the others.
3. $h_{i_1 i_2 \ldots i_d}(\boldsymbol{\theta}) \geq 0$, for all $i_1, i_2, \ldots, i_d = \in \{\ 1, \ldots, n+1\}$

Example 5.2.1 Two three-dimensional visualizations of the probabilities $h_{ijk}(\boldsymbol{\theta})$ under different dependence structures are shown in Figures 5.1 and 5.2. Each figure is based on a dataset of size $n = 4$, simulated from a trivariate Gaussian distribution with mean vector zero. Two cases are considered: a no correlation case where the covariance matrix is the identity matrix and a high correlation case where all off-diagonal entries of the covariance matrix are set to 0.9 as given in Table 5.1.

The pseudo maximum likelihood estimation method is used to estimate the parameters and the selected copula is the Frank FNAC. As mentioned in Section 2.5, FNAC is restricted to the Archimedean family and the Frank copula belongs to this family. The relationship between the parameter values and their associated Kendall τ values demonstrates that the dependence structure is controlled by two parameters in a trivariate FNAC. The first parameter θ_1 is for the variables X and Y and the second parameter

Pairs	τ	$\hat{\theta}$	τ	$\hat{\theta}$
(x, y)	0.69	10.90	0.42	4.49
$(; z)$	0.63	8.83	0.42	4.49

Table 5.2: Estimated parameters and corresponding Kendall's τ values from simulated data with varying correlation levels using Frank FNAC.

is θ_2 for the variables Z and the bivariate copula of (X, Y) where $\theta_1 > \theta_2$. For the two datasets, the estimated parameters with the corresponding Kendall τ values are presented in Table 5.2. Each dataset indicated that $\theta_1 \geq \theta_2$ and for the second dataset, the estimated parameter indicated a weaker positive dependence compared to the first dataset.

The estimated parameter is used to calculate the probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ as given by Equation (5.2.5). Figures 5.1 and 5.2 present the marginal probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ where each figure presents three sides: the right side $h_{.jk} = \sum_i h_{ijk}$, the left side $h_{i.k} = \sum_j h_{ijk}$ and the bottom side $h_{ij.} = \sum_k h_{ijk}$. On each side, there are large values when $i = j = k$ when the dataset is highly correlated compared to when the data is weakly correlated. There is also a symmetry around these large values on each side. This effect is due to the FNAC structure having two parameters to capture dependence.

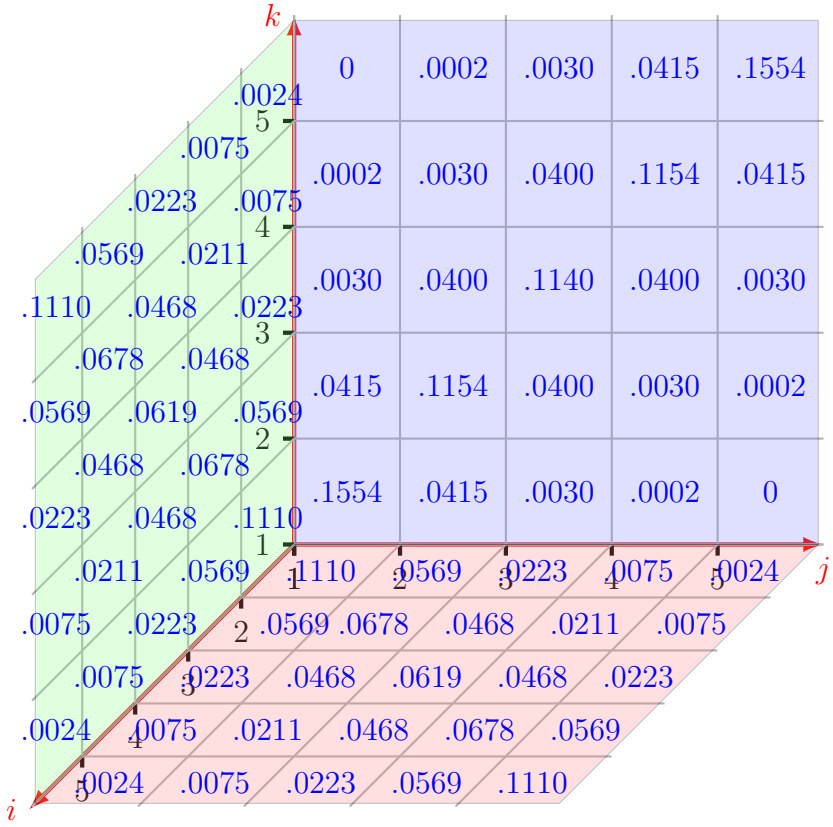


Figure 5.1: The probabilities $h_{.jk}, h_{i.k}$ and $h_{ij.}$, for the high correlation case.

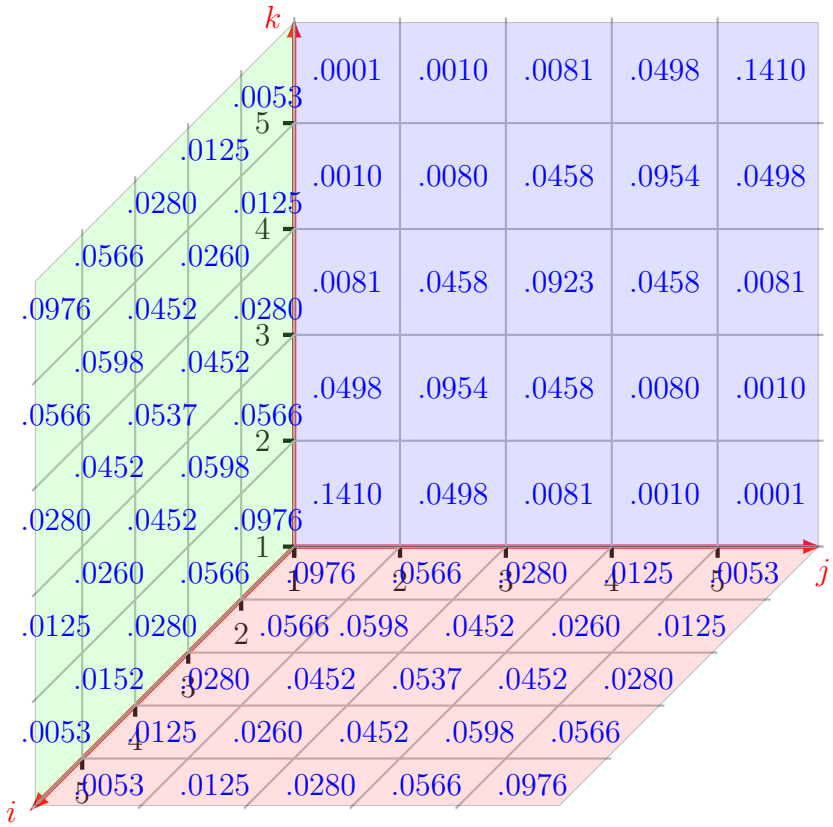


Figure 5.2: The probabilities $h_{.jk}, h_{i.k}$ and $h_{ij.}$, for the no correlation case.

5.3 Example

This section presents the NPI lower and upper probabilities to illustrate the proposed method introduced in Section 5.2. using trivariate Gaussian datasets with zero mean vectors. Three covariance matrices representing different correlation structures are considered, with sample sizes $n = 10, 25, 50$ using the R package *mvtnorm* [46]. These covariance matrices are as follows:

$$\begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0.15 & 0.15 \\ 0.15 & 1 & 0.15 \\ 0.15 & 0.15 & 1 \end{pmatrix}$$

The dependence strength is defined as High (H), Moderate (M), and Low (L) corresponding to assumed correlations of 0.9, 0.5 and 0.15, respectively. First four parametric FNAC types are used: Clayton, Gumbel, Frank and Joe, as presented in Section 2.5. The pseudo maximum likelihood method, commonly used to estimate copula parameters, is implemented via the R package *copula* [100]. All copula components within the FNAC structure belong to the same Archimedean family.

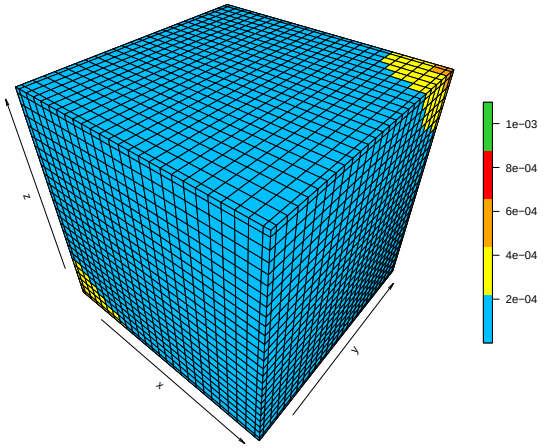
Following the method presented in Section 5.2, which applies NPI for the marginals and estimating the parameters for the assumed FNAC to capture the dependence structure, Table 5.3 presents the estimated parameters for the four assumed FNACs. The results show that the parameter estimates generally increase with stronger correlations in the generated data, is consistent with expectations. Moreover, the results satisfy the FNAC parameter condition: the parameter in the first nesting level is greater than or equal to the parameter in the second nesting level.

The probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ can be obtained using Equation (5.2.5). Figure 5.3 presents a three-dimensional plot of these probabilities for the Frank trivariate FNAC across different correlation levels and sample sizes. Generally, the figures show that the probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ are similar but not identical within each block. This pattern becomes more apparent in the top-right region of the plot as the sample size decreases. For all FNAC models, highly correlated data lead to larger values of $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ when i, j and k are close to each other, compared to cases with moderate or weak correlations. When the data are weakly correlated, the probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ still show a positive pattern

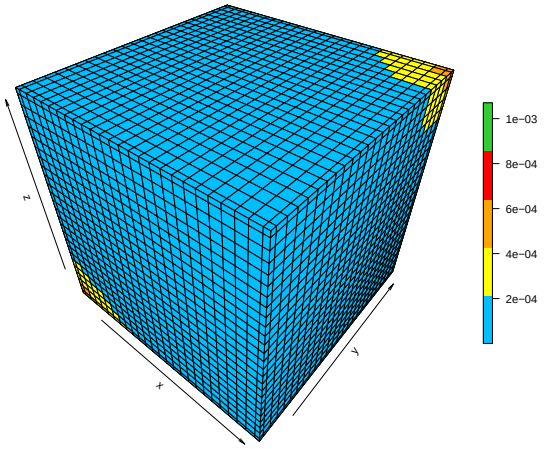
when i, j and k are close to each other, although this pattern is weaker compared to cases with stronger correlation. Similar results are observed when using other FNAC types. Full details are presented in Appendix C.

Assume that the event of interest is $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$. The NPI lower and upper probabilities are presented in Figure 5.4, which shows the NPI lower and upper probabilities with Gumbel FNAC, obtained from datasets with low, moderate and high correlations and different sample sizes. The NPI lower and upper probabilities for selected t values are presented in Table 5.4. The results show that imprecision is larger when the data exhibit a weak positive correlation than when the correlation is strong. This occurs because the event of interest, $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$, can be explained using the probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$, which play a fundamental role for inference. For positively correlated data, $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ have larger values when i, j and k are close to each other. Consequently, the calculation of lower and upper probabilities, as in Equations (3.2.9) and (3.2.10), includes more of these high probabilities. For the event of interest T_{n+1}, the NPI lower and upper probabilities are typically include these additional $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ of t resulting in less imprecision.

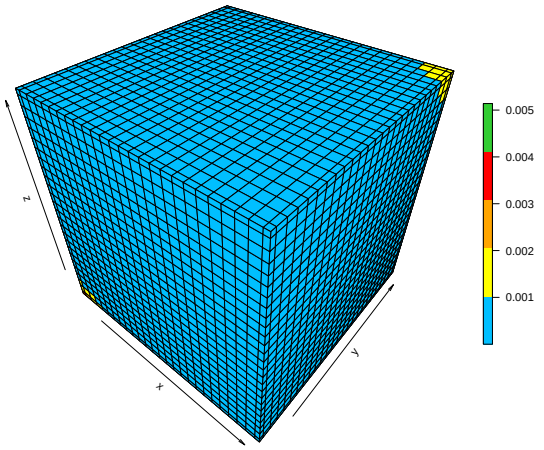
Since the data are generated from a trivariate Gaussian distribution with a zero mean vector and the event of interest involves the sum $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$, the expected value of the sum is zero. In many cases, the value 0.5 is included within the NPI lower and upper probabilities. However, in some instances such as when $t = 0$ the value 0.5 is not included due to randomness in the data.



(a) Low correlated



(b) Moderate correlated



(c) High correlated

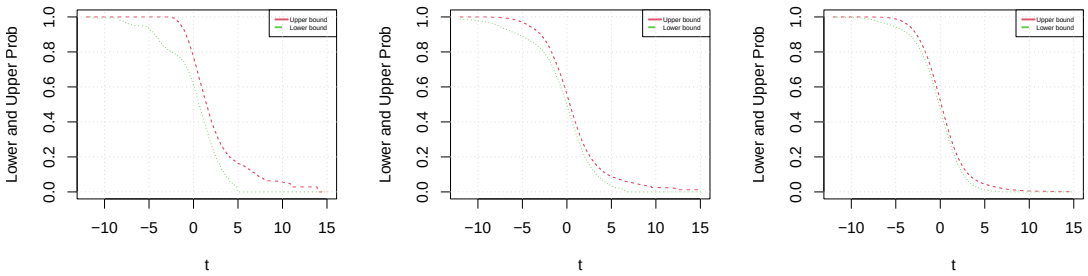
Figure 5.3: The h_{ijk} probabilities obtained from simulated data $n = 25$ using Frank FNAC with different correlation levels.

τ	n	Clayton		Gumbel		Frank		Joe	
		τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$
	10	0.67	4.04	0.64	2.81	0.61	8.09	0.55	3.32
		0.58	2.77	0.48	1.94	0.48	5.35	0.38	2.10
H	25	0.73	5.46	0.75	4.02	0.75	14.48	0.66	4.75
		0.59	2.83	0.67	3.07	0.65	9.61	0.60	3.87
	50	0.66	3.95	0.72	3.56	0.73	12.79	0.63	4.24
		0.61	3.14	0.71	3.39	0.70	11.45	0.63	4.20
	10	0.41	1.39	0.33	1.50	0.29	2.84	0.29	1.72
		0.31	0.91	0.22	1.28	0.23	2.21	0.08	1.15
M	25	0.38	1.21	0.42	1.72	0.44	4.70	0.34	1.95
		0.30	0.85	0.36	1.55	0.33	3.29	0.30	1.77
	50	0.37	1.20	0.37	1.59	0.39	3.97	0.29	1.72
		0.30	0.87	0.35	1.54	0.36	3.68	0.28	1.70
	10	0.38	1.21	0.38	1.62	0.35	3.58	0.31	1.82
		0.35	1.10	029	1.41	0.31	3.04	0.19	1.42
L	25	0.46	1.73	0.38	1.60	0.40	4.23	0.27	1.67
		0.31	0.89	0.32	1.47	0.35	3.52	0.26	1.64
	50	0.21	0.54	0.14	1.16	0.17	1.53	0.07	1.14
		0.09	0.21	0.08	1.08	0.10	0.91	0.04	1.07

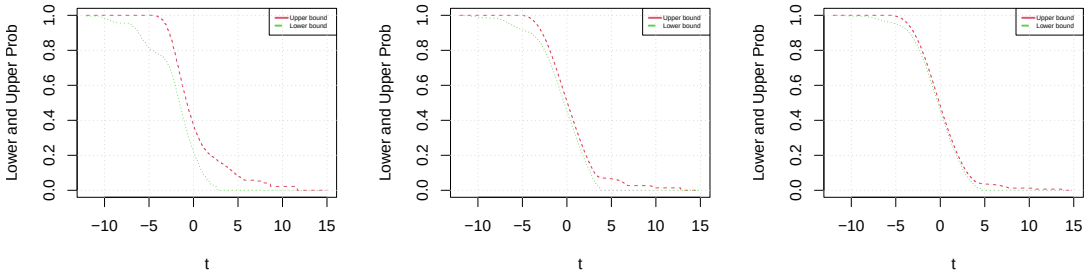
Table 5.3: Estimated parameters and corresponding Kendall's τ values from simulated data with varying sample sizes, correlation levels, and FNAC types.

τ	n	t	Clayton		Gumbel		Frank		Joe	
			$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
		-6.00	0.8816	1.0000	0.8467	1.0000	0.8374	1.0000	0.8555	1.0000
	10	-3.00	0.7619	0.8527	0.7387	0.8711	0.7326	0.8628	0.7089	0.9030
		0.00	0.2389	0.3652	0.2116	0.3184	0.2301	0.3460	0.2021	0.3039
H		-6.00	0.9506	1.0000	0.9302	1.0000	0.9186	1.0000	0.9237	1.0000
	25	-3.00	0.8263	0.8647	0.8232	0.8679	0.8126	0.8591	0.8141	0.8884
		0.00	0.5104	0.5512	0.4816	0.5229	0.4900	0.5290	0.4640	0.5034
		-6.00	0.9748	1.0000	0.9614	1.0000	0.9520	1.0000	0.9534	1.0000
	50	-3.00	0.8269	0.8461	0.8244	0.8468	0.8137	0.8360	0.8221	0.8587
		0.00	0.5147	0.5357	0.4843	0.5052	0.4940	0.5137	0.4661	0.4857
		-6.00	0.8675	1.0000	0.8859	1.0000	0.8776	1.0000	0.9135	1.0000
	10	-3.00	0.7794	0.8922	0.7333	0.9335	0.7303	0.9273	0.7230	0.9573
		0.00	0.2573	0.4260	0.2222	0.3733	0.2397	0.4017	0.1990	0.3706
M		-6.00	0.9375	1.0000	0.9387	1.0000	0.9351	1.0000	0.9567	1.0000
	25	-3.00	0.8577	0.8999	0.8438	0.9188	0.8319	0.9107	0.8453	0.9460
		0.00	0.4999	0.5603	0.4533	0.5081	0.4717	0.5239	0.4253	0.4837
		-6.00	0.9654	1.0000	0.9635	1.0000	0.9609	1.0000	0.9759	1.0000
	50	-3.00	0.8639	0.8837	0.8656	0.9038	0.8486	0.8884	0.8803	0.9337
		0.00	0.5091	0.5402	0.4533	0.4823	0.4757	0.5021	0.4237	0.4553
		-6.00	0.9202	1.0000	0.9491	1.0000	0.9477	1.0000	0.9677	1.0000
	10	-3.00	0.8507	1.0000	0.8189	1.0000	0.8123	1.0000	0.8248	1.0000
		0.00	0.6625	0.7704	0.6113	0.7741	0.6153	0.7591	0.5918	0.7996
L		-6.00	0.9166	0.9605	0.9109	0.9853	0.9032	0.9856	0.9197	0.9949
	25	-3.00	0.8493	0.8979	0.8223	0.9024	0.8136	0.8901	0.8168	0.9175
		0.00	0.5576	0.6120	0.4959	0.5579	0.5159	0.5680	0.4659	0.5365
		-6.00	0.9533	0.9852	0.9601	0.9965	0.9576	0.9962	0.9650	0.9980
	50	-3.00	0.8692	0.9053	0.8709	0.9227	0.8624	0.9136	0.8790	0.9344
		0.00	0.5154	0.5582	0.4739	0.5211	0.4854	0.5288	0.4694	0.5219

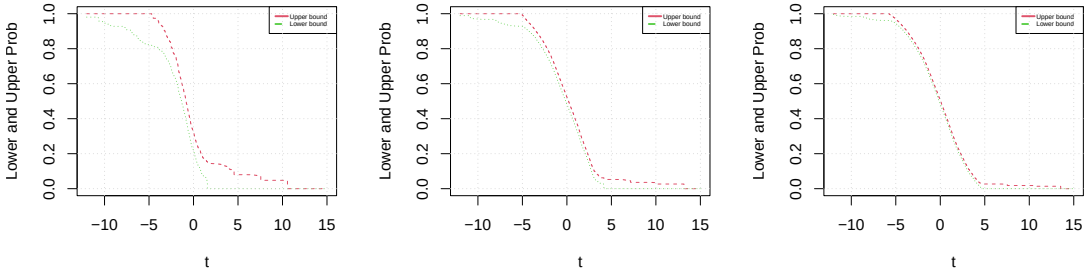
Table 5.4: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and FNAC types.



(a) Low, $n = 10$ (b) Low, $n = 25$ (c) Low, $n = 50$



(d) Moderate, $n = 10$ (e) Moderate, $n = 25$ (f) Moderate, $n = 50$



(g) High, $n = 10$ (h) High, $n = 25$ (i) High, $n = 50$

Figure 5.4: The NPI lower and upper probabilities of the event $T_{n+1} > t$, based on simulated data with different sample sizes, correlations, and Gumbel FNAC.

5.4 Predictive performance

This section presents simulation studies to evaluate the performance of the proposed methods in Section 5.2. This section uses a similar evaluation procedure to that described in Section 3.5, but FNAC is used to model the dependence structure.

A total of $N = 100$ datasets, each of size $n + 1$, are generated. For each dataset, the first n observations are used to implement the proposed method and the last observation is used to assess predictive performance. To reduce computational demands, simulations are conducted with $N = 100$ runs and sample sizes $n = 10$ and $n = 25$. Since the FNAC structure contains nested Archimedean copulas, this study uses four types of FNAC for generating data: Clayton, Gumbel, Frank and Joe. In this study, Kendall's tau values are $\tau = 0.25, 0.50, 0.75$. The terms High (H), Moderate (M) and Low (L) are used as abbreviations to indicate the strength of dependence in the results. The parameters are estimated by using the pseudo maximum likelihood estimation method and are presented in Section 2.3.

The estimated parameters of each type of FNAC are shown in Table 5.9. This table shows the average of 100 parameter estimates together with the associated Kendall's τ values. The estimated parameters closely align with the true values, particularly for the sample size of $n = 25$, indicating strong estimation accuracy. Similarly, the estimated Kendall's τ values closely match the true values, reflecting an accurate representation of the dependence structure and overall, the dependence is well captured.

Applying the method introduced in Section 5.2, the trivariate FNAC is assumed to be from the same parametric family as that used for data generation. The results are presented in Tables 5.5-5.8 where p_1 and p_2 are obtained using Equations (3.5.9) and (3.5.10). The method performs effectively when q lies within $[p_1, p_2]$. The results in these tables highlight the good performance of the proposed method in general. This is consistent with expectations, as the same parametric FNAC family is used in both the method and the data simulation.

In the case of $n = 25$, the imprecision, measured by the difference $p_2 - p_1$, is reduced compared to when $n = 10$. This aligns with the general principle that larger sample sizes lead to less imprecision, especially in methods involving imprecise probabilities, which is a common feature of methods based on imprecise probabilities. A notable feature of the

τ	q	$n=10$		$n=25$	
		p_1	p_2	p_1	p_2
L	0.25	0.15	0.30	0.14	**0.19**
	0.50	0.39	0.60	0.44	0.54
	0.75	0.69	0.85	0.73	0.78
M	0.25	0.15	0.29	0.24	0.27
	0.50	0.37	0.52	0.54	0.57
	0.75	0.69	0.81	0.74	0.77
H	0.25	0.25	0.32	0.19	0.27
	0.50	0.43	0.55	0.41	**0.46**
	0.75	0.66	0.81	0.70	0.70

Table 5.5: Predictive performance, Clayton FNAC

τ	q	$n=10$		$n=25$	
		p_1	p_2	p_1	p_2
L	0.25	0.22	0.36	0.22	0.35
	0.50	0.49	0.69	0.50	0.62
	0.75	0.75	0.87	0.79	0.81
M	0.25	0.25	0.34	0.24	0.25
	0.50	0.44	0.61	0.50	0.54
	0.75	0.73	0.84	0.72	0.77
H	0.25	0.18	0.28	0.24	0.28
	0.50	0.50	0.58	0.49	0.54
	0.75	0.71	0.80	0.73	0.77

Table 5.6: Predictive performance, Gumbel FNAC

proposed method is observed when comparing cases with high correlation to those with weak correlation. The imprecision is reduced when the correlation in the generated data is higher, compared to when it is weak. This occurs due to considering an event of interest $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$ and this can be illustrated through the probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$, which are fundamental to the proposed method for making inferences.

As discussed earlier in this chapter, the level of dependence influences these probabilities, as demonstrated in Example 5.2.1, where two parameters control the dependence in trivariate FNAC. When there is a high positive correlation, the probabilities $h_{ijk}(\hat{\theta}_1; \hat{\theta}_2)$ are large when i, j, k are close to each other. When considering an event of interest as the sum, the lower and upper probabilities, as defined in Equations (3.2.9) and (3.2.10), tend to include additional $h_{ijk}(\hat{\boldsymbol{\theta}})$ and with a positive high correlation, these additional $h_{ijk}(\hat{\boldsymbol{\theta}})$ include large values for most values of t. In contrast, for moderate correlations, the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ are smaller when i, j, k are close to each other, compared to the case of a high positive correlation. Also, when assuming no correlation, the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ become more scattered.

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.25	0.36	**0.26**	0.36
	0.50	0.43	0.55	**0.54**	0.61
	0.75	0.64	0.82	**0.81**	0.87
M	0.25	0.23	0.30	**0.26**	0.30
	0.50	0.41	0.51	0.46	0.52
	0.75	0.65	0.78	0.75	0.80
H	0.25	0.18	0.36	0.17	0.26
	0.50	0.50	0.62	**0.53**	0.59
	0.75	0.73	0.82	0.70	**0.74**

Table 5.7: Predictive performance, Frank FNAC

τ	q	$n = 10$		$n = 25$	
		p_1	p_2	p_1	p_2
L	0.25	0.13	0.25	0.19	**0.21**
	0.50	0.38	0.56	0.45	0.54
	0.75	**0.63**	0.79	0.67	0.78
M	0.25	0.22	0.30	0.24	0.28
	0.50	0.40	0.54	**0.52**	0.56
	0.75	0.67	0.79	**0.76**	0.80
H	0.25	0.16	0.25	0.20	**0.22**
	0.50	0.36	**0.44**	0.50	0.52
	0.75	0.55	**0.69**	0.72	0.79

Table 5.8: Predictive performance, Joe FNAC

Copula Family	τ	$n = 10$				$n = 25$			
		$\hat{\theta}_1$	τ_1	$\hat{\theta}_2$	τ_2	$\hat{\theta}_1$	τ_1	$\hat{\theta}_2$	τ_2
Clayton	L	1.24	0.30	0.77	0.22	0.66	0.23	0.47	0.17
	M	3.43	0.56	2.36	0.47	2.37	0.52	1.79	0.45
	H	5.32	0.69	3.95	0.62	4.91	0.69	3.76	0.63
Gumbel	L	1.58	0.31	1.42	0.24	1.36	0.24	1.23	0.17
	M	2.58	0.56	2.21	0.49	2.19	0.52	1.93	0.46
	H	3.85	0.70	3.10	0.64	3.49	0.69	2.81	0.62
Frank	L	2.84	0.25	1.43	0.13	2.00	0.20	1.26	0.13
	M	7.32	0.53	5.52	0.43	5.97	0.50	5.20	0.45
	H	13.99	0.71	10.83	0.65	11.59	0.69	9.08	0.63
Joe	L	1.98	0.28	1.74	0.22	1.55	0.21	1.40	0.16
	M	3.63	0.53	3.10	0.46	2.98	0.49	2.57	0.44
	H	6.06	0.68	4.40	0.57	6.08	0.71	4.59	0.63

Table 5.9: Estimated parameters and corresponding Kendall's τ values from simulated data with varying sample sizes, correlation levels, and FNAC types.

5.5 Applications

In this section, two examples based on the same datasets presented in Section 3.6 are considered to illustrate the application of the proposed method. The first dataset is used for the trivariate case, while the second is applied to the four-variate case.

5.5.1 Typical Meteorological Year Data

In this example, the proposed method in Section 5.2 is illustrated using the Typical Meteorological Year (TMY), as in Example 3.6.1 with the same event of interest. The random quantities are air temperature (X), relative humidity (Y) and wind speed (Z) and the apparent temperature is calculated as $AT = T_a + 0.33e - 0.70ws - 4$.

The variables are ordered according to their correlation strength, with the strongest correlations determining the order. As previously explained in Section 2.5, a key condition of the FNAC model is the restriction on parameters, where the first parameter must be greater than or equal to the second. These parameters represent the level of dependence between the variables. As shown in Figure 3.9, Example 3.6.1 shows that the air temperature and relative humidity have the highest correlation, while wind speed and humidity have the lowest correlation. In the FNAC model, these correlations determine the order of the variables.

The NPI lower and upper probabilities for this event of interest are presented in Table 5.11 and Figure 5.6. Table 5.11 clearly show that imprecision decreases as the correlation among variables increases, regardless of the FNAC type. This can be attributed to two main factors related to the event $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$, which is explained through the probabilities h_{ijk}, fundamental for inference. For high positive correlation, and with the event defined as the sum, the lower and upper probabilities in Equations (3.2.9) and (3.2.10) tend to include additional $h_{ijk}(\hat{\theta})$. These additional terms generally have large values for most t under positive correlation. Conversely, for weak correlation, the h_{ijk} probabilities where i, j, k are close are not as large as in the high positive correlation case.

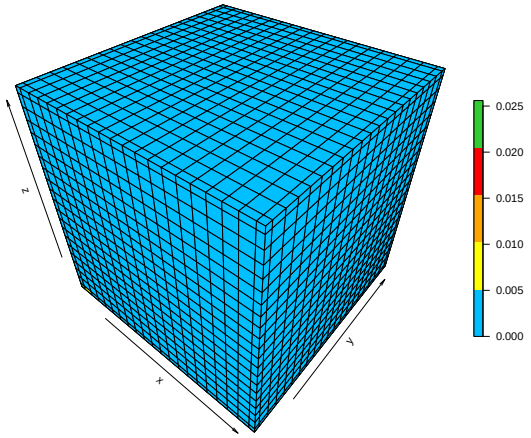
Copula family	Clayton		Gumbel		Frank		Joe	
Pairs	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$
(x, y)	0.81	8.34	0.89	9.12	0.89	35	0.87	13.57
$(.; z)$	0.37	1.18	0.30	1.43	0.33	3.25	0.19	1.42

Table 5.10: Estimated parameters and corresponding Kendall's τ values using different FNAC types.

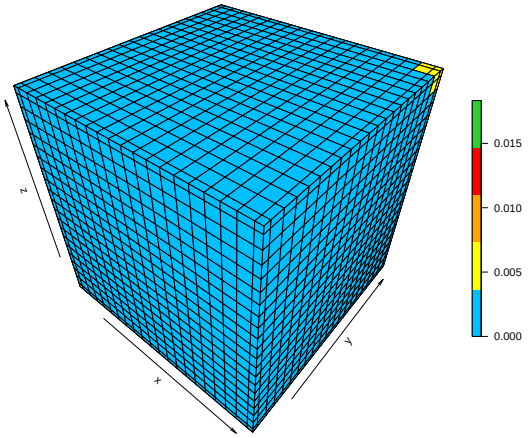
	Clayton		Gumbel		Frank		Joe	
t	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
-6.3570	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
-4.3570	0.9970	1.0000	0.9835	1.0000	0.9877	1.0000	0.9862	1.0000
-2.3570	0.9809	1.0000	0.9623	1.0000	0.9645	1.0000	0.9618	1.0000
-0.3570	0.9524	0.9986	0.9524	0.9986	0.9527	0.9982	0.9529	0.9973
1.6430	0.9474	0.9961	0.9390	0.9878	0.9387	0.9905	0.9270	0.9848
3.6430	0.8490	0.8943	0.8182	0.8541	0.8285	0.8582	0.7959	0.8287
5.6430	0.6395	0.6436	0.6617	0.6623	0.6583	0.6549	0.6528	0.6553
7.6430	0.2390	0.2640	0.2748	0.2945	0.2633	0.2835	0.2946	0.3113
9.6430	0.0148	0.0807	0.0283	0.0720	0.0227	0.0758	0.0442	0.0839
11.6430	0.0003	0.0441	0.0022	0.0288	0.0018	0.0419	0.0045	0.0298
13.6430	0.0000	0.0152	0.0000	0.0258	0.0000	0.0271	0.0000	0.0278
15.6430	0.0000	0.0022	0.0000	0.0038	0.0000	0.0038	0.0000	0.0059

Table 5.11: The NPI lower and upper probabilities of the event $AT_{n+1} > t$ using different FNAC types at selected values of t.

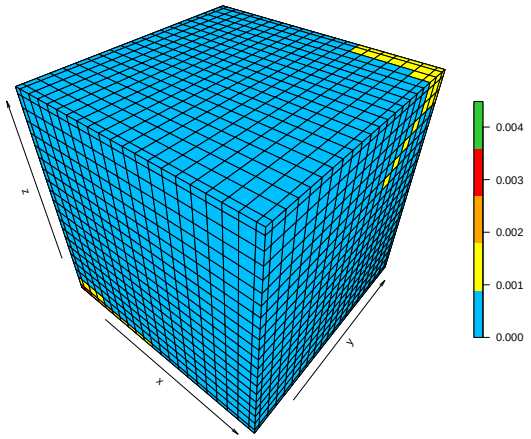
Figure 5.5 shows the probabilities $h_{ijk}(\hat{\boldsymbol{\theta}})$ for each FNAC model. As explained in Example 5.2.1, the probabilities indicate large values when i, j, k close to each other and symmetric results around these values. For a positive correlation, these large values are included in the lower and upper probabilities. Different FNAC models lead to different NPI lower and upper probabilities, as presented in Table 5.11 for selected values of t.



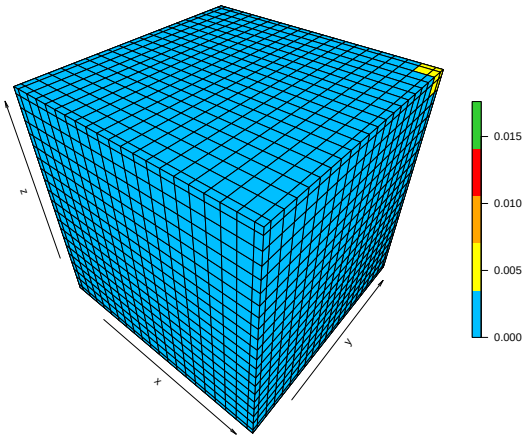
(a) Clayton FNAC



(b) Gumbel FNAC

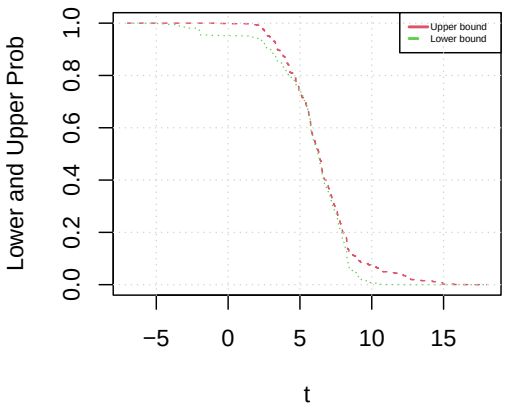


(c) Frank FNAC

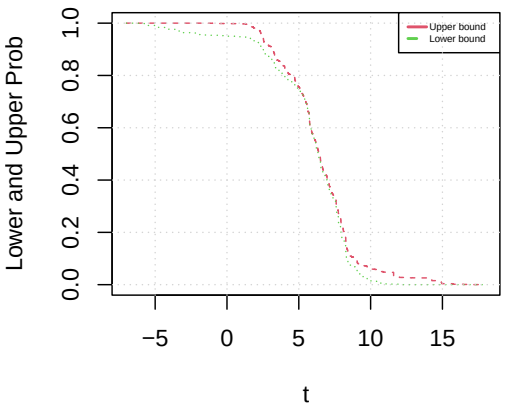


(d) Joe FNAC

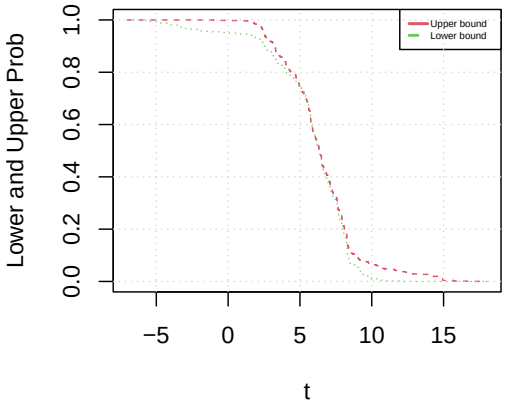
Figure 5.5: The h_{ijk} probabilities, Example 5.5.1



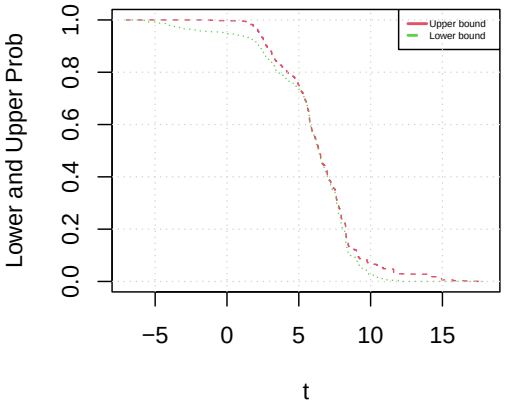
(a) Clayton FNAC



(b) Gumbel FNAC



(c) Frank FNAC



(d) Joe FNAC

Figure 5.6: The NPI lower and upper probabilities of the event $AT_{n+1} > t$ using different FNAC types.

Copula family	Clayton		Gumbel		Frank		Joe	
Pairs	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$
(x, y)	0.50	2.02	0.65	2.84	0.62	8.56	0.61	3.93
$(.; z)$	0.49	1.94	0.65	2.83	0.60	7.81	0.59	3.75
$(.; w)$	0.36	1.13	0.52	2.08	0.48	5.38	0.49	2.79

Table 5.12: Estimated parameters and corresponding Kendall's τ values using different FNAC types.

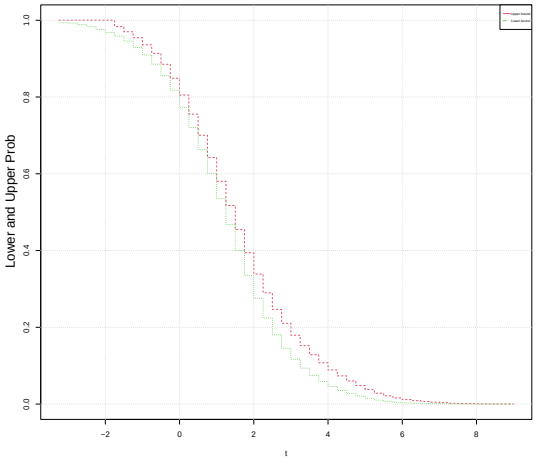
5.5.2 Weekly return data

Consider the same dataset and event of interest for the portfolio returns (PR) as in Example 3.6.2. The portfolio consist of four stocks denoted by X, Y,Z and W. The portfolio return formula is defined as $PR = w_1X + w_2Y + w_3Z + w_4W$, where w_i are the weights with equally weighted 0.25. Assume that one is interested in the next portfolio return PR_{n+1} that exceed a value t, i.e $PR_{n+1} > t$. The generalized form of the proposed method from Section 5.2 is applied to the case $d = 4$, where NPI is combined with a four-variate fully nested Archimedean copula. Table 5.13 and Figure 5.7 present the NPI lower and upper probabilities for the event of interest $PR_{n+1} > t$ in the context of a four-variate case. The results are obtained using the same parametric FNAC and estimation method as in Example 5.5.1, but applied to the four-variate case. The parameters and their corresponding τ values are estimated using the pseudo maximum likelihood estimation method and presented in Table 5.12, which outlines the dependence structure used in this method.

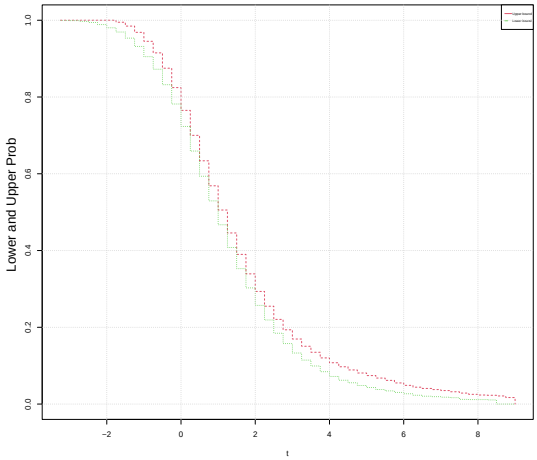
The differences are not easily seen in Figure 5.7, as the results appear similar across the parametric FNAC models. However, Table 5.13 makes these differences more evident, as the copulas yield different probability values $h_{ijkl}(\hat{\boldsymbol{\theta}})$. Although these probabilities are challenging to visualize directly, they are fundamental to the inference process, as they are used to compute the NPI lower and upper probabilities. These lower and upper probabilities are directly affected by the values of $h_{ijkl}(\hat{\boldsymbol{\theta}})$ due to the event of interest involving the sum, this behavior is explained by the effect of strong positive correlation between the random variables discussed in Section 5.2.

	Clayton		Gumbel		Frank		Joe	
t	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
-3.25	0.9935	1.0000	0.9992	1.0000	0.9999	1.0000	1.0000	1.0000
-2.75	0.9885	1.0000	0.9971	1.0000	0.9993	1.0000	0.9998	1.0000
-2.25	0.9763	1.0000	0.9886	1.0000	0.9942	1.0000	0.9982	1.0000
-1.75	0.9587	0.9834	0.9693	0.9949	0.9751	0.9981	0.9891	0.9996
-1.25	0.9292	0.9544	0.9318	0.9685	0.9305	0.9730	0.9572	0.9884
-0.75	0.8856	0.9139	0.8725	0.9150	0.8631	0.9117	0.8867	0.9413
-0.25	0.8179	0.8489	0.7818	0.8245	0.7718	0.8141	0.7736	0.8338
0.25	0.7207	0.7553	0.6592	0.7001	0.6599	0.6969	0.6332	0.6839
0.75	0.6011	0.6421	0.5293	0.5686	0.5467	0.5824	0.4988	0.5400
1.25	0.4676	0.5174	0.4079	0.4458	0.4372	0.4739	0.3844	0.4199
1.75	0.3352	0.3942	0.3026	0.3392	0.3339	0.3739	0.2896	0.3220
2.25	0.2244	0.2897	0.2190	0.2550	0.2441	0.2891	0.2149	0.2457

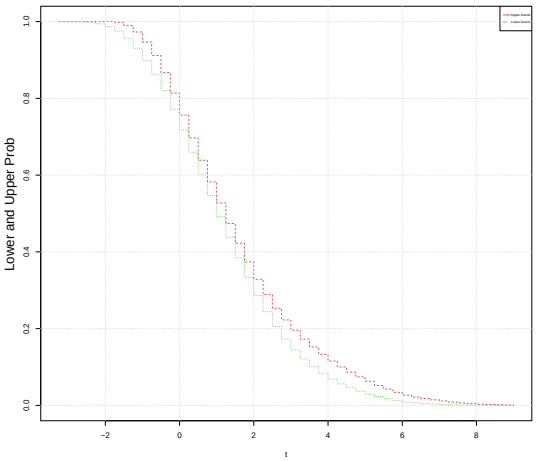
Table 5.13: The NPI lower and upper probabilities of the event $PR_{n+1} > t$ using different types of FNAC at selected values of t.



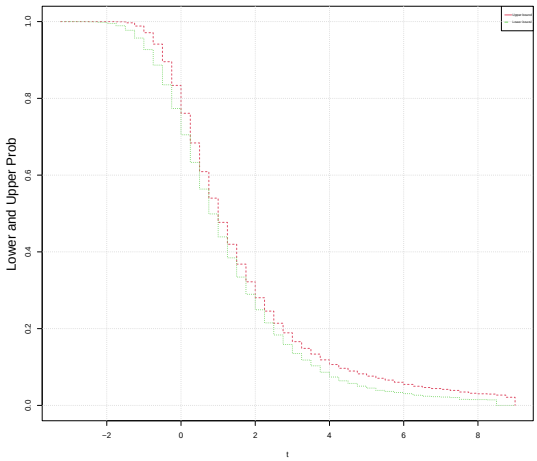
(a) Clayton FNAC



(b) Gumbel FNAC



(c) Frank FNAC



(d) Joe FNAC

Figure 5.7: The NPI lower and upper probabilities of the event $PR_{n+1} > t$ using different FNAC types.

5.6 Comparison study

This section presents a comparison study of all the methods discussed in the thesis, based on simulation studies that evaluate the performance of the proposed NPI methods, using a method similar to the one described in Section 3.5 to assess the predictive performance of the approaches. In this section, N datasets of size $n+1$ are generated. The first n observations from each dataset are used to apply the proposed methods, while the last observations is used to evaluate the predictive performance. To reduce computational time, $N = 100$ simulated datasets are generated with a sample size of $n = 10, 25$. For the simulation, a trivariate Gaussian distribution with mean vector zero and a covariance matrix as:

$$\begin{pmatrix} 1 & 0.9 & 0.5 \\ 0.9 & 1 & 0.15 \\ 0.5 & 0.15 & 1 \end{pmatrix}$$

For the parametric methods, four common copulas—Clayton, Frank, Gumbel and Joe are used. The first method applies classical copulas with a single parameter. The second method uses vine copulas, which include Clayton, Frank, Gumbel and Joe copulas within the vine structure pairwise. The fourth method applies the FNAC dependence model to each of these copulas. The fourth method is a nonparametric approach that uses a kernel-based copula, along with least squares cross-validation (LSCV) and normal reference rule-of-thumb bandwidth selection. For all the parametric methods, the parameters are estimated using the pseudo maximum likelihood estimation method.

The averages of the estimated parameters and their corresponding Kendall's τ values for each parametric method are as follows. For the first method when using classical copula, when $n = 10$: Clayton ($\hat{\theta}_C = 1.23$, $\tau = 0.38$), Gumbel ($\hat{\theta}_G = 1.70$, $\tau = 0.41$), Frank ($\hat{\theta}_F = 3.74$, $\tau = 0.37$) and Joe ($\hat{\theta}_J = 2.01$, $\tau = 0.36$). For $n = 25$: Clayton ($\hat{\theta}_C = 0.94$, $\tau = 0.32$), Gumbel ($\hat{\theta}_G = 1.53$, $\tau = 0.35$), Frank ($\hat{\theta}_F = 3.34$, $\tau = 0.34$) and Joe ($\hat{\theta}_J = 1.73$, $\tau = 0.29$). Although the simulated data are generated with varying levels of pairwise dependence $0.9, 0.5, 0.15$, the Kendall's τ values obtained from the assumed copula model fall within a range 0.29 to 0.41. This is primarily due to the use of a single-parameter copula, which applies the same level of dependence across all variable pairs. A single-parameter copula, such as the Clayton, Gumbel, Frank or Joe copula, assumes

that all pairs of variables share the same dependence structure, controlled by a single parameter θ.

In the second method, when using the vine copula, the parameters $\hat{\theta_1}$ are greater than $\hat{\theta_2}$ with their corresponding Kendall's τ values. This is due to the correlation of the simulated data; the first pair indicates a highly correlated value greater than the correlation of the second pair. The third pair indicates values weaker than the first two pairs and this is consistent with the vine structure. The results are presented in Table 5.14.

In the third method, when using the FNAC structure, the parameters $\hat{\theta_1} > \hat{\theta_2}$ hold for all the FNAC types. This is expected where the correlation of the first pair in the simulated data is higher than the correlations of the other pairs. This aligns with the FNAC parameters' condition. The results are presented in Table 5.15.

In the nonparametric method, when using the normal reference rule-of-thumb and LSCV bandwidth selection, the average of the bandwidth values when $n = 10$ and using the normal reference rule-of-thumb is identical $\boldsymbol{b} = 0.209$ while using the LSCV gives $b_X = 0.064$, $b_Y = 0.183$ and $b_Z = 0.299$. When $n = 25$ and using the normal reference rule-of-thumb, it is identical $\boldsymbol{b} = 0.156$ while using the LSCV gives $b_X = 0.031$, $b_Y = 0.105$ and $b_Z = 0.187$ shows that when $n = 25$ the results decrease compared to when $n = 10$.

As discussed in Section 2.5, the inverse values of the lower and upper survival functions of T_{n+1}, $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$, for a value $q \in (0, 1)$ as defined in Equations (3.5.7) and (3.5.8) lead to the two inequalities p_1 and p_2 presented in Equations (3.5.9) and (3.5.10) for testing performance. The same quantile values $q = 0.25, 0.50, 0.75$ are applied to assess the performance of the four methods and for good performance $p_1 < q < p_2$ must be satisfied.

The results of the predictive performance are presented in Tables 5.16-5.19. The results in general show good performance for the parametric copulas methods. The results of the predictive performance for the first parametric method when using a classical copula with a single parameter are presented in Table 5.16. This table show a good performance, with most results satisfying the condition $p_1 \leq q \leq p_2$, except for a few cases when $n = 10$ and $n = 25$, as highlighted in Table 5.16. These occur when the value of q falls outside the interval $[p_1, p_2]$.

	$n = 10$						$n = 25$					
Copula	τ_1	$\hat{\theta}_1$	τ_2	$\hat{\theta}_2$	τ_3	$\hat{\theta}_3$	τ_1	$\hat{\theta}_1$	τ_2	$\hat{\theta}_2$	τ_3	$\hat{\theta}_3$
Clayton	0.563	3.648	0.355	1.426	0.023	0.055	0.507	2.192	0.322	1.022	0.001	0.003
Gumbel	0.620	3.247	0.387	1.850	0.015	1.018	0.578	2.466	0.358	1.609	0	1
Frank	0.591	9.671	0.344	4.112	-0.228	-2.597	0.578	7.778	0.350	3.686	-0.279	-2.815
Joe	0.559	4.393	0.339	2.236	0.015	1.030	0.501	3.018	0.298	1.838	0.002	1.003

Table 5.14: Estimated parameters $\theta_1, \theta_2, \theta_3$ and their corresponding Kendall's τ_1, τ_2, τ_3 from simulated data with varying sample sizes, using vine copula types.

The results of the predictive performance for the parametric methods when using vine copulas and FNAC are presented in Tables 5.17 and 5.18. Table 5.17 shows two cases where the value q is not included in the interval $[p_1, p_2]$ when $n = 10$ and the vine structure is Gumbel or Joe copulas for $q = 0.75$. Table 5.18 shows one case $q < p_1$ when the FNAC type is Joe. Both methods show more cases where $q \notin [p_1, p_2]$ when $n = 25$. This is reasonable, as a larger sample size tends to introduce small imprecision, so it is expected that some values may fall outside the interval. The tables also show large imprecision when the sample size is $n = 10$, which tends to decrease as the sample size increases to $n = 25$. Similarly, with the nonparametric method using the bandwidth selection method when $n = 10$, there is one case that $q < p_1$ for $q = 0.75$ and increasing the sample size to $n = 25$ shows that cases where q is outside the interval are more compared to the parametric methods as shown in Table 5.19. This occurs due to the event of interest $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$ and the probabilities h_{ijk} that are calculated using the bandwidth selection method which is fundamental for inference, as explained in Section 2.5. In general, for all these methods, since the event of interest involves the sum $T_{n+1} = X_{n+1} + Y_{n+1} + Z_{n+1} > t$ and in the case of positive correlation, the calculated probabilities h_{ijk} tend to be larger when i, j and k are close to each other. Calculating the lower and upper probabilities for this event of interest tends to include additional of high values for most values of t.

	$n = 10$				$n = 25$			
Copula	τ_1	$\hat{\theta}_1$	τ_2	$\hat{\theta}_2$	τ_1	$\hat{\theta}_1$	τ_2	$\hat{\theta}_2$
Clayton	0.554	3.235	0.262	0.892	0.517	2.273	0.225	0.649
Gumbel	0.619	2.996	0.289	1.512	0.582	2.495	0.239	1.357
Frank	0.594	9.118	0.227	2.439	0.583	7.889	0.227	2.257
Joe	0.558	4.045	0.252	1.757	0.504	3.041	0.196	1.487

Table 5.15: Estimated parameters θ_1, θ_2 and their corresponding Kendall's τ_1, τ_2 from simulated data with varying sample sizes, using FNAC types.

n	q	Clayton		Gumbel		Frank		Joe	
		p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
	0.25	0.20	0.40	0.21	0.47	0.21	0.41	**0.26**	0.47
10	0.50	**0.53**	0.67	**0.60**	0.70	**0.57**	0.70	**0.62**	0.72
	0.75	0.75	0.92	0.75	0.93	0.75	0.94	0.75	0.96
	0.25	0.21	0.28	0.24	0.28	0.21	0.28	**0.26**	0.29
25	0.50	0.36	**0.40**	0.40	**0.47**	0.38	**0.44**	0.41	0.51
	0.75	0.68	0.77	0.71	0.80	0.72	0.83	0.69	0.81

Table 5.16: Predictive performance; classical copulas.

n	q	Clayton		Gumbel		Frank		Joe	
		p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
	0.25	0.18	0.28	0.13	0.29	0.12	0.35	0.13	0.31
10	0.50	0.35	0.54	0.40	0.59	0.38	0.60	0.42	0.62
	0.75	0.74	0.89	**0.77**	0.90	0.71	0.90	**0.78**	0.94
	0.25	0.15	**0.21**	0.19	**0.23**	0.19	0.25	0.21	0.25
25	0.50	0.38	**0.49**	0.49	0.52	0.42	0.51	0.50	0.54
	0.75	0.68	0.77	0.72	0.82	0.66	0.79	0.72	0.82

Table 5.17: Predictive performance; vine copulas.

n	q	Clayton		Gumbel		Frank		Joe	
		p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
	0.25	0.20	0.40	0.23	0.42	0.22	0.40	0.25	0.40
10	0.50	0.44	0.59	0.49	0.64	0.48	0.63	**0.52**	0.65
	0.75	0.71	0.90	0.73	0.91	0.75	0.94	0.71	0.98
	0.25	0.21	0.27	0.22	0.28	0.22	0.27	0.24	0.29
25	0.50	0.35	**0.40**	0.39	**0.46**	0.38	**0.45**	0.41	0.50
	0.75	0.70	0.78	0.72	0.81	0.72	0.83	0.72	0.82

Table 5.18: Predictive performance; FNAC types.

n	q	Normal reference		LSCV	
		p_1	p_2	p_1	p_2
	0.25	0.18	0.33	0.20	0.33
10	0.50	0.39	0.62	0.44	0.70
	0.75	**0.78**	0.92	**0.81**	0.93
	0.25	**0.27**	0.31	**0.27**	0.31
25	0.50	**0.58**	0.62	**0.58**	0.61
	0.75	**0.81**	0.90	**0.81**	0.89

Table 5.19: Predictive performance; nonparametric copulas.

5.7 Concluding remarks

In this chapter, the method of combining NPI for the marginals with a trivariate FNAC to capture the dependence among variables is introduced. FNAC is a nested structure to model the dependence between variables. FNAC is more flexible than classical copula with one parameter due to the nested structure and contains multiple parameters but it is less flexible than vine copula because FNAC is limited to the Archimedean family. This method is introduced for predictive inference about a single future trivariate observation. A multivariate expression ($d > 3$) is constructed in a similar manner. This method is

illustrated via examples and the performance of this method is evaluated through simulations, showing good results in general. Throughout this work, the focus was on assuming an FNAC from the same type of Archimedean family. However, it might be of interest to explore the use of FNAC with different Archimedean copulas in a nested structure. This is left for future work. The performance of the method using FNAC is compared with classical copulas, nonparametric copulas and vine copulas via simulation. The results show that the methods with either FNAC or vine copulas perform well compared to other methods, while the nonparametric method shows relatively weaker performance.

The benefits offered by FNAC become more complex as the dimension increases. This is due to the increased level of nesting, which adds more parameters. In high dimensional cases, FNAC becomes limited in capturing dependence, where the range of dependence that can be modelled is reduced. As the dataset size and dimensionality increase, the proposed method requires more computational resources and processing time.

Chapter 6

Conclusions

This thesis mainly aims to introduce NPI for multivariate data using copulas. To do this, we approach the problem in two ways: semi-parametrically, by assuming a parametric copulas, and nonparametrically, by assuming a nonparametric copulas to model the dependence structure—while using NPI for the marginals. The NPI-based marginals are then combined with the chosen copula (parametric or nonparametric) to form the full multivariate model. Four methods are introduced in this thesis using different types of copulas for the predictive performance of a single future observation. The focus is on parametric copulas with one parameter, nonparametric copulas, vine copulas, and FNAC. These proposed methods using these types of copulas are illustrated their superiority through simulation studies and real data applications. Through simulation studies, the performance of these methods is evaluated and a comparison study of the proposed methods is conducted. The main strength of the multivariate copula model is that it provides flexibility where the dependence structure and the marginals are separated. This benefit allows for choosing different dependence structures with uniform marginals.

In Chapter 3, two methods are introduced for predictive inference: NPI combined with parametric classical copula and NPI combined with nonparametric copulas using kernel-based copula. The first method depends on using a parametric copula with a single parameter. This method performs well regardless of the sample size, copula type and the level of dependence. For further research, it could be of interest to study classical copulas with more than one parameter, as they are better suited to capturing complex dependence structures than single-parameter copulas. One more interesting topic that can be explored

using this method is by studying the imprecise discrete copulas using patchwork techniques [65]. This is left for future research. The second method depends on using kernel-based copulas and the bandwidth selection type. This method is demonstrated using simulated and real data sets. The predictive performance of this method was evaluated and found to be suboptimal. Future research could explore alternative nonparametric copulas to potentially enhance prediction. Also, one could study the smoothed bootstrap of this method [5]. This is left as future research.

In Chapter 4, the method of combining NPI with vine copulas is introduced. The main benefit of the vine copula structure is that it contains several bivariate copulas in a model. This feature allows for the flexibility of specifying each bivariate copula, which allows for capturing different dependencies in a model. A numerical example shows that the probabilities h_{ijk} conditions are not satisfied. The IPFP algorithm is able to solve this problem by adjusting the marginals and preserving the dependence among variables. The proposed method is illustrated through a real data application and simulated data in different scenarios. The performance is evaluated through simulations and the method performs well. For future research, one may study the method of combining NPI with discrete vine copulas, e.g. similar to the ones developed by Panagiotelis *et al.* [87].

In Chapter 5, the method of combining NPI with FNAC is introduced. In this method, the NPI is incorporated with FNAC, capturing the dependence structure. FNAC is suitable for modeling complex high-dimensional dependence but is limited to the Archimedean copula family. The proposed method is illustrated via examples and real-data applications. The performance is evaluated via simulation studies and the results are satisfactory. A comparison study between the four methods introduced in this thesis. The results show that the performance of the method when using either FNAC or vine copulas is better than when using classical copulas, while the nonparametric method did not perform well. An advantage of FNAC is that it allows the use of different Archimedean copulas within a model. It may be interesting to explore this type of FNAC combined with NPI. This is left for future research.

In this thesis, these methods can be applied to any event of interest involving future observations. Integrating NPI with a dependence model for predictive inference in multivariate scenarios provides valuable insights into data relationships. However,

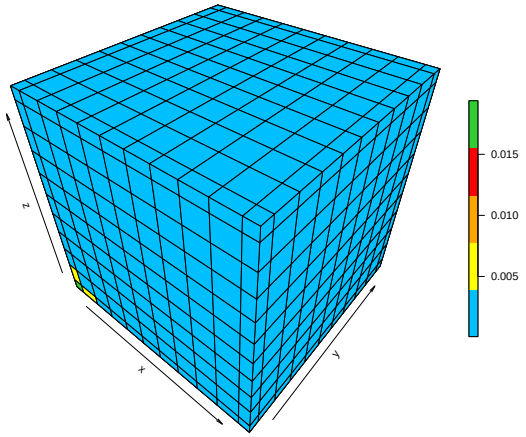
as dimensionality increases, the dependence model grows more complex. Calculating the discretized probabilities becomes increasingly computationally intensive, and overall computation time also rises with larger sample sizes.

As a further potential direction for future research, a new way to combine NPI with multivariate data called the partially nested Archimedean copula (PNAC) is introduced by Joe [59], which is unlike the FNAC. The FNAC structure nests one dimension at a time, requiring at least three variables for modelling dependence. Whereas PNAC requires at least four variables in its structure [59]. For instance, assume there are four random variables, then model each pair by the copula and then couple these two bivariate copulas by another copula. Another interesting approach for combining NPI with multivariate data is to use the General Nested Archimedean copula (GNAC). GNAC is used to model the dependence structure with arbitrary nesting levels, which is left as future work [68].

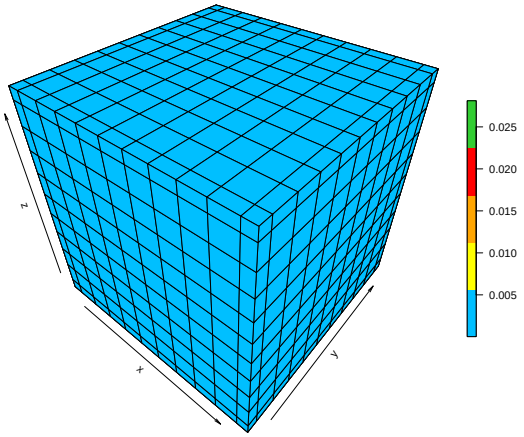
Appendix A

Visualizations of the probabilities h_{ijk}; Classical Copulas

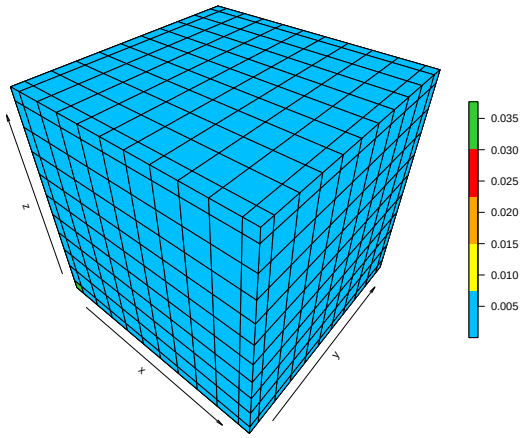
This appendix displays visualizations of the probabilities h_{ijk} under various trivariate copula types. It demonstrates how the estimation method, the strength of the dependence between variables, and the selected copula type all affect the outcomes. These visualizations are part of the simulation study discussed in Section 3.4.



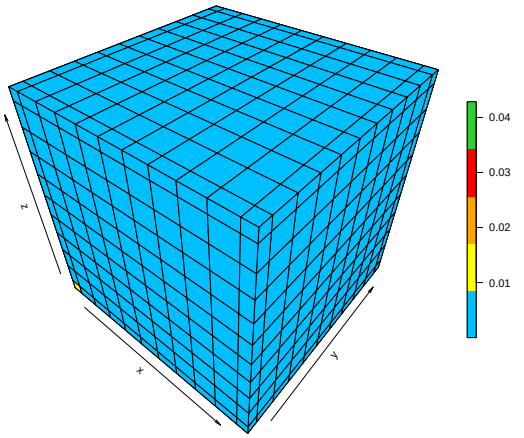
(a) Low, Clayton copula



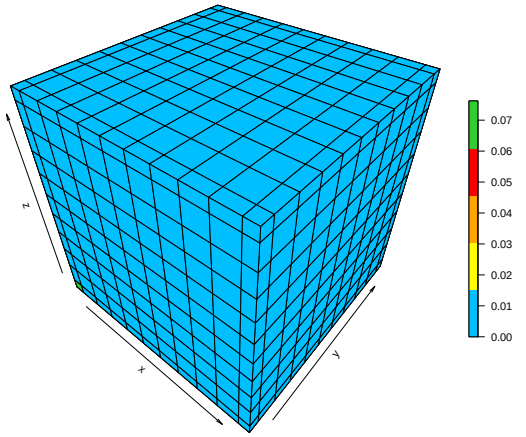
(b) Low, Gumbel copula



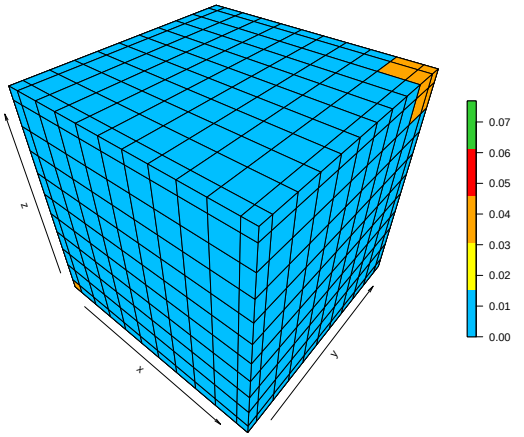
(c) Moderate, Clayton copula



(d) Moderate, Gumbel copula

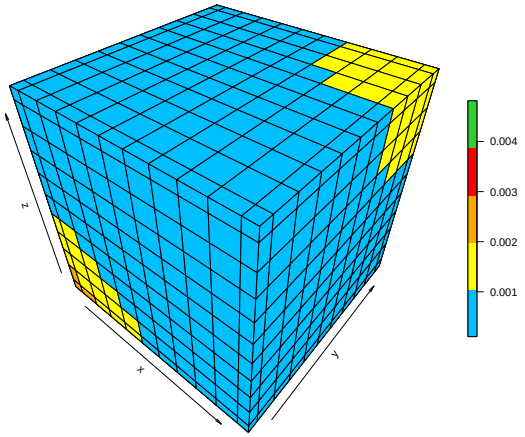


(e) High, Clayton copula

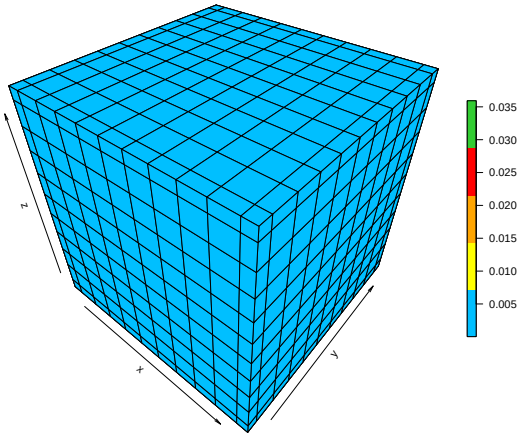


(f) High, Gumbel copula

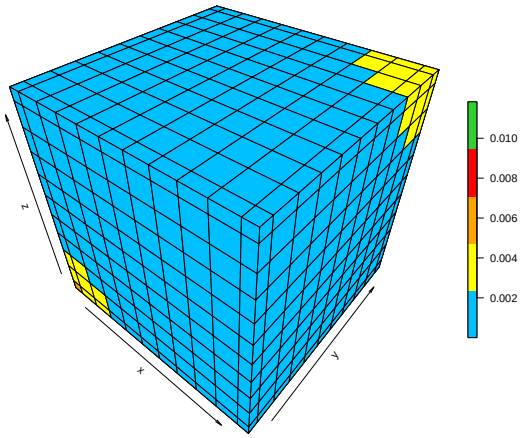
Figure A.1: The h_{ijk} probabilities were obtained from simulated data of size $n = 10$, generated using the Clayton copula and the Gumbel copula with varying correlation coefficients.



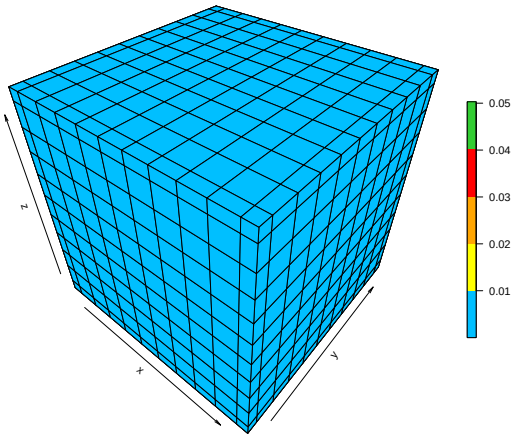
(a) Low, Frank copula



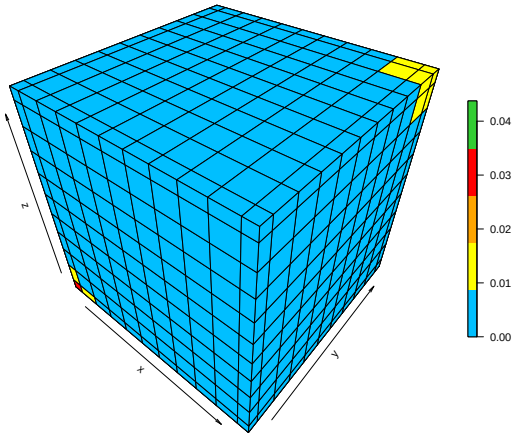
(b) Low, Joe copula



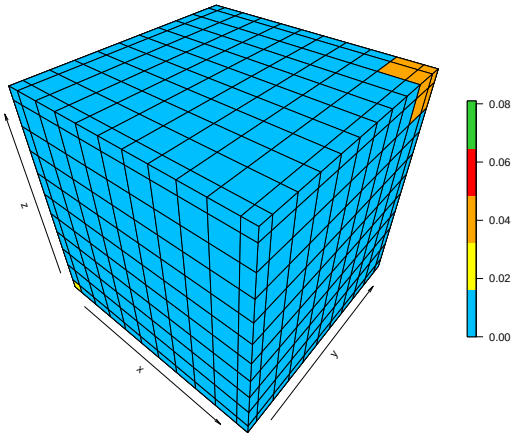
(c) Moderate, Frank copula



(d) Moderate, Joe copula

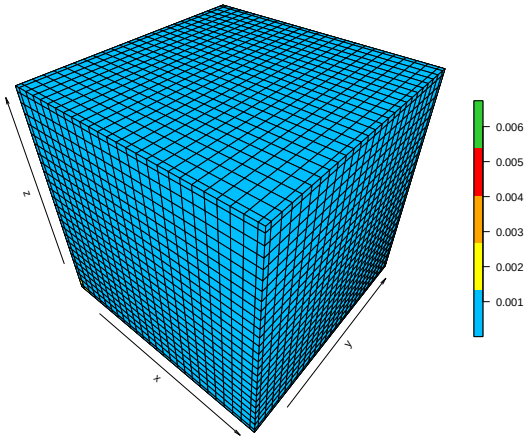


(e) High, Frank copula

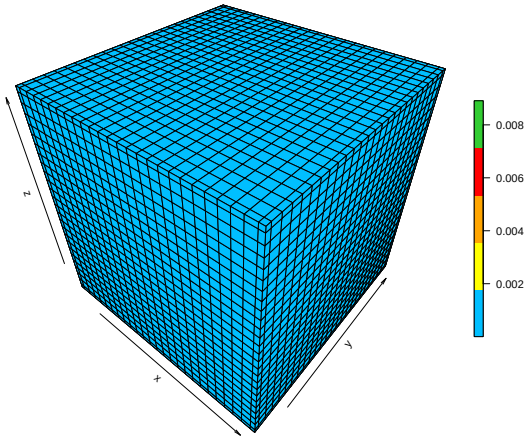


(f) High, Joe copula

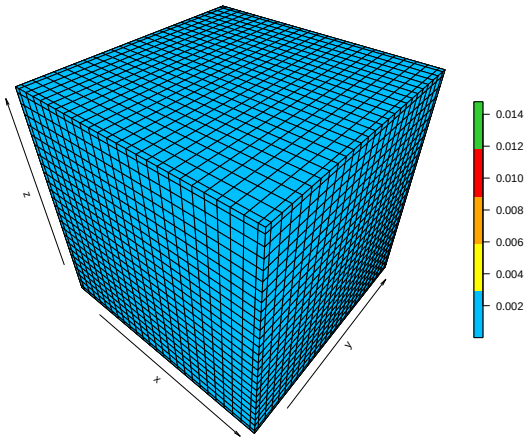
Figure A.2: The h_{ijk} probabilities were obtained from simulated data of size $n = 10$, generated using the Frank copula and the Joe copula with varying correlation coefficients.



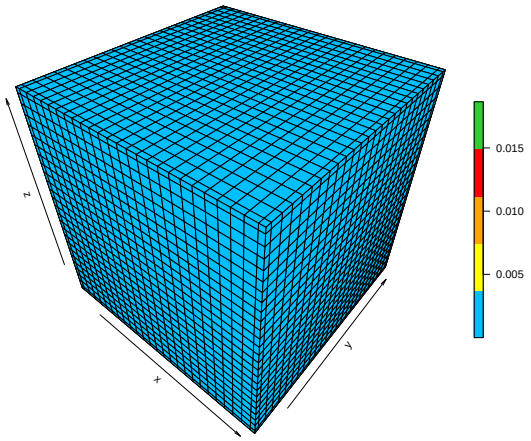
(a) Low, Clayton copula



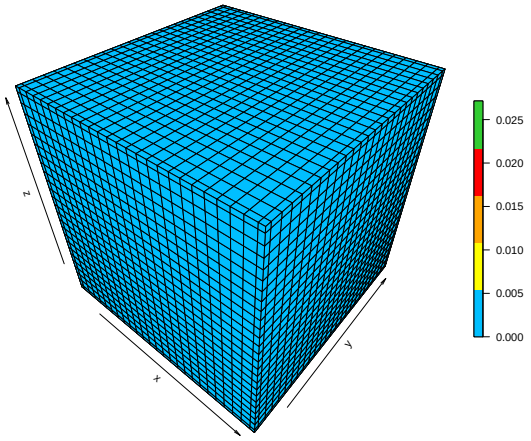
(b) Low, Gumbel copula



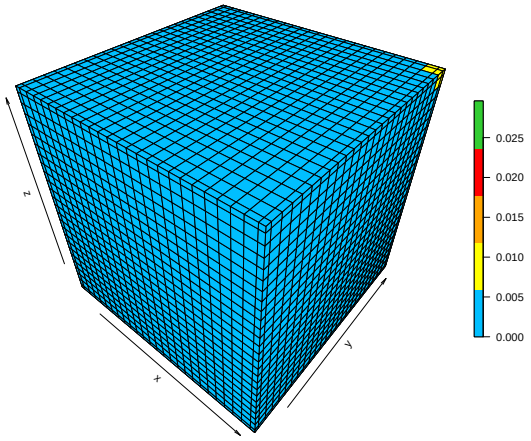
(c) Moderate, Clayton copula



(d) Moderate, Gumbel copula

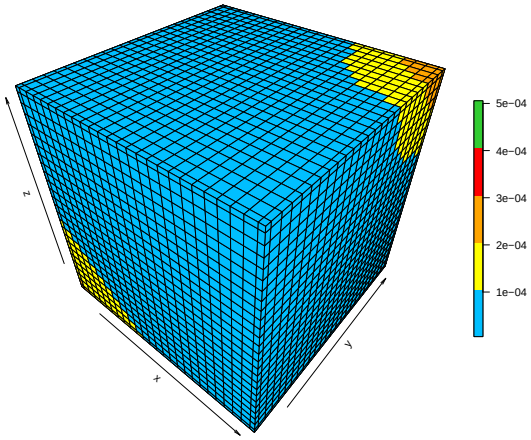


(e) High, Clayton copula

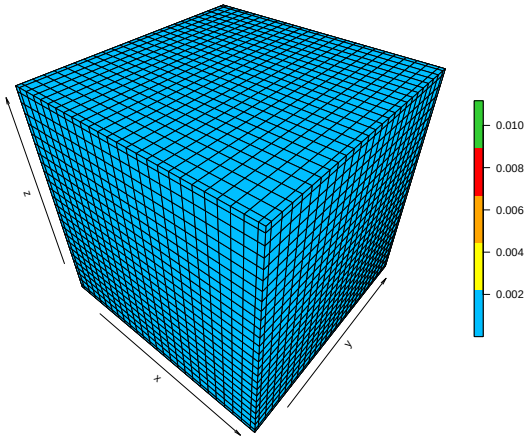


(f) High, Gumbel copula

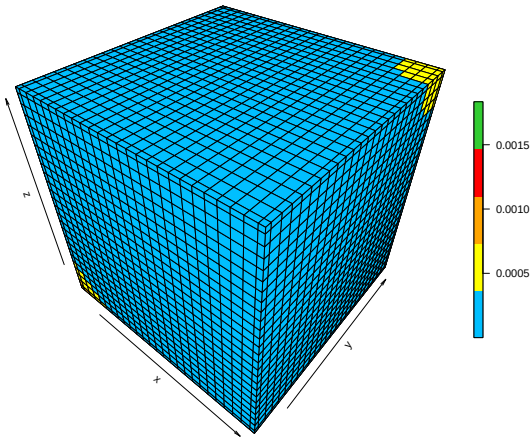
Figure A.3: The h_{ijk} probabilities were obtained from simulated data of size $n = 25$, generated using the Clayton copula and the Gumbel copula with varying correlation coefficients.



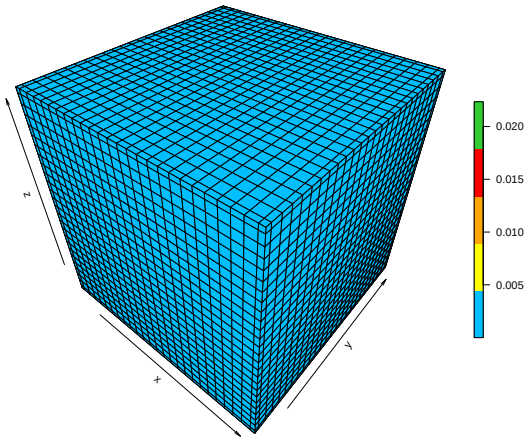
(a) Low, Frank copula



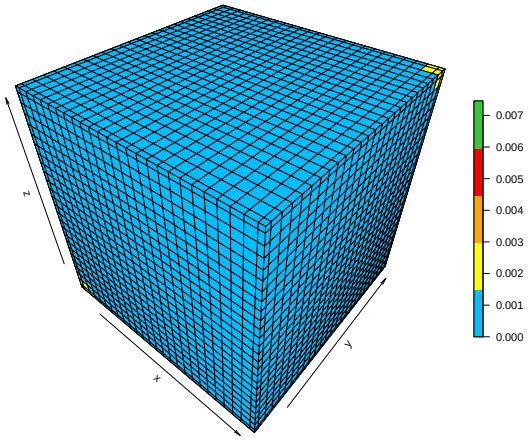
(b) Low, Joe copula



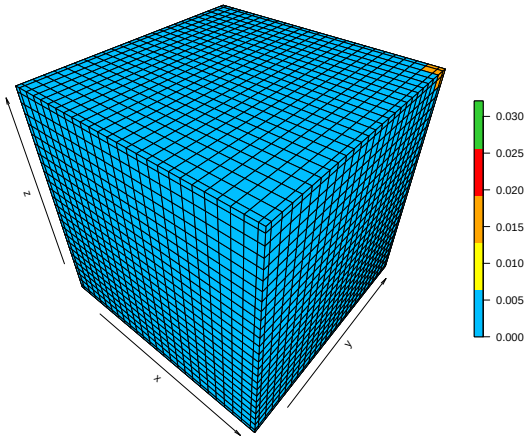
(c) Moderate, Frank copula



(d) Moderate, Joe copula

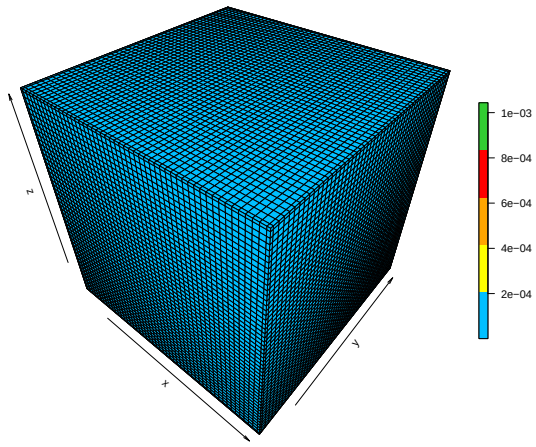


(e) High, Frank copula

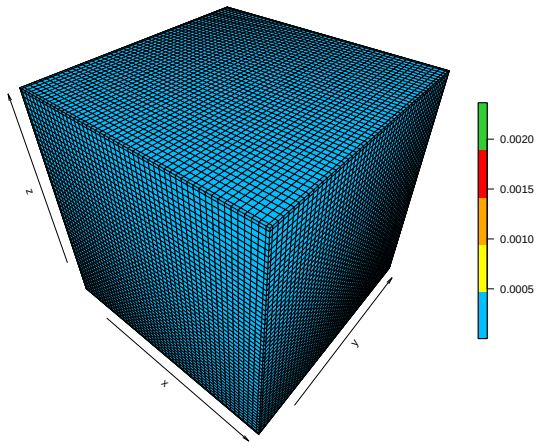


(f) High, Joe copula

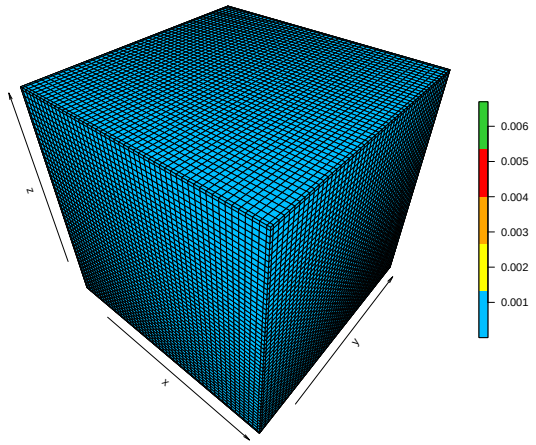
Figure A.4: The h_{ijk} probabilities were obtained from simulated data of size $n = 25$, generated using the Frank copula and the Joe copula with varying correlation coefficients.



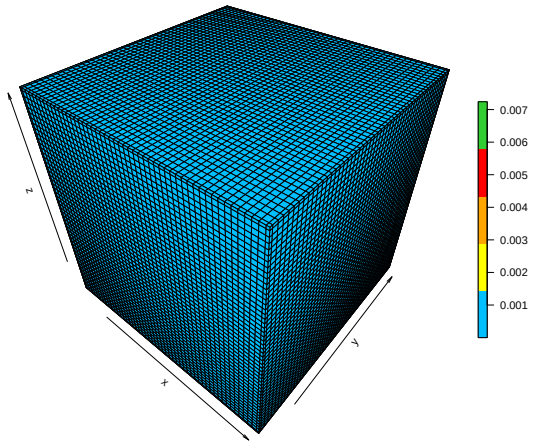
(a) Low, Clayton copula



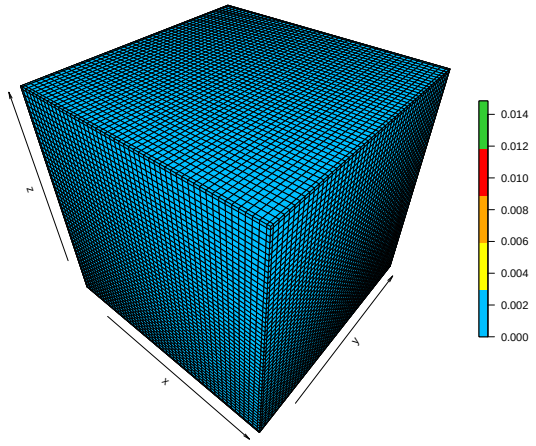
(b) Low, Gumbel copula



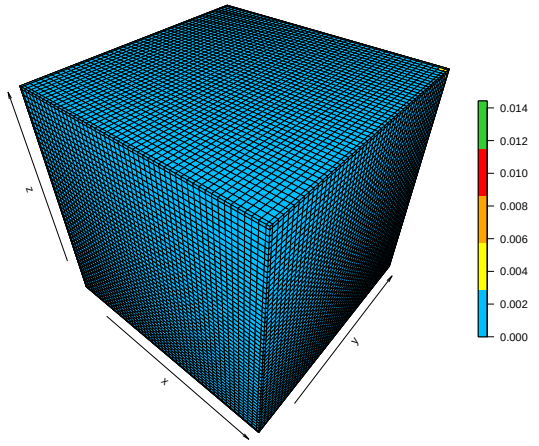
(c) Moderate, Clayton copula



(d) Moderate, Gumbel copula

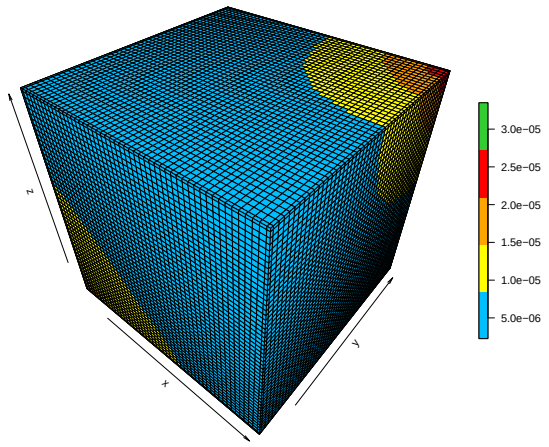


(e) High, Clayton copula

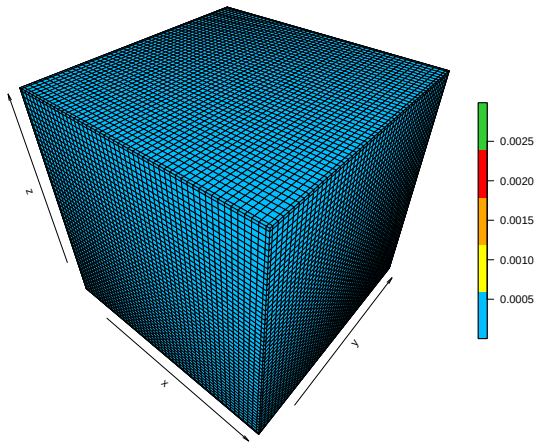


(f) High, Gumbel copula

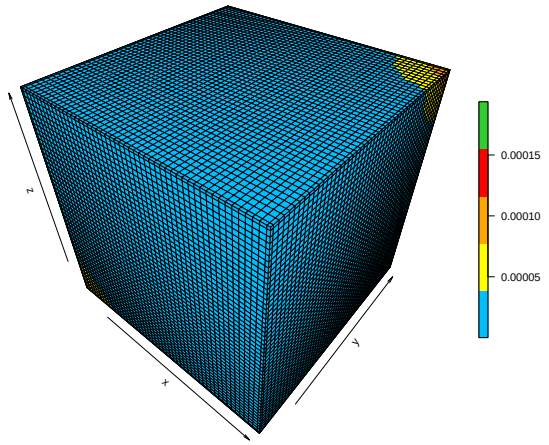
Figure A.5: The h_{ijk} probabilities were obtained from simulated data of size $n = 50$, generated using the Clayton copula and the Gumbel copula with varying correlation coefficients.



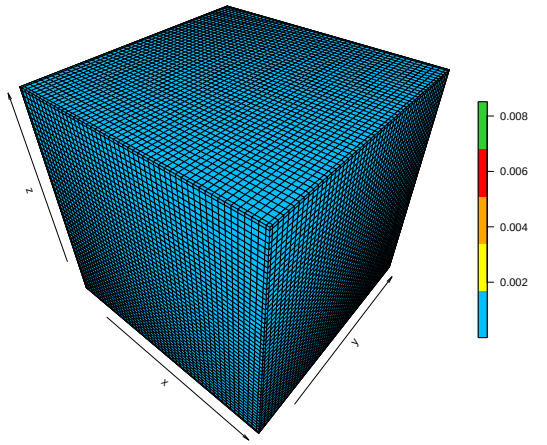
(a) Low, Frank copula



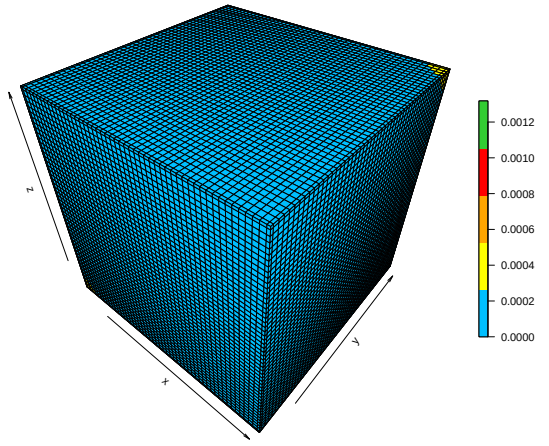
(b) Low, Joe copula



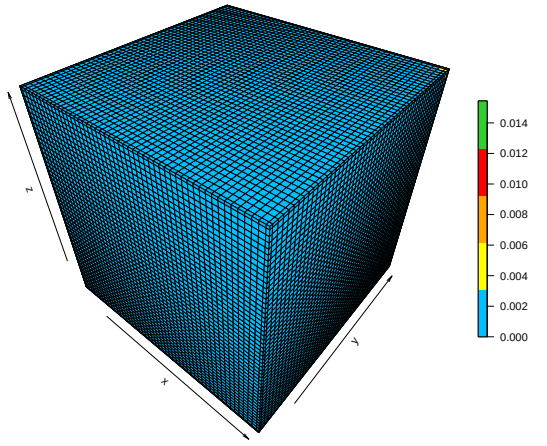
(c) Moderate, Frank copula



(d) Moderate, Joe copula

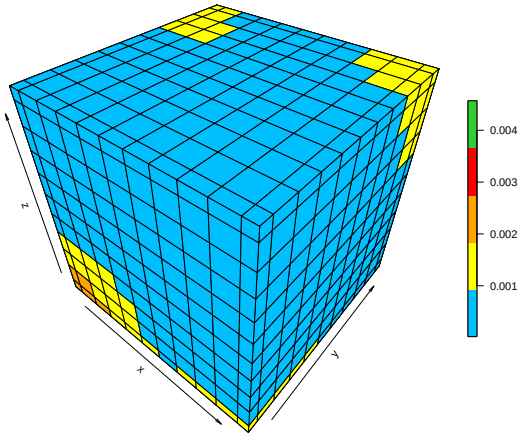


(e) High, Frank copula

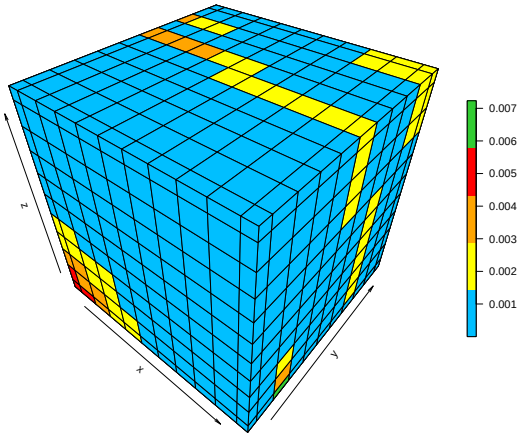


(f) High, Joe copula

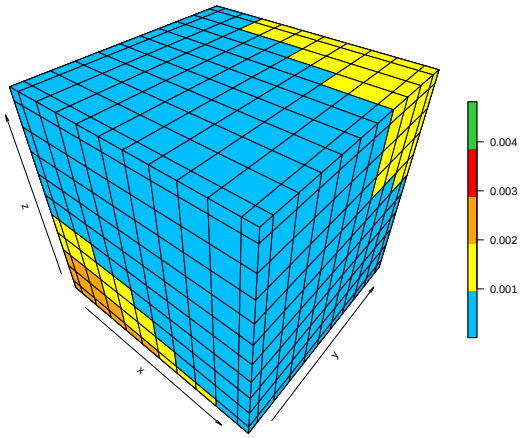
Figure A.6: The h_{ijk} probabilities were obtained from simulated data of size $n = 50$, generated using the Frank copula and the Joe copula with varying correlation coefficients.



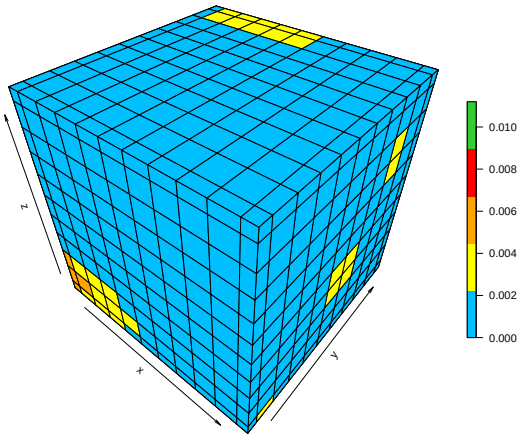
(a) Low, normal reference



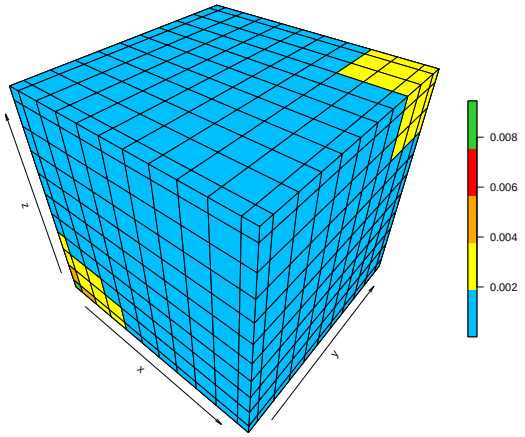
(b) Low, LSCV



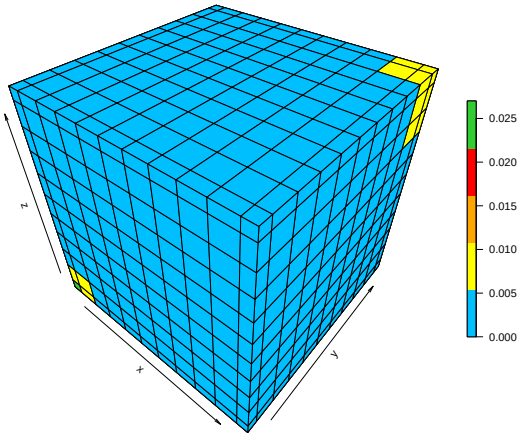
(c) Moderate, normal reference



(d) Moderate, LSCV

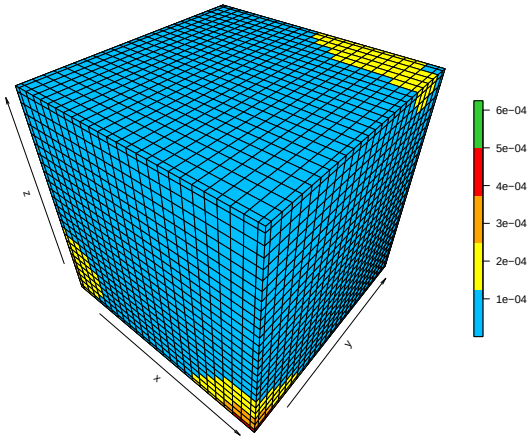


(e) High, normal reference

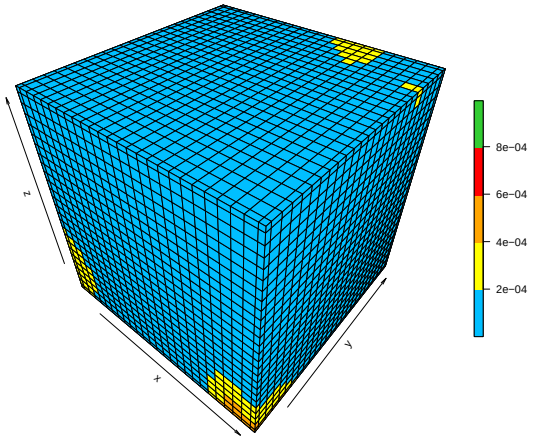


(f) High, LSCV

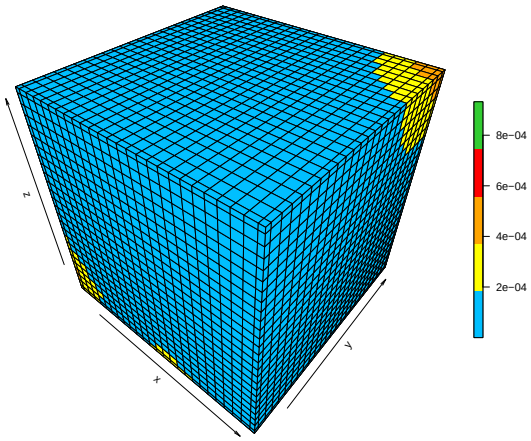
Figure A.7: The h_{ijk} probabilities were obtained from simulated data of size $n = 10$ using different types of bandwidths with varying correlation coefficients.



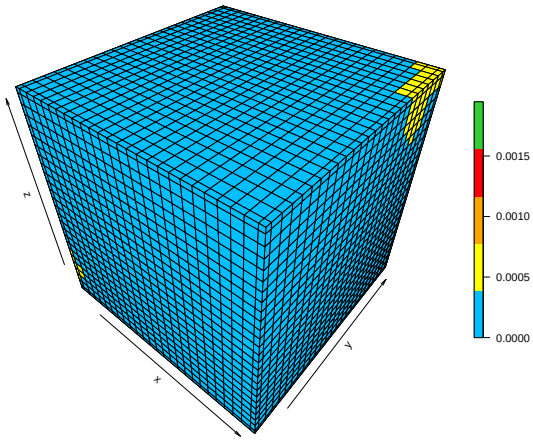
(a) Low, normal reference



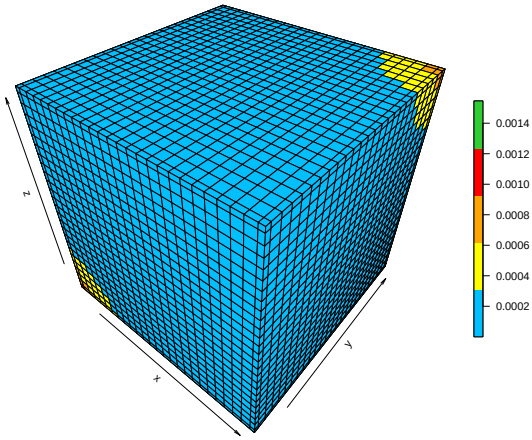
(b) Low, LSCV



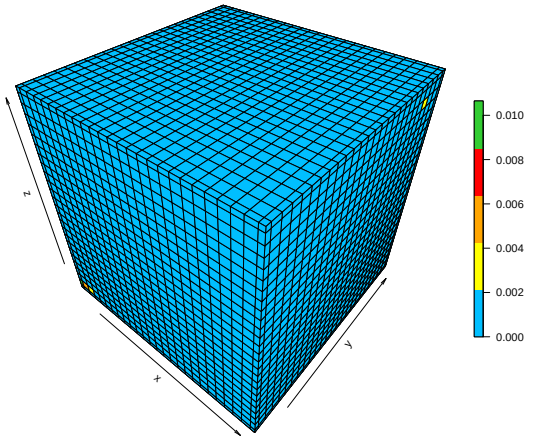
(c) Moderate, normal reference



(d) Moderate, LSCV

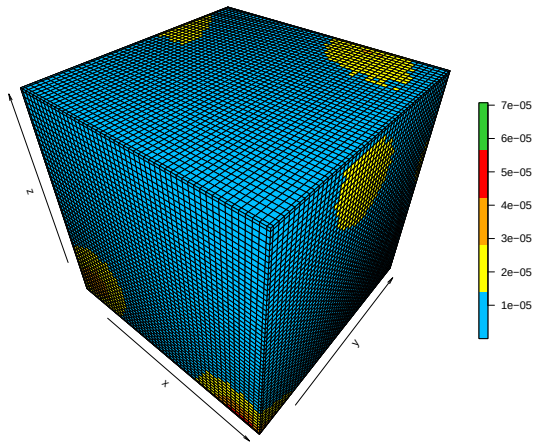


(e) High, normal reference

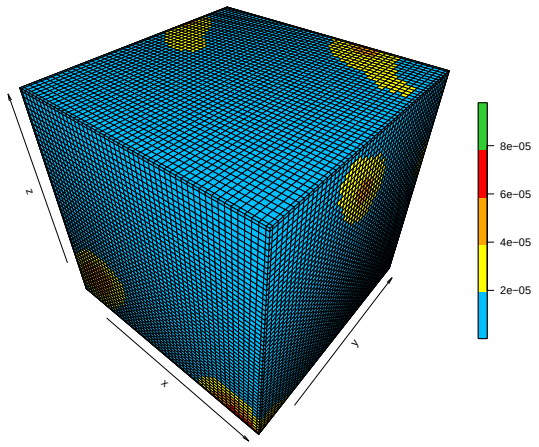


(f) High, LSCV

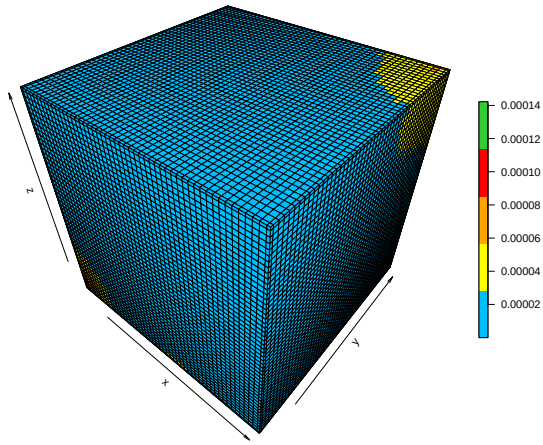
Figure A.8: The h_{ijk} probabilities were obtained from simulated data of size $n = 25$ using different types of bandwidths with varying correlation coefficients.



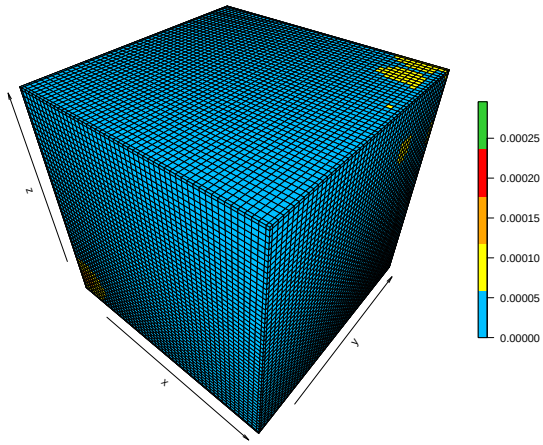
(a) Low, normal reference



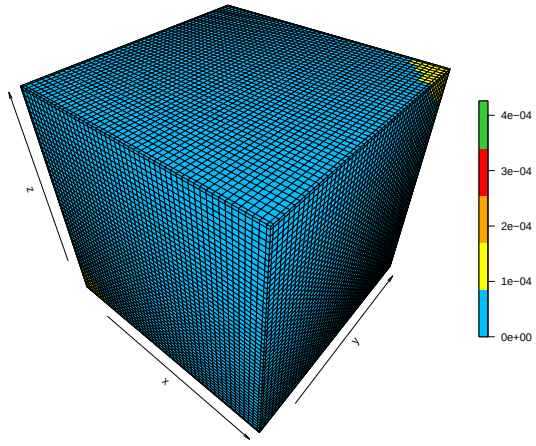
(b) Low, LSCV



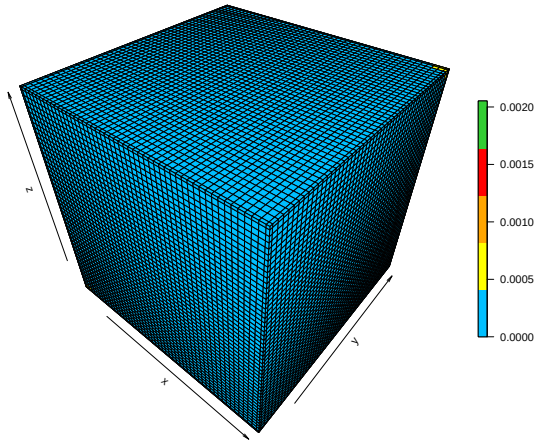
(c) Moderate, normal reference



(d) Moderate, LSCV



(e) High, normal reference



(f) High, LSCV

Figure A.9: The h_{ijk} probabilities were obtained from simulated data of size $n = 50$ using different types of bandwidths with varying correlation coefficients.

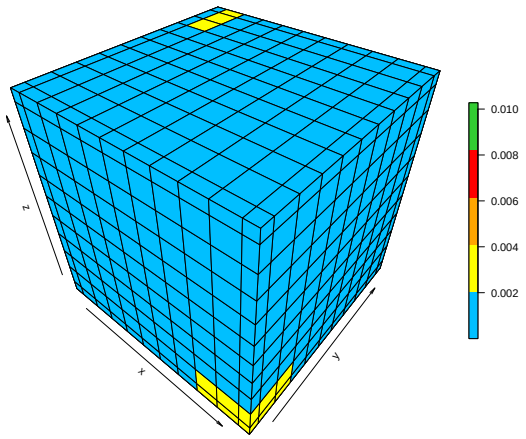
Appendix B

Visualizations of the probabilities h_{ijk}; Vine Copula

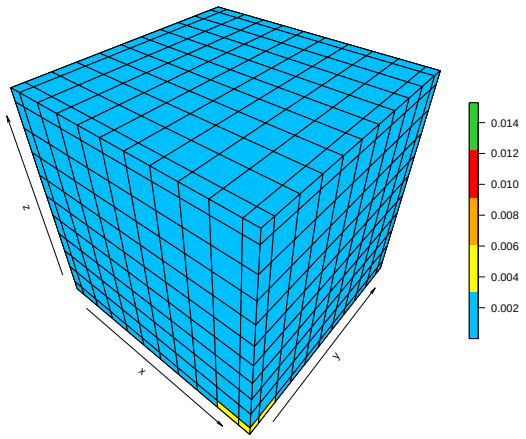
This appendix presents the estimated parameters and visualizations of the probabilities h_{ijk}, obtained using a Gaussian copula based on the scenarios outlined in Section 4.3.

n	Correlation	Pair	Case I		Case II		Case III		Case IV	
			τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$	τ	$\hat{\theta}$
		1st pair	0.28	0.43	0.28	0.43	-0.57	-0.78	0.28	0.43
10	No correlation	2nd pair	-0.57	-0.78	-0.10	-0.15	-0.10	-0.15	-0.57	-0.78
		3rd pair	0.08	0.13	-0.55	-0.76	0.24	0.36	-0.10	-0.15
		1st pair	0.50	0.71	0.50	0.71	-0.19	-0.29	0.50	0.71
10	Moderate	2nd pair	-0.19	-0.29	0.28	0.42	0.28	0.42	-0.19	-0.29
		3rd pair	0.29	0.44	-0.24	-0.37	0.48	0.69	0.28	0.42
		1st pair	0.86	0.97	0.86	0.97	0.64	0.84	0.86	0.97
10	High	2nd pair	0.64	0.84	0.69	0.88	0.69	0.88	0.64	0.84
		3rd pair	0.33	0.50	-0.10	-0.16	0.73	0.91	0.69	0.88
		1st pair	-0.01	-0.01	-0.01	-0.01	-0.37	-0.55	-0.01	-0.01
25	No correlation	2nd pair	-0.37	-0.55	0.09	0.09	0.14	-0.30	-0.37	-0.55
		3rd pair	0.09	0.14	-0.37	-0.55	0.03	0.05	0.09	0.14
		1st pair	0.35	0.53	0.35	0.53	-0.03	-0.05	0.35	0.53
25	Moderate	2nd pair	-0.03	-0.05	0.33	0.50	0.33	0.50	-0.03	-0.05
		3rd pair	0.37	0.54	-0.20	-0.30	0.56	0.38	0.33	0.50
		1st pair	0.67	0.87	0.67	0.87	0.58	0.79	0.67	0.87
25	High	2nd pair	0.58	0.79	0.70	0.89	0.70	0.89	0.58	0.79
		3rd pair	0.46	0.67	0.06	0.09	0.40	0.58	0.70	0.89
		1st pair	0.01	0.02	0.01	0.02	-0.09	-0.14	0.01	0.02
50	No correlation	2nd pair	-0.09	-0.14	0.04	0.06	0.04	0.06	-0.09	-0.14
		3rd pair	0.04	0.06	-0.09	-0.14	0.02	0.03	0.04	0.06
		1st pair	0.61	0.81	0.36	0.53	0.41	0.27	0.36	0.53
50	Moderate	2nd pair	0.73	0.91	0.36	0.54	0.36	0.54	0.27	0.41
		3rd pair	0.50	0.70	0.12	0.19	0.26	0.40	0.36	0.54
		1st pair	0.70	0.89	0.70	0.89	0.69	0.89	0.70	0.89
50	High	2nd pair	0.69	0.89	0.72	0.90	0.72	0.90	0.69	0.89
		3rd pair	0.35	0.53	0.27	0.41	0.32	0.48	0.72	0.90

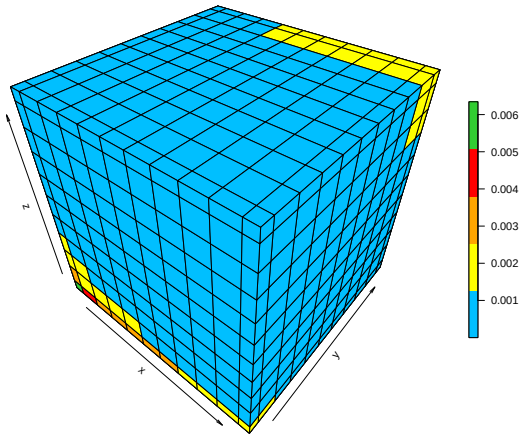
Table B.1: Estimated Kendall's τ and copula parameters $\hat{\theta}$ for different pairs across four cases, correlation levels, and varying sample sizes using Gaussian vine copulas.



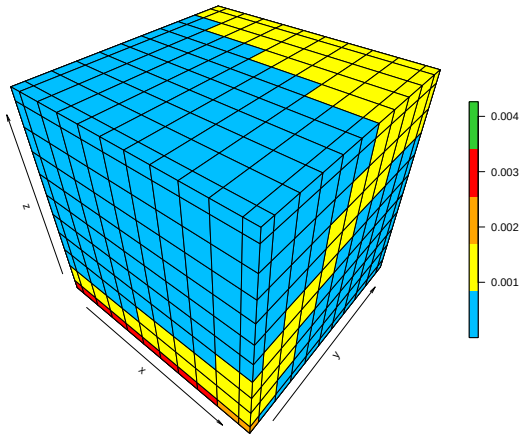
(a) Low, Case I



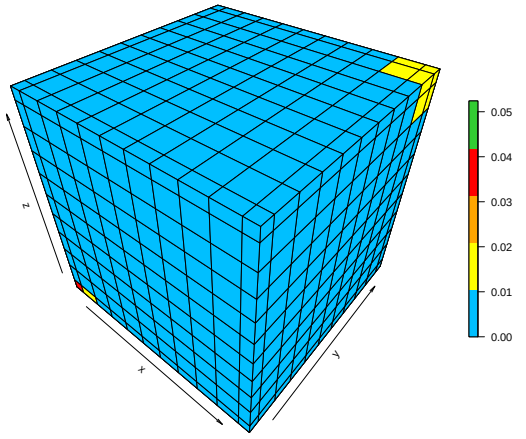
(b) Low, Case II



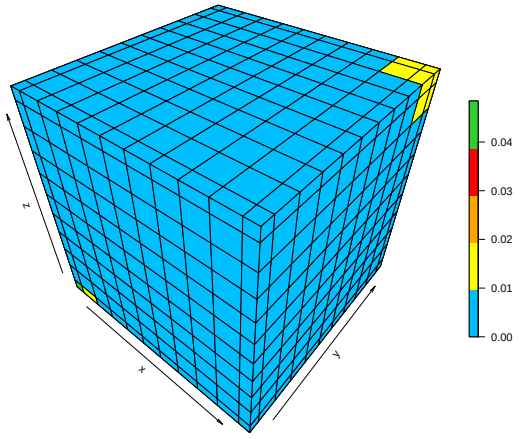
(c) Moderate, Case I



(d) Moderate, Case II

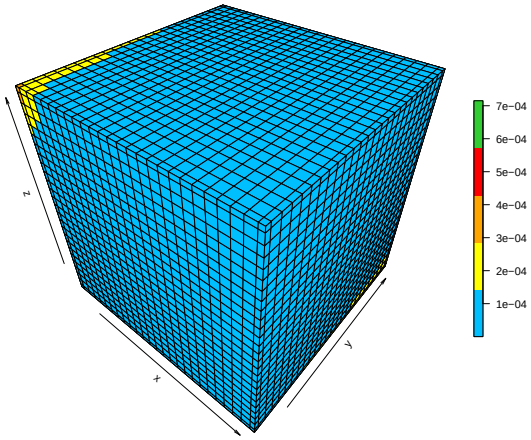


(e) High, Case I

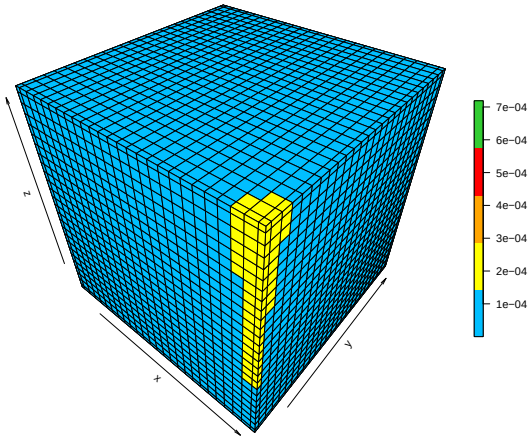


(f) High, Case II

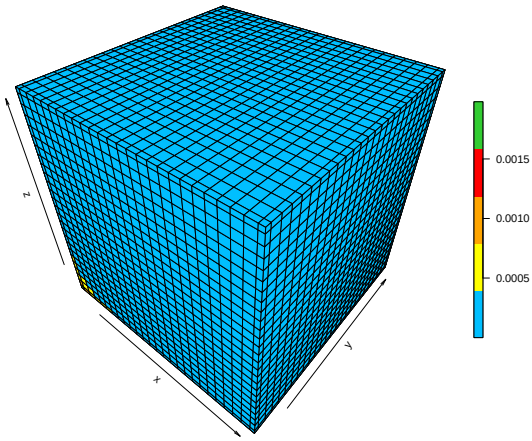
Figure B.1: The h_{ijk} probabilities were obtained from simulated data of size $n = 10$, generated using Gaussian copulas with varying correlation coefficients.



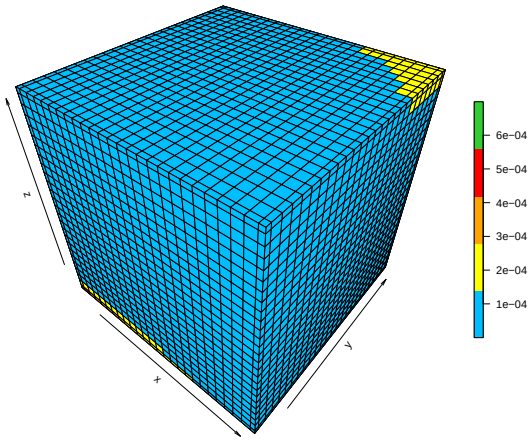
(a) Low, Case I



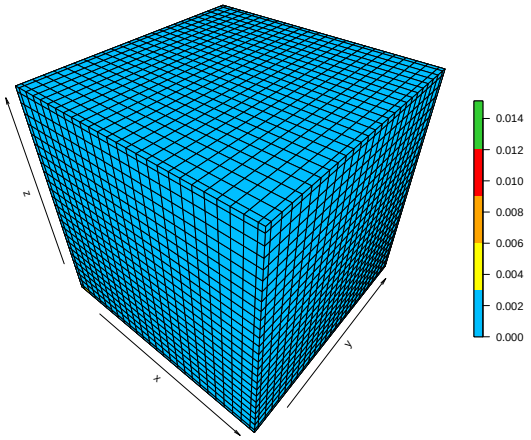
(b) Low, Case II



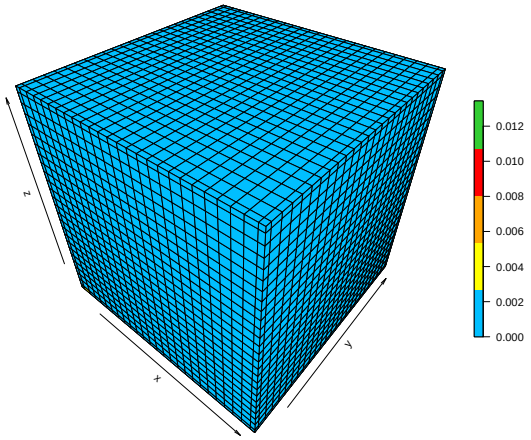
(c) Moderate, Case I



(d) Moderate, Case II

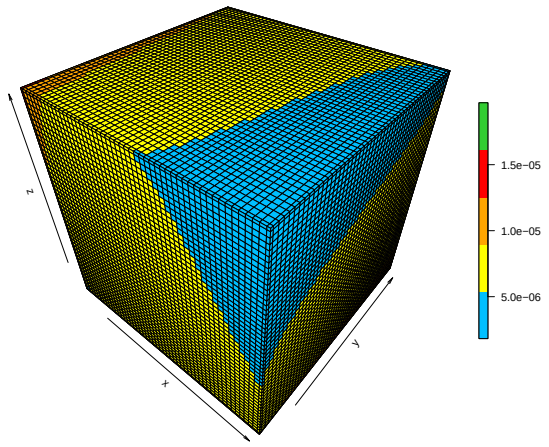


(e) High, Case I

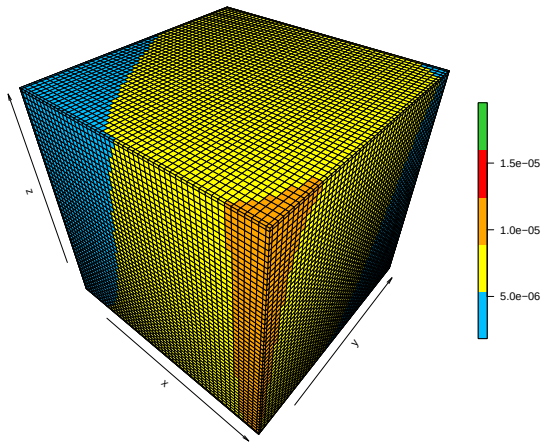


(f) High, Case II

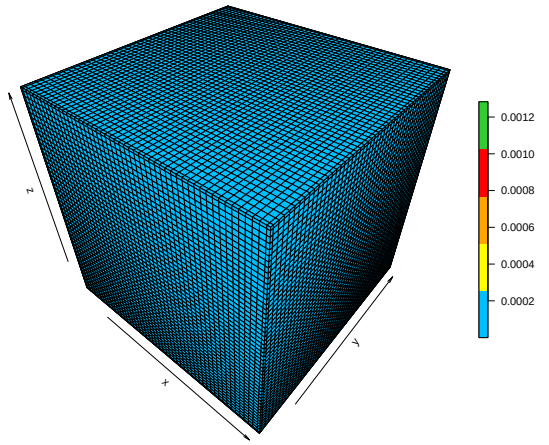
Figure B.2: The h_{ijk} probabilities were obtained from simulated data of size $n = 25$, generated using Gaussian copulas with varying correlation coefficients.



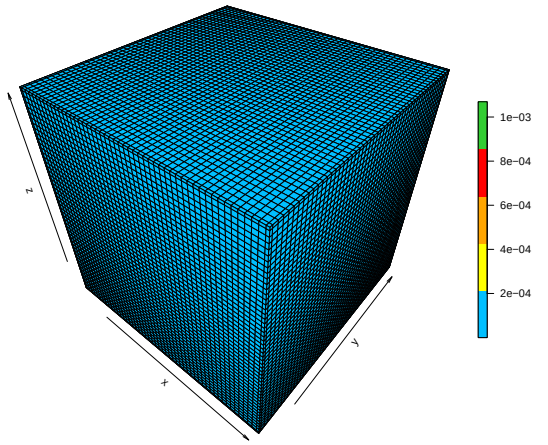
(a) Low, Case I



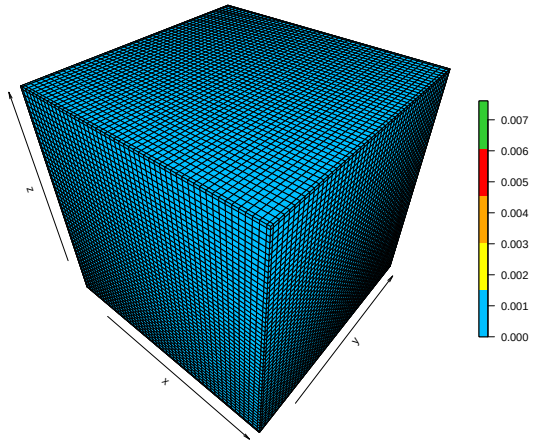
(b) Low, Case II



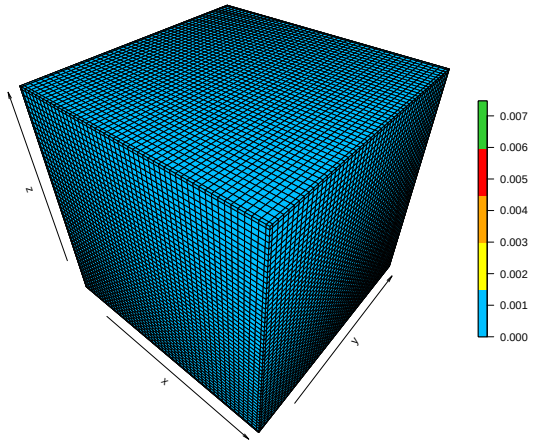
(c) Moderate, Case I



(d) Moderate, Case II

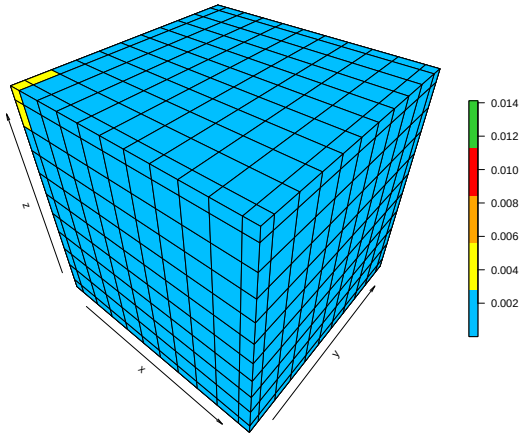


(e) High, Case I

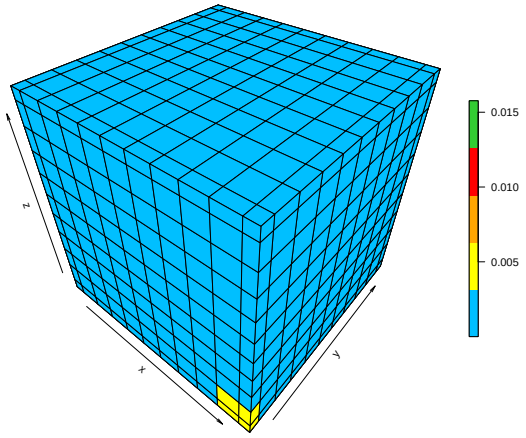


(f) High, Case II

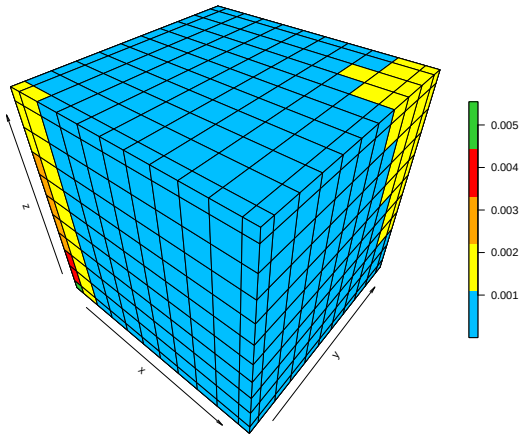
Figure B.3: The h_{ijk} probabilities were obtained from simulated data of size $n = 50$, generated using Gaussian copulas with varying correlation coefficients.



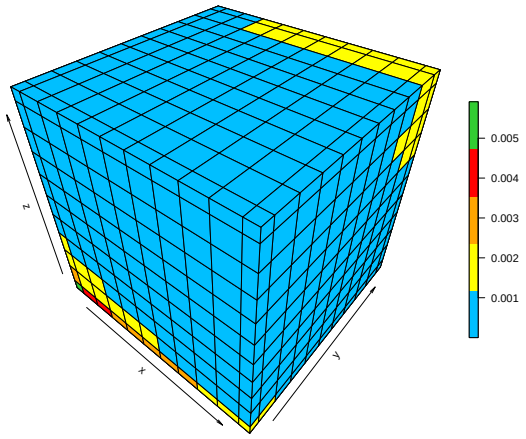
(a) Low, Case III



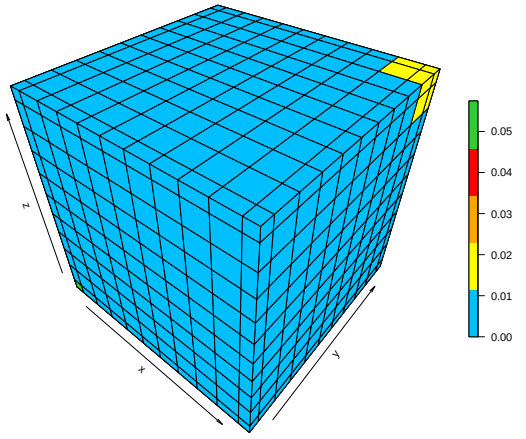
(b) Low, Case IV



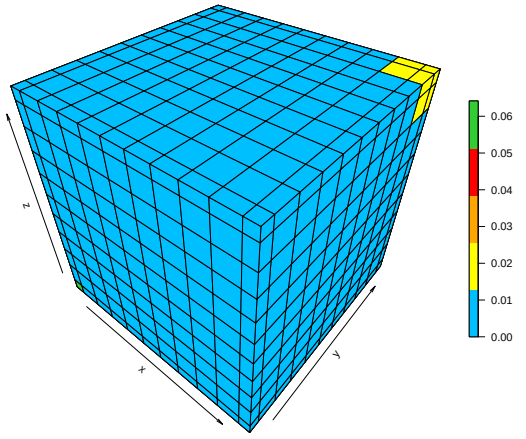
(c) Moderate, Case III



(d) Moderate, Case IV

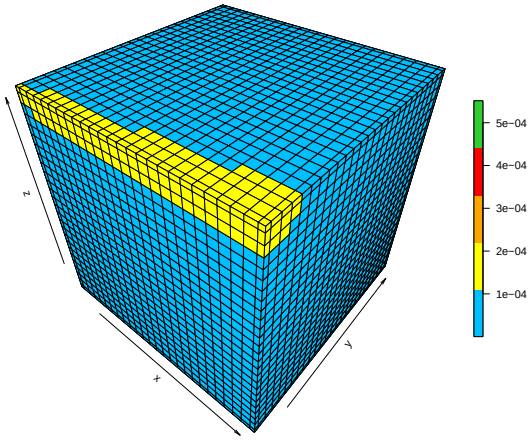


(e) High, Case III

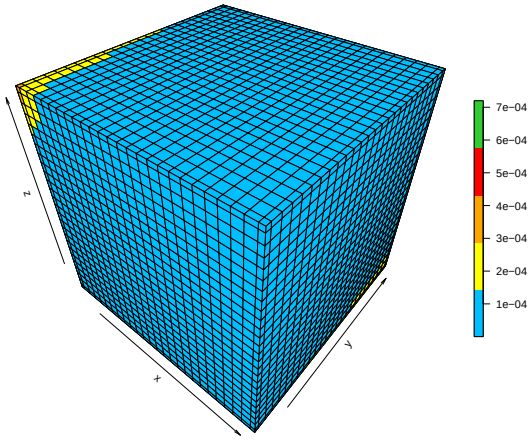


(f) High, Case IV

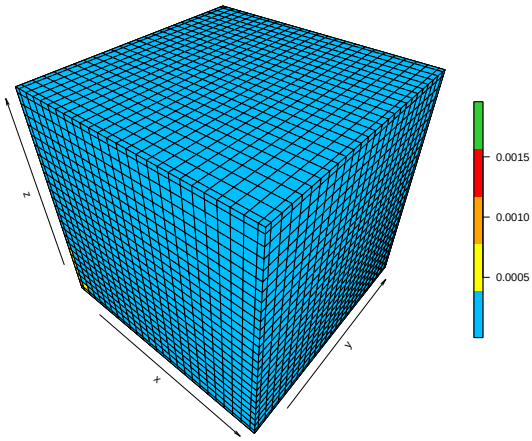
Figure B.4: The h_{ijk} probabilities were obtained from simulated data of size $n = 10$, generated using Gaussian copulas with varying correlation coefficients.



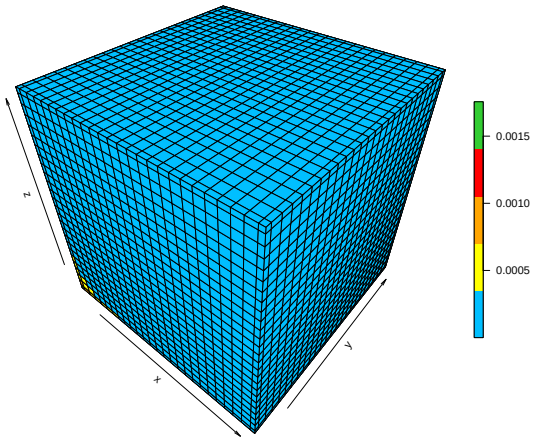
(a) Low, Case III



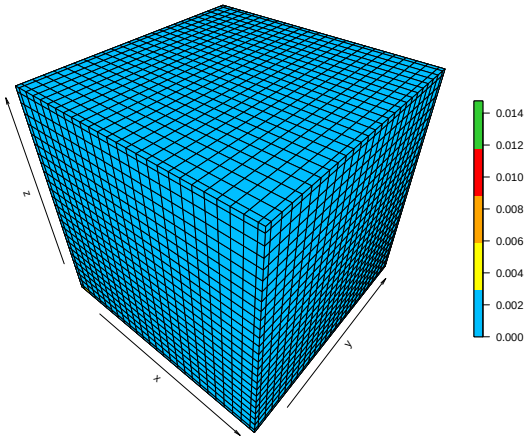
(b) Low, Case IV



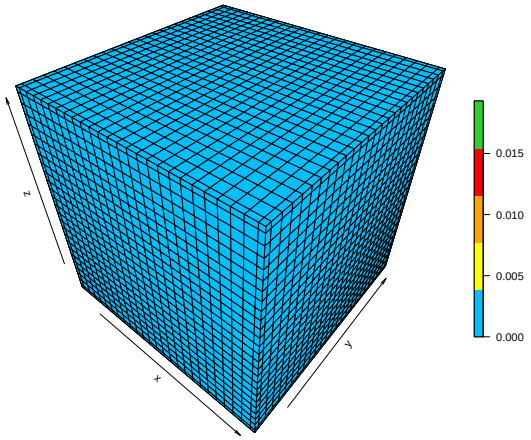
(c) Moderate, Case III



(d) Moderate, Case IV

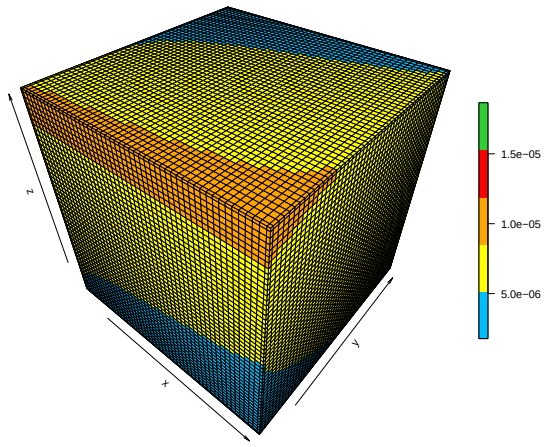


(e) High, Case III

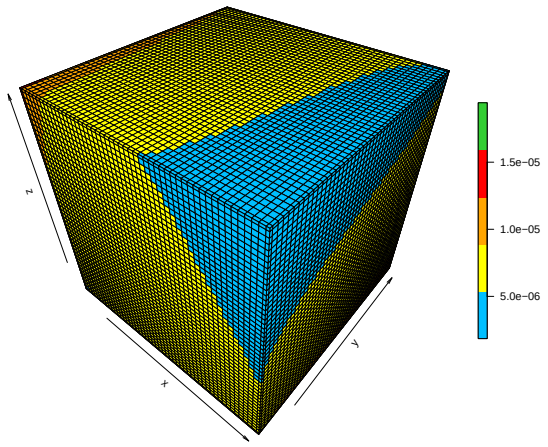


(f) High, Case IV

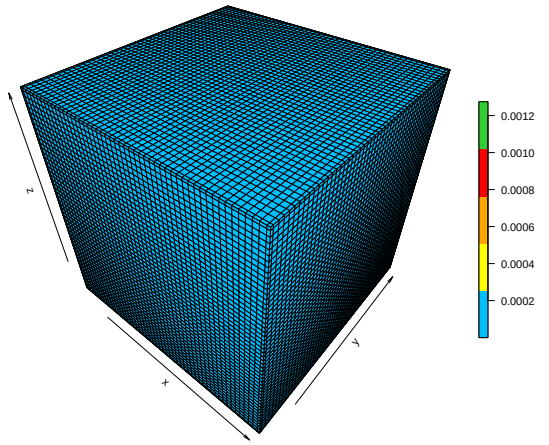
Figure B.5: The h_{ijk} probabilities were obtained from simulated data of size $n = 25$, generated using Gaussian copulas with varying correlation coefficients.



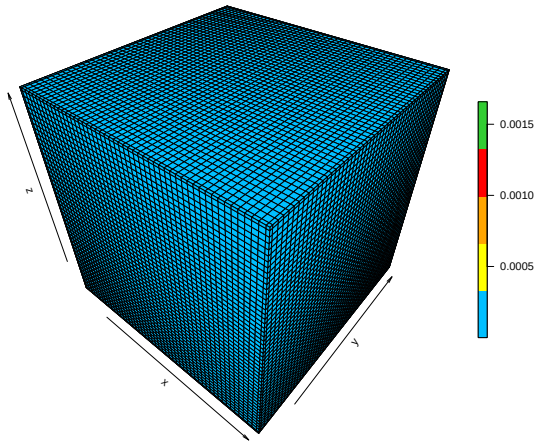
(a) Low, Case III



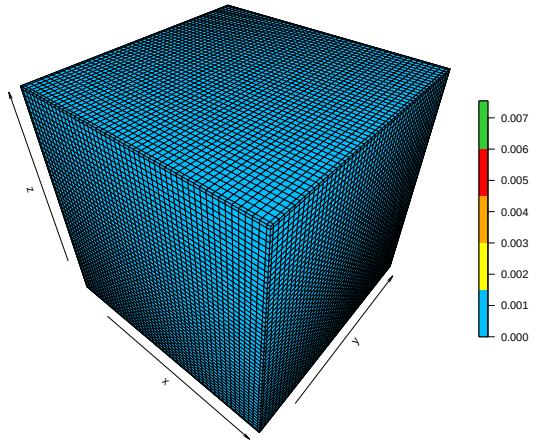
(b) Low, Case IV



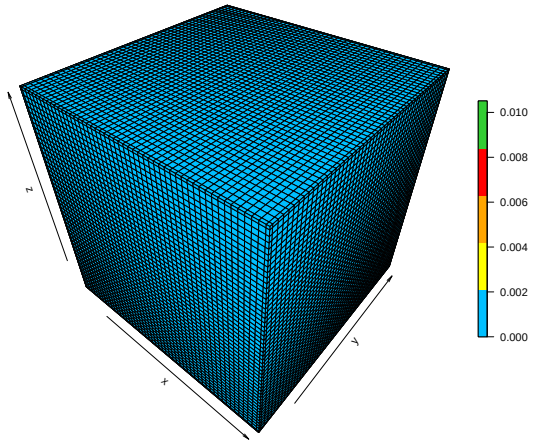
(c) Moderate, Case III



(d) Moderate, Case IV



(e) High, Case III



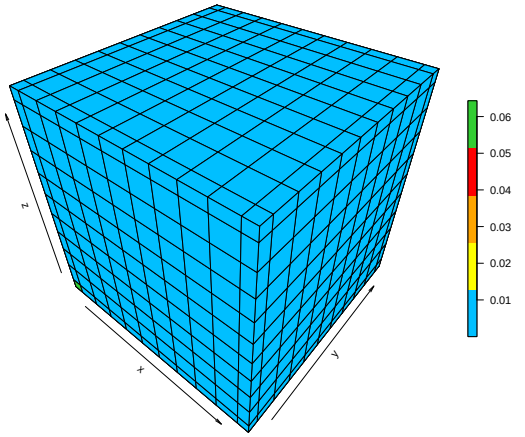
(f) High, Case IV

Figure B.6: The h_{ijk} probabilities were obtained from simulated data of size $n = 50$, generated using Gaussian copulas with varying correlation coefficients.

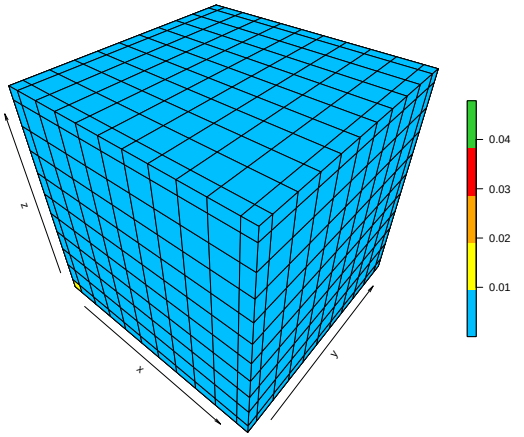
Appendix C

Visualizations of the probabilities h_{ijk}; FNAC

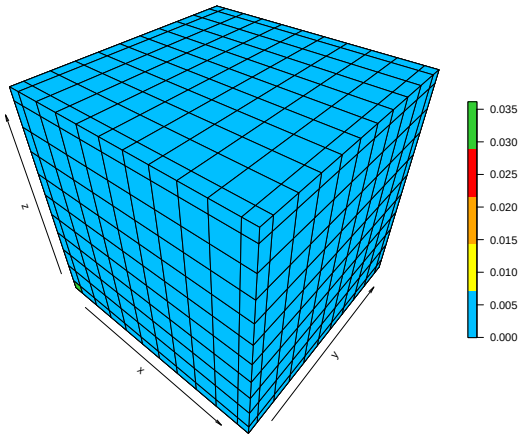
In this appendix, the probabilities h_{ijk} under several types of trivariate FNAC are visualized. It illustrates how the chosen FNAC type, the degree of dependence among variables and the estimation technique both influence the results. Section 5.3 presents the simulation study, which uses these graphical results.



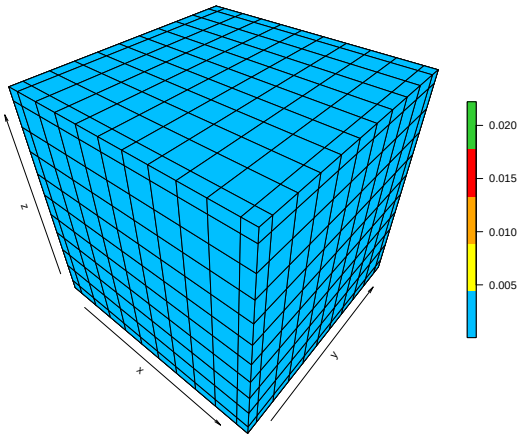
(a) High, Clayton FNAC



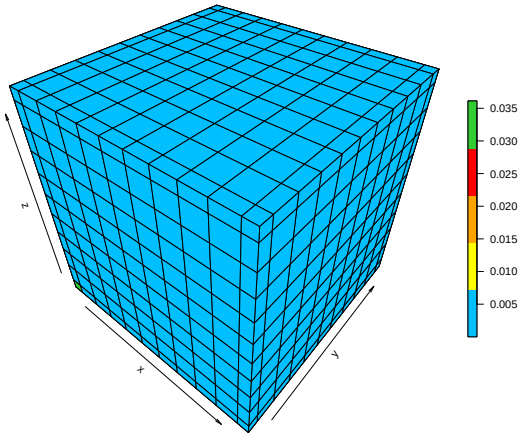
(b) High, Gumbel FNAC



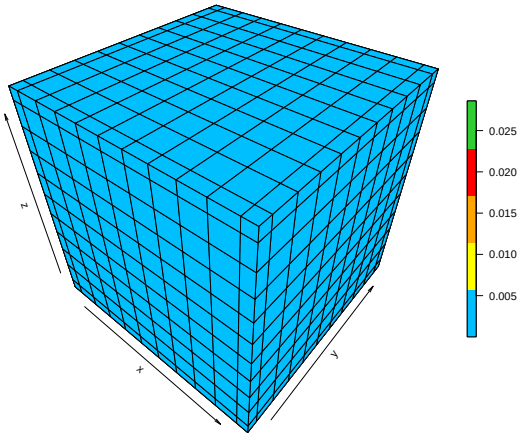
(c) Moderate, Clayton FNAC



(d) Moderate, Gumbel FNAC

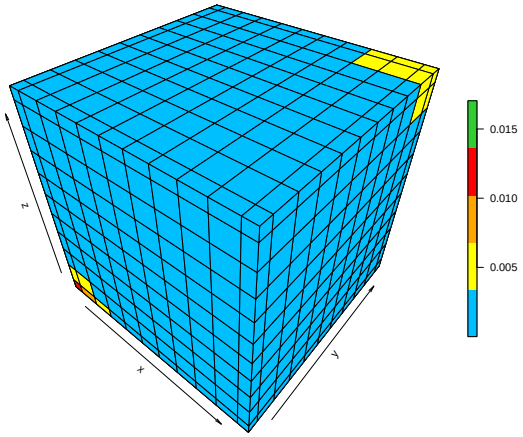


(e) Weak, Clayton FNAC

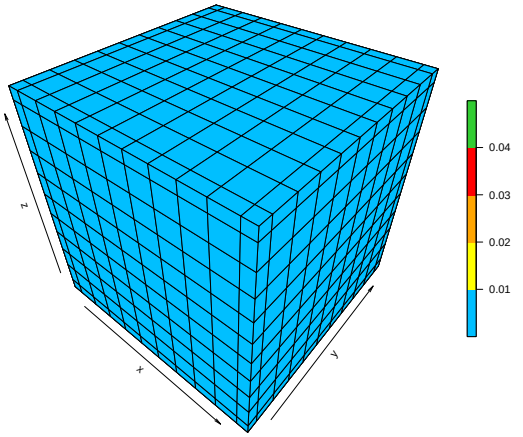


(f) Weak, Gumbel FNAC

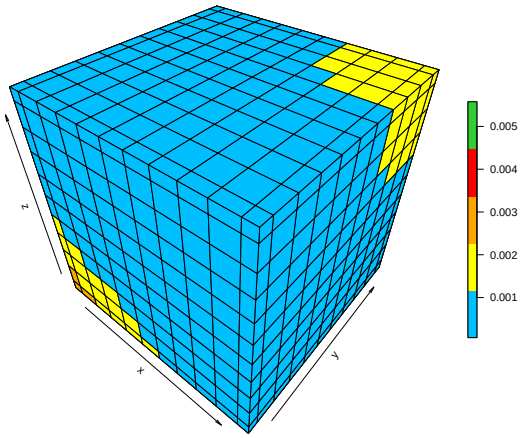
Figure C.1: The h_{ijk} probabilities were obtained from simulated data of size $n = 10$, generated using the Clayton FNAC and the Gumbel FNAC with varying correlation coefficients.



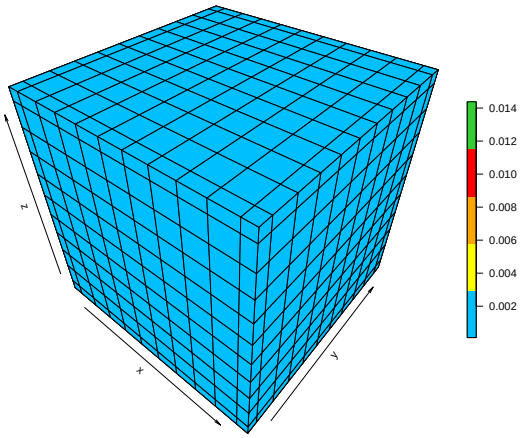
(a) High, Frank FNAC



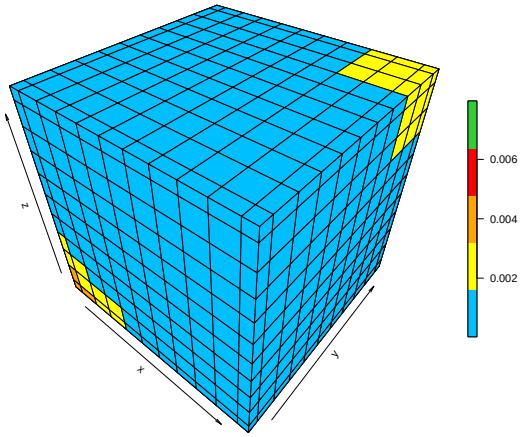
(b) High, Joe FNAC



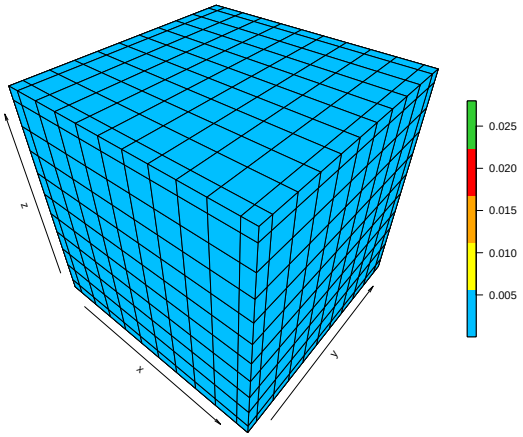
(c) Moderate, Frank FNAC



(d) Moderate, Joe FNAC

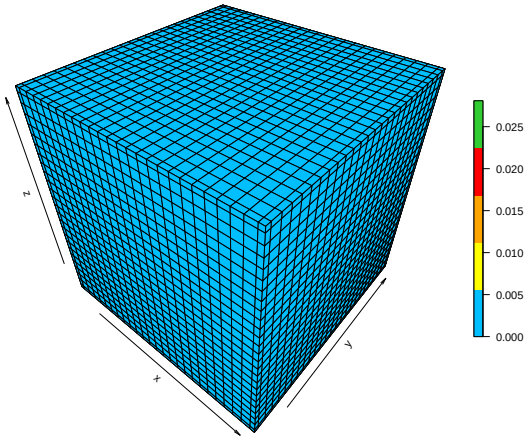


(e) Weak, Frank FNAC

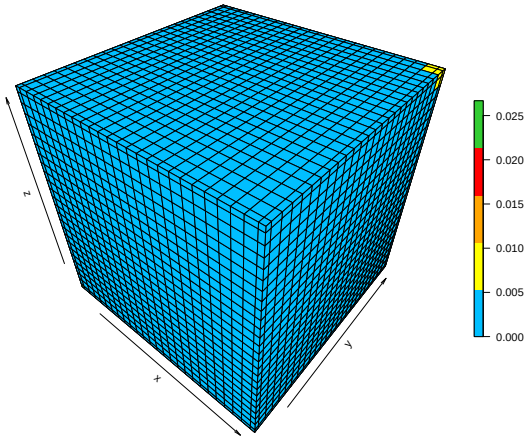


(f) Weak, Joe FNAC

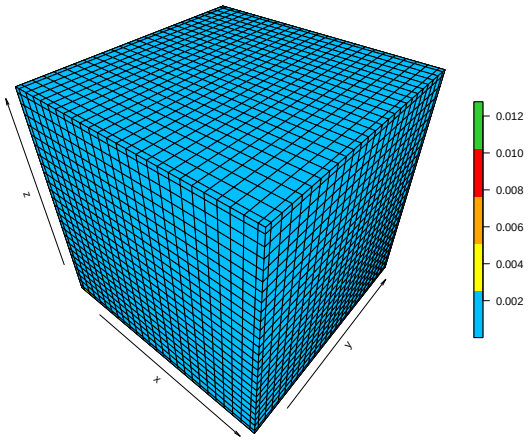
Figure C.2: The h_{ijk} probabilities were obtained from simulated data of size $n = 10$, generated using the Frank FNAC and the Joe FNAC with varying correlation coefficients.



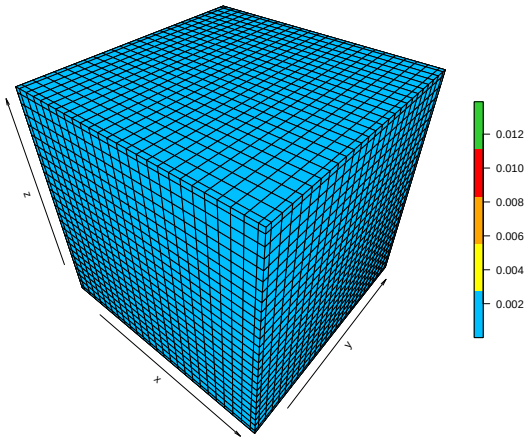
(a) High, Clayton FNAC



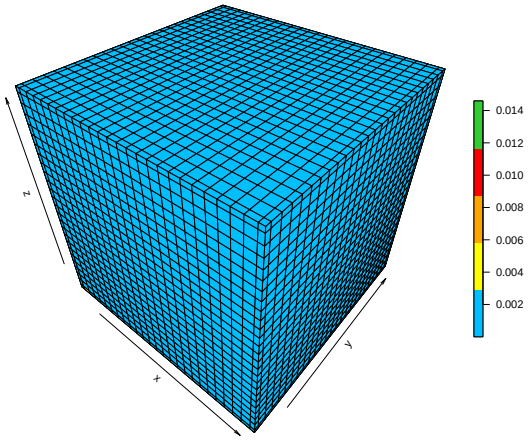
(b) High, Gumbel FNAC



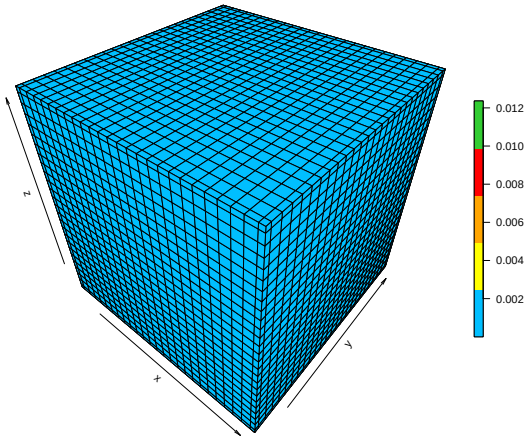
(c) Moderate, Clayton FNAC



(d) Moderate, Gumbel FNAC

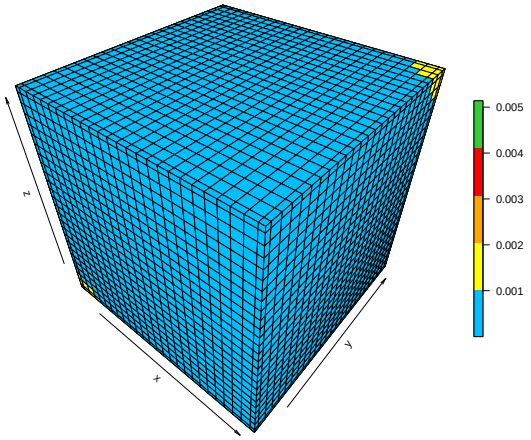


(e) Weak, Clayton FNAC

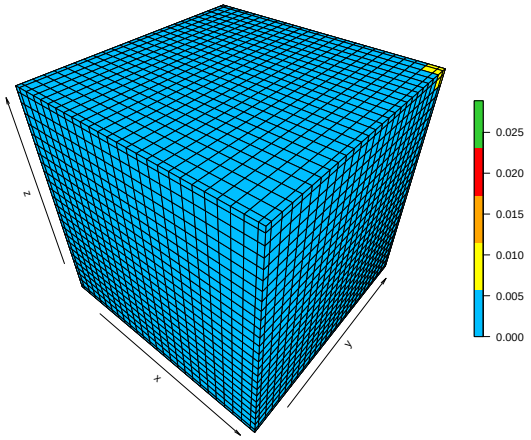


(f) Weak, Gumbel FNAC

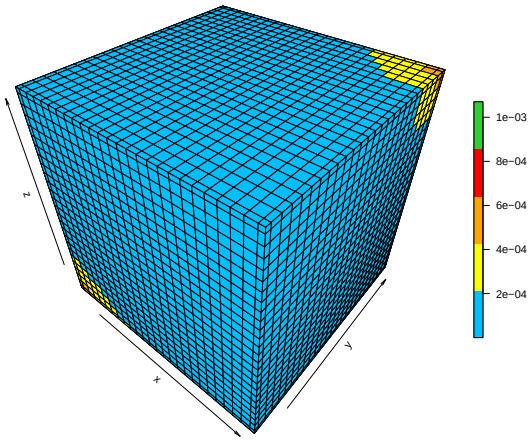
Figure C.3: The h_{ijk} probabilities were obtained from simulated data of size $n = 25$, generated using the Clayton FNAC and the Gumbel FNAC with varying correlation coefficients.



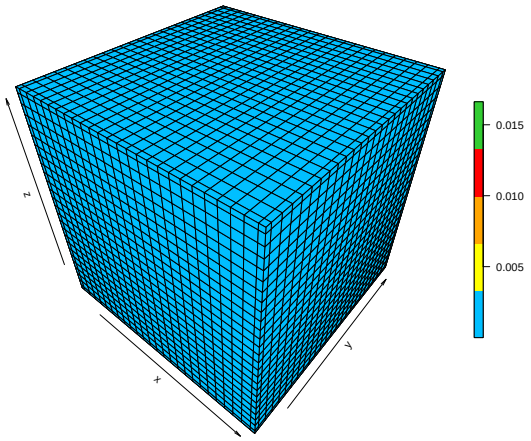
(a) High, Frank FNAC



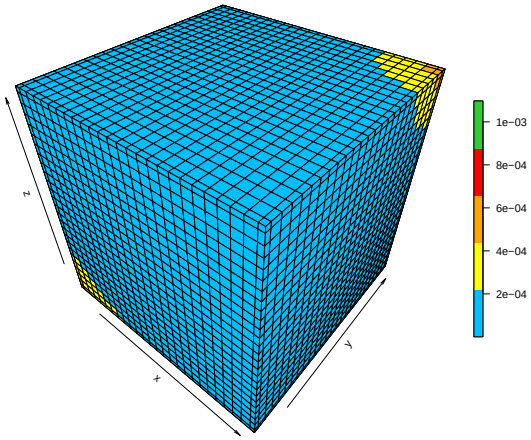
(b) High, Joe FNAC



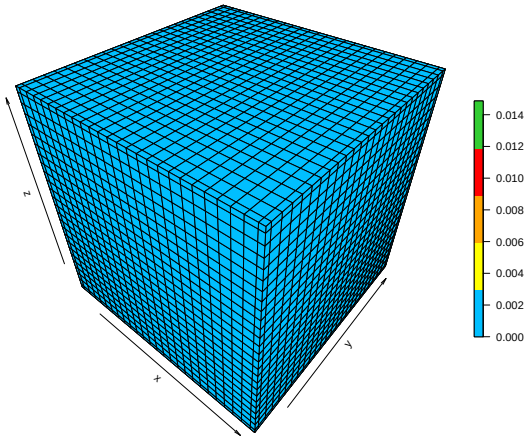
(c) Moderate, Frank FNAC



(d) Moderate, Joe FNAC

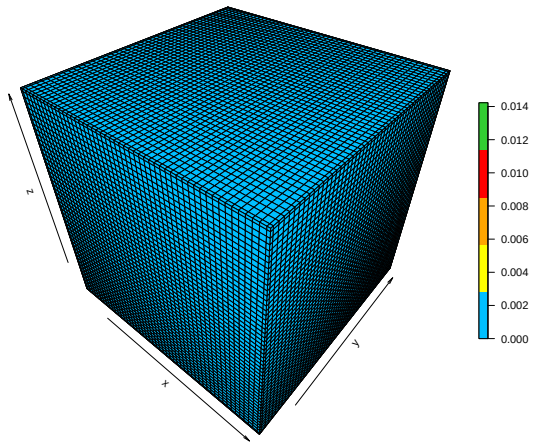


(e) Weak, Frank FNAC

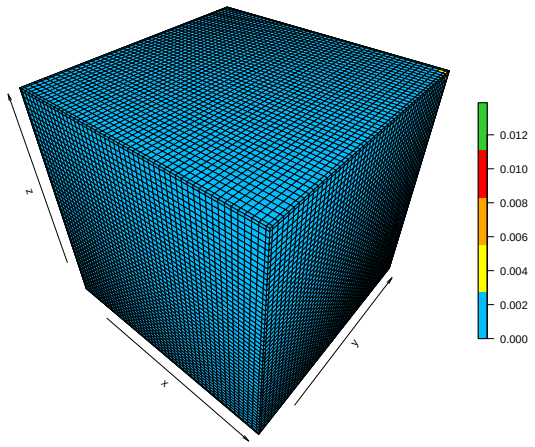


(f) Weak, Joe FNAC

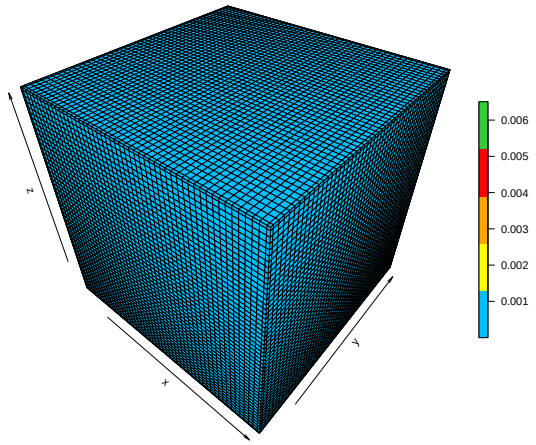
Figure C.4: The h_{ijk} probabilities were obtained from simulated data of size $n = 25$, generated using the Frank FNAC and the Joe FNAC with varying correlation coefficients.



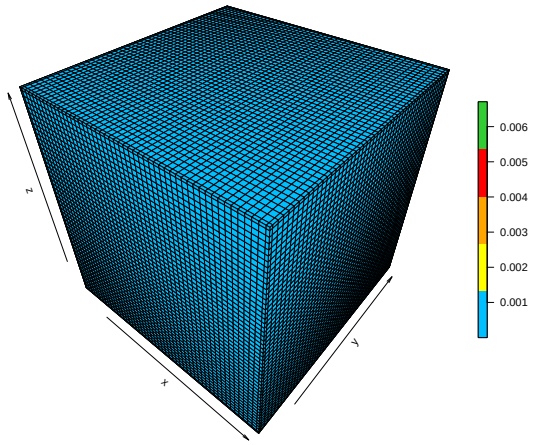
(a) High, Clayton FNAC



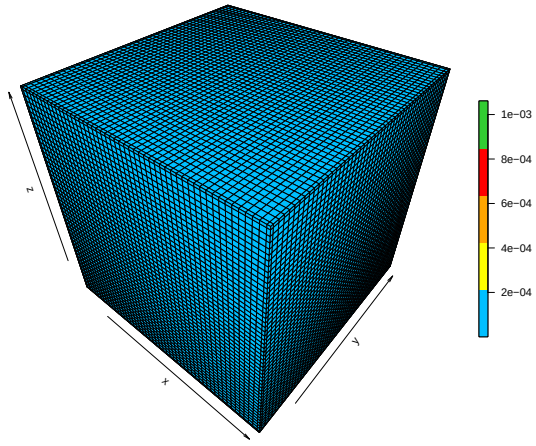
(b) High, Gumbel FNAC



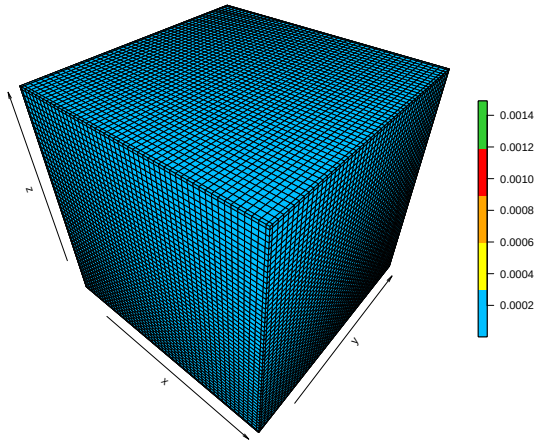
(c) Moderate, Clayton FNAC



(d) Moderate, Gumbel FNAC

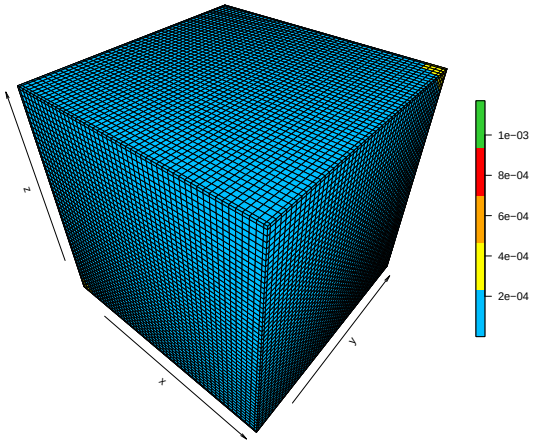


(e) Weak, Clayton FNAC

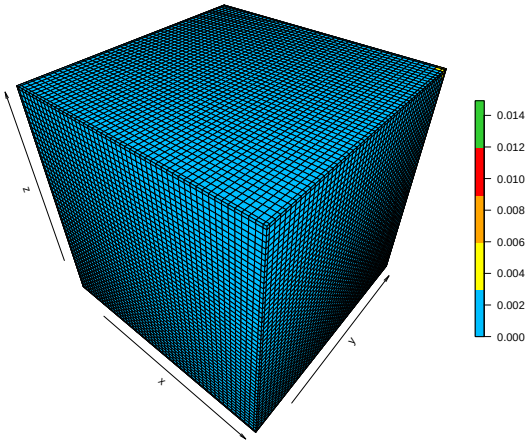


(f) Weak, Gumbel FNAC

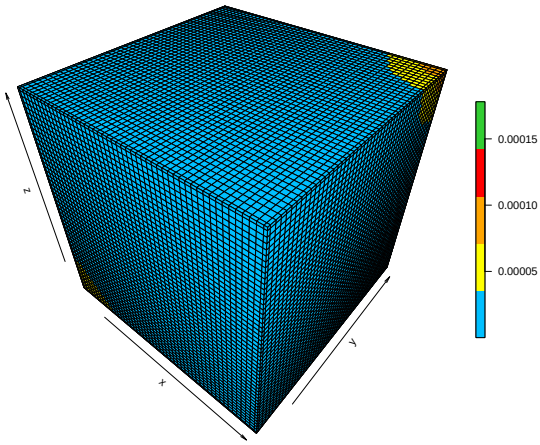
Figure C.5: The h_{ijk} probabilities were obtained from simulated data of size $n = 50$, generated using the Clayton FNAC and the Gumbel FNAC with varying correlation coefficients.



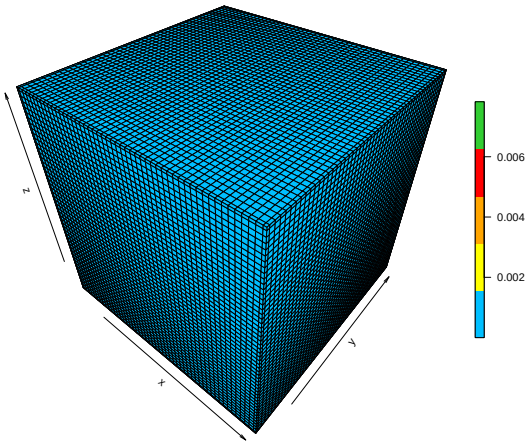
(a) High, Frank FNAC



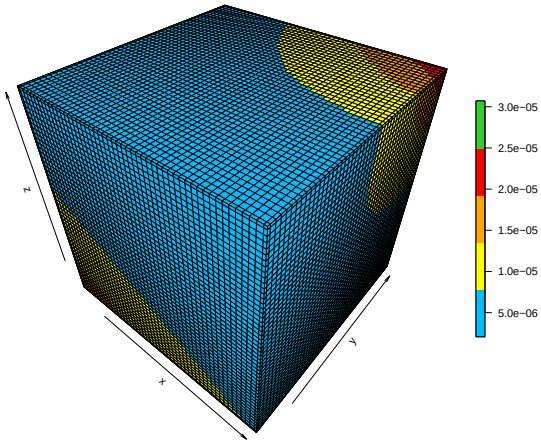
(b) High, Joe FNAC



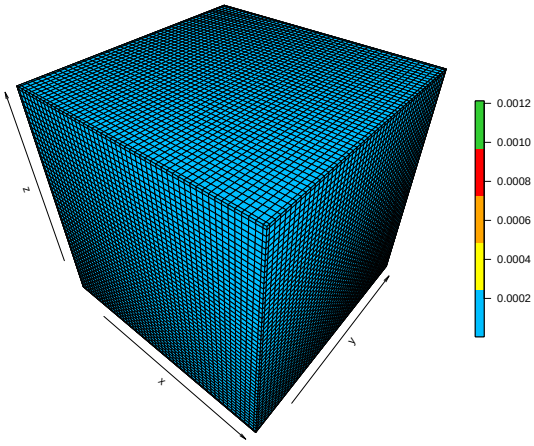
(c) Moderate, Frank FNAC



(d) Moderate, Joe FNAC



(e) Weak, Frank FNAC



(f) Weak, Joe FNAC

Figure C.6: The h_{ijk} probabilities were obtained from simulated data of size $n = 50$, generated using the Frank FNAC and the Joe FNAC with varying correlation coefficients.

Bibliography

[1] Aas, K. (2016). Pair-copula constructions for financial applications: A review. *Econometrics*, 4(4):43.

[2] Aas, K. and Berg, D. (2009). Models for construction of multivariate dependence: A comparison study. *The European Journal of Finance*, 15(7-8):639–659.

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

[4] Acar, E. F., Genest, C., and Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90.

[5] Al Luhayb, A. S. M., Coolen-Maturi, T., and Coolen, F. P. A. (2023). Smoothed bootstrap methods for bivariate data. *Journal of Statistical Theory and Practice*, 17(3):1–37.

[6] Al Luhayb, A. S. M., Coolen-Maturi, T., and Coolen, F. P. A. (2024). Smoothed bootstrap methods for hypothesis testing. *Journal of Statistical Theory and Practice*, 18(1):16.

[7] Ali, M. M., Mikhail, N. N., and Haq, M. S. (1978). A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, 8(3):405–412.

[8] Allen, D. E., McAleer, M., and Singh, A. K. (2017). Risk measurement and risk modelling using applications of vine copulas. *Sustainability*, 9(10):1762.

[9] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Hoboken, N.J.

[10] Augustin, T. and Coolen, F. P. A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124(2):251–272.

[11] Augustin, T., Coolen, F. P. A., De Cooman, G., and Troffaes, M. C. M. (2014). *Introduction to Imprecise Probabilities*. John Wiley & Sons, Chichester, West Sussex.

[12] Ayantobo, O. O., Li, Y., and Song, S. (2019). Multivariate drought frequency analysis using four-variate symmetric and asymmetric Archimedean copula functions. *Water Resources Management*, 33(1):103–127.

[13] Barthélemy, J. and Suesse, T. (2018). mipfp: An R package for multidimensional array fitting and simulating multivariate Bernoulli distributions. *Journal of Statistical Software*, 86(2):1–20.

[14] Bedford, T. J. and Cooke, R. M. (2001a). Monte Carlo simulation of vine dependent random variables for applications in uncertainty analysis. In *Proceedings of ESREL2001*, Turin, Italy. https://strathprints.strath.ac.uk/9662/.

[15] Bedford, T. J. and Cooke, R. M. (2001b). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1):245–268.

[16] Bedford, T. J. and Daneshkhah, A. (2010). Approximating multivariate distributions with vines. Working paper, Management Science Department, University of Strathclyde. Submitted to Operations Research, https://strathprints.strath.ac.uk/40237/.

[17] Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000). Copulas for finance – a reading guide and some applications. Groupe de Recherche Opérationnelle. Working paper. `http://ssrn.com/abstract=1032533`.

[18] Brechmann, E. and Czado, C. (2013). Risk management with high-dimensional vine copulas: An analysis of the Euro Stoxx 50. *Statistics & Risk Modeling*, 30(4):307–342.

[19] Breymann, W., Dias, A., and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1):1–14.

[20] Carriere, J. F. (2000). Bivariate survival models for coupled lives. *Scandinavian Actuarial Journal*, 2000(1):17–32.

[21] Charpentier, A., Fermanian, J.-D., and Scaillet, O. (2007). The estimation of copulas: Theory and practice. In *Copulas: From theory to application in finance*. Risk Books, London. 35-–64.

[22] Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & Sons, Chichester, West Sussex.

[23] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

[24] Coolen, F. P. A. (2011). Nonparametric predictive inference. In *International Encyclopedia of Statistical Science*. Springer, Berlin. 968–970.

[25] Coolen, F. P. A. and Augustin, T. (2005). Learning from multinomial data: A nonparametric predictive alternative to the imprecise dirichlet model. In *ISIPTA 2005: Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*. SIPTA. 5:125–134.

[26] Coolen, F. P. A. (2006). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, 15(1–2):21–47.

[27] Coolen-Maturi, T., Coolen, F. P. A., and Muhammad, N. (2016). Predictive inference for bivariate data: Combining nonparametric predictive inference for marginals with an estimated copula. *Journal of Statistical Theory and Practice*, 10(3):515–538.

[28] Czado, C. (2010). Pair-Copula constructions of multivariate copulas. In *Copula Theory and Its Applications*. Springer, Berlin, Heidelberg. 93–109.

[29] Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R*. Springer, Cham.

[30] Czado, C., Bax, K., Sahin, Ö., Nagler, T., Min, A., and Paterlini, S. (2022). Vine copula based dependence modeling in sustainable finance. *The Journal of Finance and Data Science*, 8:309–330.

[31] Datta, R. and Reddy, M. J. (2023). Trivariate frequency analysis of droughts using copulas under future climate change over Vidarbha region in India. *Stochastic Environmental Research and Risk Assessment*, 37(10):3855–3877.

[32] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Bulletins de l'Académie Royale de Belgique*, 65(1):274–292.

[33] Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The annals of mathematical statistics.*, 11(4):427–444.

[34] Dewick, P. R. and Liu, S. (2022). Copula modelling to analyse financial data. *Journal of Risk and Financial Management*, 15(3):104.

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

[36] Embrechts, P., Lindskog, F., and Mcneil, A. (2003). Modelling Dependence with Copulas and Applications to Risk Management. In *Handbook of Heavy Tailed Distributions in Finance*. Elsevier/North-Holland, Amsterdam. 329–384.

[37] Everitt, B. S. and Dunn, G. (2001). *Applied multivariate data analysis*. Arnold, London.

[38] Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: Applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286.

[39] Frank, M. J. (1979). On the simultaneous associativity of f(x, y) and x + y - f(x, y). *Aequationes Mathematicae*, 19:194–226.

[40] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A*, 14:53–77.

[41] Geenens, G. (2020). Copula modeling for discrete random vectors. *Dependence Modeling*, 8(1):417–440.

[42] Genest, C. and Mackay, R. J. (1986). Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14(2):145–159.

[43] Genest, C., Masiello, E., and Tribouley, K. (2009). Estimating copula densities through wavelets. *Insurance: Mathematics and Economics*, 44(2):170–181.

[44] Genest, C. and Rivest, L. P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043.

[45] Genest, C. Ghoudi, K. and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

[46] Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, M. T. (2021). mvtnorm: Multivariate Normal and t Distributions. *Journal of Computational and Graphical Statistics*, 11:950–971.

[47] Hayfield, T. and Racine, J. S. (2020). Package 'np' nonparametric kernel smoothing methods for mixed data types. *https://github.com/JeffreyRacine/R-Package-np*.

[48] He, T. (2025). Spread option pricing method based on nonparametric predictive inference copula. *Journal of Forecasting*. https://doi.org/10.1002/for.3262.

[49] Hill, B. (1993). Parametric models for $A_{(n)}$: Splitting processes and mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2):423–433.

[50] Hill, B. M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63(322):677–691.

[51] Hill, B. M. (1988). De Finetti's Theorem, Induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference (with discussion). *Bayesian statistics*, 3:211–241.

[52] Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics & Data Analysis*, 52(12):5163–5174.

[53] Hofert, M. (2010). *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. PhD thesis, Universität Ulm, Germany. https://doi.org/10.18725/OPARU-1787.

[54] Hofert, M. and Pham, D. (2013). Densities of nested Archimedean copulas. *Journal of Multivariate Analysis*, 118:37–52.

[55] Hofert, M. and Scherer, M. (2011). CDO pricing with nested Archimedean copulas. *Quantitative Finance*, 11(5):775–787.

[56] Holmgren, W. F., Hansen, C. W., and Mikofski, M. A. (2018). pvlib python: A python package for modeling solar energy systems. *Journal of Open Source Software*, 3(29):884.

[57] Jobson, J. D. (1992). *Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods*. Springer, New York.

[58] Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics*. Institute of Mathematical Statistics, Hayward, CA. 120–141.

[59] Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC, London.

[60] Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman and Hall/CRC, London.

[61] Kaishev, V. K., Dimitrova, D. S., and Haberman, S. (2007). Modelling the joint distribution of competing risks survival times using copula functions. *Insurance: Mathematics and Economics*, 41(3):339–361.

[62] Kauermann, G., Schellhase, C., and Ruppert, D. (2013). Flexible copula density estimation with penalized hierarchical b-splines. *Scandinavian Journal of Statistics*, 40(4):685–705.

[63] Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30:81–93.

[64] Kole, E., Koedijk, K., and Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking & Finance*, 31(8):2405–2423.

[65] Košir, T. and Perrone, E. (2025). Discrete imprecise copulas. *Fuzzy Sets and Systems*, 504:109251. https://doi.org/10.1016/j.fss.2024.109251.

[66] Kurowicka, D. and Cooke, R. M. (2006). *Uncertainty analysis with high dimensional dependence modelling.* John Wiley & Sons, Chichester, West Sussex.

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

[68] Lan, Z. and Singh, V. P. (2019). Asymmetric Copulas. In *Copulas and their Applications in Water Resources Engineering.* Cambridge University Press, Cambridge. 172–241.

[69] Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice.* Princeton University Press, New York.

[70] Li, X., Mikusiński, P., Sherwood, H., and Taylor, M. D. (1997). On approximation of copulas. In *Distributions with given Marginals and Moment Problems.* Springer, Dordrecht. 107–116.

[71] Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis*, 99(10):2234–2250.

[72] Lindskog, F., McNeil, A., and Schmock, U. (2003). Kendall's tau for elliptical distributions. In *Credit Risk.* Physica-Verlag HD, Heidelberg. https://doi.org/10.1007/978-3-642-59365-9$_8$.

[73] Madadgar, S. and Moradkhani, H. (2013). Drought analysis under climate change using copula. *Journal of Hydrologic Engineering*, 18(7):746–759.

[74] McNeil, A. J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.

[75] McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Annals of Statistics*, 37(5B):3059–3097.

[76] Muhammad, N. (2016). *Predictive Inference with Copulas for Bivariate Data*. PhD thesis, Durham University, UK. https://npi-statistics.com/pdfs/theses/NM16.pdf.

[77] Muhammad, N., Coolen, F. P. A., and Coolen-Maturi, T. (2016). Predictive inference for bivariate data with nonparametric copula. *AIP Conference Proceedings*, 1750(1):0600041—0600048. https://doi.org/10.1063/1.4954609.

[78] Muhammad, N., Coolen-Maturi, T., and Coolen, F. P. A. (2018). Nonparametric predictive inference with parametric copulas for combining bivariate diagnostic tests. *Statistics, Optimization & Information Computing*, 6(3):398–408.

[79] Muhammad, N., Roslin, A. H., Daud, H., Kadir, E. A., and Maharani, W. (2024). Predicting forest fire spots using nonparametric predictive inference with parametric copula: Malaysia case study. *AIP Conference Proceedings*, 3189(1):100003. https://doi.org/10.1063/5.0224342.

[80] Muhammad, N. and Yusoff, N. (2018). Nonparametric predictive inference with parametric copula for survival analysis. *MATEC Web of Conferences*, 189:03026. 10.1051/matecconf/201818903026.

[81] Nagler, T. (2014). Kernel methods for vine copula estimation. Master's thesis, Technische Universität München, Germany. `https://mediatum.ub.tum.de/node?id=1231221`.

[82] Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., and Erhardt, T. (2024). *VineCopula: Statistical inference of vine copulas*. R package version 2.6.1 http://cran.r-project.org/web/packages/VineCopula/.

[83] Nelsen, R. B. (1991). Copulas and Association. In *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*. Springer, Dordrecht. 51–74.

[84] Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer New York, NY.

[85] Nikoloulopoulos, A. K., Joe, H., and Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis*, 56(11):3659–3673.

[86] Okhrin, O., Okhrin, Y., and Schmid, W. (2013). Properties of hierarchical archimedean copulas. *Statistics & Risk Modeling*, 30(1):21–54.

[87] Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107(499):1063–1072.

[88] Roslin, A. H. and Muhammad, N. (2024). Nonparametric Predictive Inference Forest Fire Dashboard. *Procedia Computer Science*, 245:255–262. 9th International Conference on Computer Science and Computational Intelligence (ICCSCI 2024), https://doi.org/10.1016/j.procs.2024.10.250.

[89] Roslin, A. H., Muhammad, N., and Kadir, E. A. (2025). Forecasting locations of forest fires in indonesia through nonparametric predictive inference with parametric copula: A case study. *Journal of Quality Measurement and Analysis (JQMA)*, 21(1):237–251. https://doi.org/10.17576/jqma.2101.2025.15.

[90] Savu, C. and Trede, M. (2010). Hierarchies of Archimedean copulas. *Quantitative Finance*, 10(3):295–304.

[91] Schweizer, B. and Sklar, A. (1960). Statistical metric spaces. *Pacific Journal of Mathematics*, 10(1):313–334.

[92] Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Space*. North-Holland, New York.

[93] Shi, P. and Frees, E. W. (2010). Long-tail longitudinal modeling of insurance company expenses. *Insurance: Mathematics and Economics*, 47(3):303–314.

[94] Shi, P. and Yang, L. (2018). Pair copula constructions for insurance experience rating. *Journal of the American Statistical Association*, 113(521):122–133.

[95] Silverman, B. W. (1986). *Density estimation: For statistics and data analysis*. Chapman & Hall, London.

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

[97] Steadman, R. G. (1984). A universal scale of apparent temperature. *Journal of Applied Meteorology*, 23(12):1674–1687.

[98] Stöber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructions—Limitations and extensions. *Journal of Multivariate Analysis*, 119:101–118.

[99] Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance*, 4(3):339–352.

[100] Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21(4):1–21.

[101] Yang, L. and Czado, C. (2022). Two-part D-vine copula models for longitudinal insurance claim data. *Scandinavian Journal of Statistics*, 49(4):1534–1561.

[102] Zhang, X. and Jiang, H. (2019). Application of copula function in financial risk analysis. *Computers & Electrical Engineering*, 77:376–388.

[103] Émile, M. and Gumbel, J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 9:171–173.

[104] Žežula, I. (2009). On multivariate Gaussian copulas. *Journal of Statistical Planning and Inference*, 139(11):3942–3946.