

Modelling Tail Risk with Empirical Copulas

Tong Pun Yat

STAT4798 Statistics and Actuarial Science Project

Department of Statistics and Actuarial Science

The University of Hong Kong

30 April 2024

Abstract

The empirical copula is a nonparametric method to model dependence, without the need of pre-specifying marginal distributions. Over the years, various statisticians have proposed new variations of the empirical copula, such as the empirical beta copula, empirical checkerboard copula, and most recently, the data-adaptive empirical copulas. Assessing tail risk with such copulas often leads to underestimation, due to the lack of available data beyond the largest / smallest observation which leads to difficulties in fitting the upper and lower tail accurately. This report provides an oversight on how well empirical copulas can fit the upper and lower tail, by computing the exceedance and cumulative probabilities, and comparing the estimated probabilities with the corresponding true copula. It is found that the data-adaptive empirical copulas proposed by [Kojadinovic and Yi, 2024](#) exhibits the best performance despite some erratic patterns and systematic underestimation. The low-Kendall's-tau region is also heavily affected by the limited sample size.

Preface

Prof. Hofert has provided inspiration and motivation behind this project (understanding the overestimation and underestimation of tail risk). I have developed the algorithm in its entirety, reviewed by Prof. Hofert, where he has given comments on how to improve the simulation. The report is also typed entirely by me, where the first half of the report are prerequisites, which are in reference with various relevant publications (most notably [Hofert et al., 2018](#), [Kojadinovic and Yi, 2024](#), [Segers et al., 2017](#)); and the second half of the report is original, with parts of the R code referencing [Hofert et al., 2023](#), [Kojadinovic and Yi, 2024](#) with some modifications from me.

Acknowledgements

I would like to thank Prof. Hofert, my supervisor for STAT4798, for his continued guidance in the entirety of this project, where he has introduced the underlying motivation around this topic and on copulas overall. I would also like to especially thank my friend Yao Hui for assisting in running the (super long) R algorithm for me (as my computer did not have sufficient memory / RAM to handle the simulation). I would also like to thank my friends for their continued support during this course, and lastly my girlfriend Tiffannie, or I wouldn't have stood till now.

Contents

1	Introduction	6
1.1	Motivation	6
2	Empirical Copulas	8
2.1	Empirical copula	9
2.2	Empirical checkerboard copula	10
2.3	Empirical beta copula	11
2.4	Smoothed and data-adaptive copulas based on the empirical beta copula: Generalisations	12
3	Model specifications	14
3.1	Copulas and Parameters	14
3.2	Sampling from the true copula	17
3.3	Simulating exceedance and cumulative probabilities	18
3.4	Model Evaluation: Cramer-von-Mises (CvM) Statistic	26
4	Results	27
4.1	Preparation	27
4.2	Plots: exceedance probabilities vs \mathbf{u} (evaluation points)	28
4.3	CvM: Exceedance probabilities	32
4.4	Analysis: Exceedance probabilities	36
4.5	Plots: \mathbf{u} (evaluation points) vs cumulative probabilities	41
4.6	CvM: Cumulative probabilities	45
4.7	Analysis: Cumulative probabilities	49
4.8	Recommendations	51
5	Conclusion	52
1	Appendix: Copula Prerequisites	56
1.1	Definition and Construction	56
1.2	Common Copula distributions	59

1.3	A Primer in Correlation Measures	63
1.4	Survival Copulas and Rotated Copulas	69
2	Appendix: Empirical Copulas	71
2.1	Empirical beta copula	71
2.2	Smoothed and data-adaptive copulas proposed by Kojadinovic and Yi, 2024	74

1 Introduction

1.1 Motivation

In introductory statistics, we borne the need to understand the relationship between multiple underlying variables to make statistical inference. By analysing them in a univariate way, we gain intuition on the distributional characteristics of the marginal distributions, while multivariate analysis allow us to understand the dependence between variables. To begin with, we can start with linear dependence. Notably, we generalise the linear relationship between variables with correlation coefficients:

$$\rho = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1d} \\ \rho_{21} & 1 & \dots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{pmatrix}, \rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}}$$

However, we realise that this metric is only useful when evaluating linear relationships between variables (or transformed variables). A low correlation coefficient does not imply a weak relationship as variables might exhibit non-linear dependence.

Copulas, as investigated in this report, can describe the dependence between the components of the random vectors. However in reality, while dealing with messy datasets, it is challenging to infer the unknown copula and marginal distributions by observation. Parametric or semiparametric methods of estimating the copula are often biased as the parametric family of copulas have to be specified beforehand, which is often unrealistic. A possible workaround of this problem is to utilise nonparametric estimation of copulas, which does not require prior knowledge on the family of copulas when undergoing statistical inference.

Empirical copulas, first proposed by [Deheuvels, 1979](#), analyse the multi-dimensional pseudo-observations (estimated nonparametrically as the ranks of the components of the random vector) componentwise, and can be shown to have a similar shape to the real estimated copula. However, tbhe classical empirical copula is not a copula itself as it has

discrete uniform margins instead of continuous ones.

To overcome this shortcoming, multiple smoothed copulas built on the basis of empirical copulas are proposed. They include a multilinear extension of the empirical copula, the empirical checkerboard copula (as defined by [Carley and Taylor, 2002](#) and [Hofert et al., 2018](#)), the empirical beta copula ([Segers et al., 2017](#)), and recently, generalised smooth copula estimators (that contain the empirical beta copula) proposed by [Kojadinovic and Yi, 2024](#).

In present time, tail risk is still little understood and often underestimated. It is known that the usage of the no-tail-dependent Gaussian copula in the modelling of default probability of securities (and tranches) with mortgage underlyings have indirectly led to the financial crisis in 2007, where securities were mispriced due to the underestimation of tail default risk. The usage of the wordings “1-in-500” in describing the massive flood that happened in Hong Kong during late 2023 is widely misunderstood, as there are no 500-year worth of data in understanding an event of this scale. Yet, due to global warming, these large scale national disasters will be recurring (or even commonplace) in the near future, which brings the urgency in developing tail risk modelling techniques.

In this report, we try to understand whether empirical copulas systematically underestimate tail risk with respect to the true underlying copula by computing the up-tail and low-tail probability of the class of empirical copulas and compare it to the true up-tail and low-tail probability. This could extend the possibility that even for a small sample size empirical copulas can be more widely applied in the modelling of tail risks without the risk of misspecifying the class of copulas and with added accuracy compared to other parametric / semiparametric methods.

2 Empirical Copulas

There are a myriad of parametric or semiparametric methods for estimating copulas. However, there are multiple pitfalls. In order to make an accurate estimation of a joint distribution, it is necessary to specify the marginal distributions. After this procedure, it is needed to specify the copula that quantifies the dependence. The parameters of the underlying marginal distributions and the copula can be estimated by classical parametric methods such as maximum likelihood estimation. A slight getaround for the potential issues of parametric methods involves using the empirical distribution function to model the marginal distributions. This still does not solve the problem of specifying the underlying copula. One potential problem of misspecification is that, parametric or semiparametric methods become useless as even if the parameter is estimated, it still does not provide an accurate or meaningful description due to the existence of bias in the estimation of the underlying marginal distributions or copulas. Moreover, the usage of maximum likelihood estimation methods becomes computationally prohibitive (due to the need of numerical optimization techniques and even they will have restrictions), which exacerbates the possibility of creating bias in the parameter estimation.

Therefore, this report specifically focuses on a type of nonparametric estimation, which are empirical copulas. Below, the types of empirical copulas will be introduced, as they are extensively used in this report.

2.1 Empirical copula

One of the advantages of extending copulas to an empirical setting is that we do not have to specify parametric models *a priori*. The following definition is with reference to [Segers et al., 2017](#) first proposed by [Deheuvels, 1979](#).

Definition 2.1.1 ([Deheuvels, 1979](#), [Segers et al., 2017](#)) ***Empirical copula***

Consider i.i.d random vectors $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,d})$ for $i = \{1, \dots, n\}$. Denote $R_{i,j}$ be the rank of $X_{i,j}$ within $X_{1,j}, \dots, X_{n,j}$.

Consider an empirical distributions:

$$\hat{F}_{n,j}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_{i,j} \leq x\}.$$

Substitute x with $X_{i,j}$ with the reasoning that $\hat{F}_{n,j}(X_{i,j}) \sim \mathcal{U}(0, 1)$ asymptotically then

$$\hat{F}_{n,j}(X_{k,j}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_{i,j} \leq X_{k,j}\} = \frac{R_{k,j}}{n}.$$

As $X_{i,1}, \dots, X_{i,d}$ are i.i.d random vectors, the empirical copula is the empirical distribution function of $(\hat{F}_{n,1}(X_{i,1}), \dots, \hat{F}_{n,d}(X_{i,d}))$ given by

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\left\{\frac{\mathbf{R}_i}{n} \leq \mathbf{u}\right\} = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{1}\left\{\frac{R_{i,j}}{n} \leq u_j\right\}.$$

Proposition 2.1 ***Empirical copula is a multivariate step function***

It can be seen that, by the nature of ranking observations based on relative size of the component samples $X_{1,j}, \dots, X_{n,j}$, $\forall u_j$, there always exist $u_j + \epsilon$, where ϵ is an arbitrary constant where the copula will output the same value. Hence, it follows that the empirical copula is a d -variate step function.

The name empirical copula is actually (and ironically) an oxymoron. As the empirical copula is a step function, it violates the continuity requirements needed for it to be classified as a copula.

2.1.1 Survival function of the empirical copula

One important implication of this is we couldn't construct an equivalent version of the survival function of the empirical copula based on Appendix Theorem 1.4.2, as it will result in a non-monotone exceedance probability. To act as a proxy, an alternative formulation of the survival function that corresponds to the survival copula of the empirical copula counterpart is constructed below:

Definition 2.1.2 *Survival function of the empirical copula*

The proxy survival function for the empirical copula is defined as:

$$\bar{H}(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left\{ \frac{\mathbf{R}_i}{n} \geq \mathbf{u} \right\} = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbf{1} \left\{ \frac{R_{i,j}}{n} \geq u_j \right\}.$$

Due to its step nature, it often has high bias compared with the true underlying copula. Hence, statisticians have proposed some smoothed versions of the empirical copula as a getaround.

2.2 Empirical checkerboard copula

The empirical checkerboard copula roughly resembles the empirical copula, but instead of imposing a step function, it is based on a Latin hypercube idea. Hence, the empirical checkerboard copula is absolutely continuous in ran F .

Definition 2.2.1 *Empirical Checkerboard Copula*

The empirical checkerboard copula is defined by

$$C_n^{\boxplus}(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \min \{ \max \{ n u_j - R_{i,j} + 1, 0 \}, 1 \}.$$

If the component samples $X_{1,j}, \dots, X_{n,j}$ do not have ties, the rank that corresponds to the component samples are unique. At the same time, the empirical checkerboard copula is absolutely continuous and hence the empirical checkerboard copula is a true copula, contrary to the unsmoothed empirical copula. By observation of the analytical form of the empirical checkerboard copula, it does not seem to have an obvious form for the survival copula, hence we compute the corresponding survival function of the underlying empirical checkerboard copula from Appendix Theorem 1.4.2.

2.3 Empirical beta copula

The empirical beta copula is proposed by Segers et al., 2017, as a special case of the Bernstein copula.

Definition 2.3.1 (Segers et al., 2017) *Empirical Beta Copula*

The empirical beta copula is defined by:

$$C_m^\beta(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{n,R_{i,j}}(u_j), \quad \mathbf{u} \in [0, 1]^d$$

where $F_{n,r}$ is the cumulative distribution function of the beta distribution with shape parameters r and $n+1-r$ respectively, i.e. $\mathcal{B}(r, n+1-r)$.

This formulation seems rather unintuitive compared to the relatively simple intuition provided by the empirical copula.

2.3.1 Survival function of the empirical beta copula

Recall that the empirical beta copula can be rewritten as:

$$C_n^\beta(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d P(V_{R_{i,j},j} \leq u_j) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{n,R_{i,j}}(u_j), \quad \mathbf{u} \in [0, 1]^d$$

We propose a (proxy) survival function for the empirical beta copula.

Definition 2.3.2 *Survival function of the empirical beta copula*

The survival function of the empirical beta copula is defined by:

$$\bar{H}^\beta(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d P(V_{R_{i,j},j} \geq u_j) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d (1 - F_{n,R_{i,j}}(u_j)), \quad \mathbf{u} \in [0, 1]^d$$

To validate our proxy survival function for the corresponding empirical beta copula, we construct the survival function by Appendix Theorem 1.4.2. It can be numerically shown that they are equivalent. For computational efficiency, the defined survival function would be used as a proxy.

2.4 Smoothed and data-adaptive copulas based on the empirical beta copula: Generalisations

Recall that in the previous section, we have rewritten the empirical Bernstein copula as:

$$B_{\mathbf{m}}(C_n(\mathbf{u})) = \int_{[0,1]^d} C_n(\mathbf{w}) dD_{\mathbf{A}}(\mathbf{w})$$

where $\mathbf{A} \sim \left(\frac{S_{m_1, u_1}}{n}, \dots, \frac{S_{m_d, u_d}}{n} \right)$ and $S_{m_j, u_j} \sim \mathbb{B}(m_j, u_j)$, $j \in \{1, \dots, d\}$.

The empirical Bernstein copula is thus a mixture of empirical copulas, weighted by the smoothing distribution $D_{\mathbf{A}}(\mathbf{w})$. If the law of the random vector \mathbf{A} is modified (by changing its dependencies / moments), it might provide a better predictive performance compared to the original empirical beta copula.

A generalized class of empirical copula with data-adaptive smoothing is defined below:

Definition 2.4.1 (Kojadinovic & Yi, 2024) *Generalized smoothed empirical copula*

The generalized smoothed empirical copula is defined as:

$$C_n^v(\mathbf{u}) = \int_{[0,1]^d} C_n(\mathbf{w}) dD_{\mathbf{W}_u^X}(\mathbf{w})$$

where \mathbf{W}_u^X is an arbitrary random vector.

2.4.1 The smoothed and data-adaptive copulas proposed by Kojadinovic and Yi

Definition 2.4.2 (Kojadinovic & Yi, 2024) *Proposed smoothed and data-adaptive copulas*

1. *Empirical copula with binomial smoothed survival margins and empirical beta copula as survival copula*

$$\begin{aligned} C_n^v(\mathbf{u}) &= \frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^X \left[\bar{F}_{u_1}^X \left(\frac{R_{i,1}-1}{n} \right), \dots, \bar{F}_{u_d}^X \left(\frac{R_{i,d}-1}{n} \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[\bar{B}_{n,u_1}^X (R_{i,1}-1), \dots, \bar{B}_{n,u_d}^X (R_{i,d}-1) \right], \quad \mathbf{u} \in [0, 1]^d \end{aligned}$$

where \bar{B}_{n,u_d}^X is the survival function of the binomial distribution: $\bar{B}_{n,u_d}^X \sim \mathbb{B}(n, u_d)$

2. *Empirical copula with beta-binomial smoothed survival margins ($\rho = 4$) and empirical beta copula as survival copula*

$$C_n^v(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[\overline{BB}_{u_1}^{\mathbf{X}}(R_{i,1} - 1), \dots, \overline{BB}_{u_d}^{\mathbf{X}}(R_{i,d} - 1) \right], \quad \mathbf{u} \in [0, 1]^d$$

where $\overline{BB}_{u_j}^{\mathbf{X}}$ is the survival function of the beta-binomial distribution.

3. *Empirical copula with beta smoothed survival margins ($\rho = 4$) and empirical beta copula as survival copula*

$$C_n^v(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[\bar{\mathcal{B}}_{u_1}^{\mathbf{X}}\left(\frac{R_{i,1} - 0.5}{n}\right), \dots, \bar{\mathcal{B}}_{u_d}^{\mathbf{X}}\left(\frac{R_{i,d} - 0.5}{n}\right) \right], \quad \mathbf{u} \in [0, 1]^d$$

where $\bar{\mathcal{B}}_{u_j}^{\mathbf{X}}$ is beta distributed.

3 Model specifications

We investigate the effectiveness of estimating upper and lower tail risk for several empirical copulas, with respect to the true copula.

3.1 Copulas and Parameters

3.1.1 Programming Language and Packages

The programming language used to do the simulation is the R language. The packages “`copula`” (Hofert et al., 2023), “`extraDistr`” (Wolodzko, 2023), “`RColorBrewer`” (Neuwirth, 2022) are utilised in the simulation.

3.1.2 True Copulas

To evaluate the accuracy of the estimation of the tails, we will consider empirical copulas that are fitted to samples generated by the true copula. Then, we will benchmark the accuracy of the class of empirical copulas with respect to the true copula. For convenience, the copulas will be listed below again.

The following copulas are used:

1. **Gumbel-Hougaard copula** (only upper-tail)

$$C_{\theta}^{GH}(\mathbf{u}) = \exp \left(- \left(\sum_{j=1}^d (-\ln u_j)^{\theta} \right)^{1/\theta} \right), \quad \theta \in [1, \infty), \quad \mathbf{u} \in [0, 1]^d$$

2. **Clayton copula** (only lower-tail)

$$C_{\theta}^C(\mathbf{u}) = \begin{cases} \max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\}^{-1/\theta}, & d = 2, \theta \in [-1, \infty) \setminus \{0\}, \mathbf{u} \in [0, 1]^2 \\ (1 - d + \sum_{j=1}^d u_j^{-\theta})^{-1/\theta}, & d \geq 3, \theta \in (0, \infty), \mathbf{u} \in [0, 1]^d \end{cases}$$

3. **Student-*t* copula with 4 degrees of freedom** (both upper- and lower-tail)

For correlation matrix Σ ,

$$\begin{aligned} C^t(\mathbf{u}) &= \int_0^{u_d} \cdots \int_0^{u_1} c^t(x_1, \dots, x_d) dx_1 \dots dx_d, \text{ where} \\ c^t(\mathbf{u}) &= \frac{\Gamma(\frac{v+d}{2})}{\Gamma(\frac{v}{2}) \sqrt{\det \Sigma}} \left(\frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} \right)^d \frac{(1 + \frac{\mathbf{x}^T \Sigma^{-1} \mathbf{x}}{v})^{-\frac{v+d}{2}}}{\prod_{j=1}^d (1 + \frac{x_j^2}{v})^{-\frac{v+1}{2}}}, \quad \mathbf{u} \in (0, 1)^d \end{aligned}$$

4. Gaussian Copula (both upper- and lower-tail)

For correlation matrix Σ ,

$$C^N(\mathbf{u}) = \int_{-\infty}^{\Phi^{-1}(u_d)} \cdots \int_{-\infty}^{\Phi^{-1}(u_1)} \frac{\exp(-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x})}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} dx_1 \dots dx_d, \quad \mathbf{u} \in (0, 1)^d$$

3.1.3 Empirical Copulas

We will evaluate 6 different types of empirical copulas that is derived from Deheuvels' classical empirical copula ([Deheuvels, 1979](#)). For convenience, the copulas will be listed below again.

1. Empirical copula

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{1} \left\{ \frac{R_{i,j}}{n} \leq u_j \right\}, \quad \mathbf{u} \in [0, 1]^d$$

2. Empirical beta copula

$$C_n^\beta(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{n, R_{i,j}}(u_j), \quad \mathbf{u} \in [0, 1]^d$$

3. Empirical checkerboard copula

$$C_n^{\boxplus}(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \min\{\max\{nu_j - R_{i,j} + 1, 0\}, 1\}, \quad \mathbf{u} \in [0, 1]^d$$

4. Empirical copula with binomial smoothed survival margins and empirical beta copula as survival copula

$$\begin{aligned} C_n^v(\mathbf{u}) &= \frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^{\mathbf{X}} \left[\bar{F}_{u_1}^{\mathbf{X}} \left(\frac{R_{i,1}-1}{n} \right), \dots, \bar{F}_{u_d}^{\mathbf{X}} \left(\frac{R_{i,d}-1}{n} \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[\bar{B}_{n,u_1}^{\mathbf{X}}(R_{i,1}-1), \dots, \bar{B}_{n,u_d}^{\mathbf{X}}(R_{i,d}-1) \right], \quad \mathbf{u} \in [0, 1]^d \end{aligned}$$

where $\bar{B}_{n,u_d}^{\mathbf{X}}$ is the survival function of the binomial distribution: $\bar{B}_{n,u_d}^{\mathbf{X}} \sim \mathbb{B}(n, u_d)$

5. Empirical copula with beta-binomial smoothed survival margins ($\rho = 4$) and empirical beta copula as survival copula

$$C_n^v(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[\bar{B}_{n,u_1}^{\mathbf{X}}(R_{i,1}-1), \dots, \bar{B}_{n,u_d}^{\mathbf{X}}(R_{i,d}-1) \right], \quad \mathbf{u} \in [0, 1]^d$$

where $\bar{B}_{n,u_j}^{\mathbf{X}}$ is the survival function of the beta-binomial distribution.

6. Empirical copula with beta smoothed survival margins ($\rho = 4$) and empirical beta copula as survival copula

$$C_n^w(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[\bar{\mathcal{B}}_{u_1}^{\mathbf{X}} \left(\frac{R_{i,1} - 0.5}{n} \right), \dots, \bar{\mathcal{B}}_{u_d}^{\mathbf{X}} \left(\frac{R_{i,d} - 0.5}{n} \right) \right], \quad \mathbf{u} \in [0, 1]^d$$

where $\bar{\mathcal{B}}_{u_j}^{\mathbf{X}}$ is beta distributed.

3.1.4 Dimensions

In this work, we mainly focus on smaller dimensions, which include $d \in \{2, 3, 4, 5\}$. This is due to the computationally prohibitive nature of computing the survival function of the target copula analytically. However, it can be expected that the characteristics of tail risk can be generalised to higher dimensions.

3.1.5 Kendall's Tau

To increase comparability between copulas and also between dimensions, Kendall's tau matching is useful to backfit the underlying copula parameters. We can also observe whether, under different values of Kendall's tau, there is a systematic underestimation / overestimation of tail risk with respect to the true copula. For simplicity, 7 values of Kendall's tau are considered: $\tau \in \{-0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75\}$. Kendall's tau matching is done via the `iTau` function from the R package “copula” (Hofert et al., 2023).

Kendall's tau matching is done instead of Spearman's Rho matching, as there is no analytical form of spearman's rho for most archimedean copulas and can only be done by numerical root finding.

3.1.6 Range of evaluation points

We only focus on upper and lower tail probabilities, hence we set the upper-tail evaluation points to values in $u \in [0.95, \dots, 1]$, and lower-tail evaluation points to values in $u \in [0, \dots, 0.05]$.

3.1.7 Simulation properties

We investigate how close and thus accurate the empirical copulas are to the true copula. We set a small number of data points (n) and replications (B): $n = 125$ and $B = 30$, where

both numbers are not too big nor too small, to prevent overfitting the empirical copula.

3.2 Sampling from the true copula

To evaluate whether an empirical copula overestimates or underestimates tails, we first have to generate samples from the true copula to fit empirical copulas. Below, the sampling methods from each type of (true) copula is provided:

3.2.1 Sampling Gaussian copulas

Algorithm 3.1 (Hofert et al., 2018) *Sampling Gaussian copulas*

1. Compute the cholesky factor A from Σ
2. Generate a standard normal vector $\mathbf{Z} = (Z_1, \dots, Z_d)$, where $Z_1, \dots, Z_d \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, 1)$.
3. Compute $\mathbf{X} = A\mathbf{Z}$
4. Return $\mathbf{U} = (\Phi(X_1), \dots, \Phi(X_d))$

3.2.2 Sampling Student- t copulas

Algorithm 3.2 (Hofert et al., 2018) *Sampling Student- t copulas*

1. Compute A from Σ
2. Generate standard normal vector $\mathbf{Z} = (Z_1, \dots, Z_d)$, where $Z_1, \dots, Z_d \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, 1)$.
3. Generate an Inverse Gaussian random variable $W \sim \text{IG}\left(\frac{v}{2}, \frac{v}{2}\right)$
4. Compute $\mathbf{X} = \sqrt{W}A\mathbf{Z}$
5. Return $\mathbf{U} = (t_v(X_1), \dots, t_v(X_d))$

3.2.3 Sampling Clayton and Gumbel copulas

Clayton and Gumbel copulas are Archimedean copulas with completely monotone generators and can easily be sampled.

It can be shown that (Hofert et al., 2018) for Archimedean Clayton copulas,

$$\mathbf{U} = \left(\psi\left(\frac{X_1}{V}\right), \dots, \psi\left(\frac{X_d}{V}\right) \right) \sim C$$

where X_1, \dots, X_d are independent $\text{Exp}(1)$ random variables, and independent from the frailty V . For Clayton copulas, $V \sim \Gamma(\theta^{-1}, 1)$ and $\psi(t) = (1+t)^{-1/\theta}$; for Gumbel-Hougaard copulas V is stable distributed and $\psi(t) = \exp(-t^{1/\theta})$. This easily allows us to sample (most) Archimedean copulas; this approach is known as the Marshall-Olkin algorithm (Marshall and Olkin, 1988, Hofert et al., 2018):

Algorithm 3.3 (Marshall & Olkin, 1988) *Sampling Archimedean copulas with the Marshall-Olkin Algorithm*

1. Generate $V \sim F = LS^{-1}[\psi]$
2. Generate $X_1, \dots, X_d \stackrel{\text{ind.}}{\sim} \text{Exp}(1)$
3. Return $\mathbf{U} = \left(\psi\left(\frac{X_1}{V}\right), \dots, \psi\left(\frac{X_d}{V}\right) \right)$.

3.3 Simulating exceedance and cumulative probabilities

The following algorithms describes how we assess exceedance and cumulative probabilities to understand the extent of the underestimation / overestimation of tail risk of empirical copulas.

First we provide an algorithm on the fitting of the survival function for different empirical copulas, then we provide the algorithm for computing exceedance and cumulative probabilities.

Algorithm 1 Survival function of the empirical copula

```
procedure SURVIVALFUNCTIONEC( $\mathbf{u}$ ,  $\mathbf{U}$ )
    n ← nrow( $\mathbf{U}$ )                                ▷ nrow refers to number of rows
    d ← ncol( $\mathbf{U}$ )                               ▷ ncol refers to number of columns
    for i in 1:n do
        for j in 1:d do
             $\mathbf{R}_{i,j} \leftarrow \sum_{k=1}^n \mathbb{1}\{U_{i,j} \leq U_{k,j}\}$ 
        end for
    end for
     $\hat{\mathbf{U}} \leftarrow \frac{\mathbf{R}}{n+1}$                   ▷ As  $\mathbf{U}$  is unknown, pseudo-observations  $\hat{\mathbf{U}}$  are created
     $\bar{H}(\mathbf{u}) \leftarrow \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\hat{\mathbf{U}}_i \geq \mathbf{u}_i\}$           ▷ Empirical survival function
    return  $\bar{H}(\mathbf{u})$ 
end procedure
```

Algorithm 2 Survival function of the empirical beta copula

```
procedure SURVIVALFUNCTIONEBC( $\mathbf{u}$ ,  $\mathbf{U}$ )
    n ← nrow( $\mathbf{U}$ )
    d ← ncolumn( $\mathbf{U}$ )
    for i in 1:n do
        for j in 1:d do
             $\mathbf{R}_{i,j} \leftarrow \sum_{k=1}^n \mathbb{1}\{U_{i,j} \leq U_{k,j}\}$ 
        end for
    end for
     $C_n^\beta(\mathbf{u}) \leftarrow \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d P(\mathcal{B}(R_{i,j}, n+1 - R_{i,j}) \geq u_{i,j})$ 
    return  $C_n^\beta(\mathbf{u})$ 
end procedure
```

Algorithm 3 Survival function of the empirical checkerboard copula

```

procedure CUMULATIVEFUNCTIONECC( $\mathbf{u}$ ,  $\hat{\mathbf{U}}$ ) ▷  $\hat{\mathbf{U}}$  instead of  $\mathbf{U}$ 
     $\mathbf{R} \leftarrow \hat{\mathbf{U}} \times (n + 1)$ 
     $C_n^{\boxplus}(\mathbf{u}) \leftarrow \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \min\{\max\{nu_{i,j} + 1 - R_{i,j}\}, 1\}$ 
    return  $C_n^{\boxplus}(\mathbf{u})$ 
end procedure

procedure SURVIVALFUNCTIONECC( $\mathbf{u}$ ,  $\mathbf{U}$ )
     $n \leftarrow \text{nrow}(\mathbf{U})$ 
     $d \leftarrow \text{ncol}(\mathbf{U})$ 
    for i in 1:n do
        for j in 1:d do
             $\mathbf{R}_{i,j} \leftarrow \sum_{k=1}^n \mathbb{1}\{U_{i,j} \leq U_{k,j}\}$ 
        end for
    end for
     $\hat{\mathbf{U}} \leftarrow \frac{\mathbf{R}}{n+1}$  ▷ As  $\mathbf{U}$  is unknown, pseudo-observations  $\hat{\mathbf{U}}$  are created
     $\mathbf{u}^* \leftarrow \mathbf{1} - \hat{\mathbf{u}}$  ▷  $\hat{\mathbf{u}}$  are componentwise reflections of  $\mathbf{u}$ 
     $\Omega \leftarrow \text{PowerSet}(1, \dots, d)$ 
     $\bar{\mathbf{H}} \leftarrow \mathbf{0}$ 
    for each subset  $\Omega^*$  of  $\Omega$  do
        if length( $\Omega^*$ ) = 0 then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + \mathbf{1}$ 
        else if length( $\Omega^*$ ) = 1 then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} - (\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*)$  ▷ deduct  $\mathbf{1} - \mathbf{u}_j$  (1-dimensional) from  $\bar{\mathbf{H}}$ 
        else
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + (-1)^{\text{length}(\Omega^*)} \text{CumulativeFunctionECC}(\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*, \hat{\mathbf{U}}_{j \in \Omega^*}^*)$ 
        end if
    end for
    return  $\bar{\mathbf{H}}$ 
end procedure

```

Algorithm 4 Survival function of EBC-adapted empirical copula with binomial survival margins

```

procedure CUMULATIVEFUNCTIONEBINOM( $\mathbf{u}$ ,  $\hat{\mathbf{U}}$ ) ▷  $\hat{\mathbf{U}}$  instead of  $\mathbf{U}$ 
     $\mathbf{R} \leftarrow \hat{\mathbf{U}} \times (n + 1)$ 
     $C_n^v(\mathbf{u}) \leftarrow \frac{1}{n} \sum_{i=1}^n C_n^\beta [P(\mathbb{B}(n, u_{i,1}) \geq (R_{i,1} - 1)), \dots, P(\mathbb{B}(n, u_{i,d}) \geq (R_{i,d} - 1))]$ 
    return  $C_n^v(\mathbf{u})$ 
end procedure

procedure SURVIVALFUNCTIONEBINOM( $\mathbf{u}$ ,  $\mathbf{U}$ )
     $n \leftarrow \text{nrow}(\mathbf{U})$ 
     $d \leftarrow \text{ncol}(\mathbf{U})$ 
    for i in 1:n do
        for j in 1:d do
             $\mathbf{R}_{i,j} \leftarrow \sum_{k=1}^n \mathbb{1}\{U_{i,j} \leq U_{k,j}\}$ 
        end for
    end for
     $\hat{\mathbf{U}} \leftarrow \frac{\mathbf{R}}{n+1}$  ▷ As  $\mathbf{U}$  is unknown, pseudo-observations  $\hat{\mathbf{U}}$  are created
     $\mathbf{u}^* \leftarrow \mathbf{1} - \hat{\mathbf{u}}$  ▷  $\hat{\mathbf{u}}$  are componentwise reflections of  $\mathbf{u}$ 
     $\Omega \leftarrow \text{PowerSet}(1, \dots, d)$ 
     $\bar{\mathbf{H}} \leftarrow \mathbf{0}$ 
    for each subset  $\Omega^*$  of  $\Omega$  do
        if  $\text{length}(\Omega^*) = 0$  then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + \mathbf{1}$ 
        else if  $\text{length}(\Omega^*) = 1$  then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} - (\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*)$  ▷ deduct  $\mathbf{1} - \mathbf{u}_j$  (1-dimensional) from  $\bar{\mathbf{H}}$ 
        else
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + (-1)^{\text{length}(\Omega^*)} \text{CumulativeFunctionEBinom}(\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*, \hat{\mathbf{U}}_{j \in \Omega^*}^*)$ 
        end if
    end for
    return  $\bar{\mathbf{H}}$ 
end procedure

```

Algorithm 5 Survival function of EBC-adapted empirical copula with beta-binomial survival margins of $\rho = 4$

```

procedure CUMULATIVEFUNCTIONEBB4( $\mathbf{u}$ ,  $\hat{\mathbf{U}}$ ) ▷  $\hat{\mathbf{U}}$  instead of  $\mathbf{U}$ 
     $\mathbf{R} \leftarrow \hat{\mathbf{U}} \times (n + 1)$ 
     $C_n^v(\mathbf{u}) \leftarrow \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[ P \left( BB(n, u_{i,1}, \rho = 4) \geq \frac{R_{i,1}-1}{n} \right), \dots, P \left( BB(n, u_{i,d}, \rho = 4) \geq \frac{R_{i,d}-1}{n} \right) \right]$ 
    return  $C_n^v(\mathbf{u})$ 
end procedure

procedure SURVIVALFUNCTIONEBB4( $\mathbf{u}$ ,  $\mathbf{U}$ )
     $n \leftarrow \text{nrow}(\mathbf{U})$ 
     $d \leftarrow \text{ncol}(\mathbf{U})$ 
    for i in 1:n do
        for j in 1:d do
             $\mathbf{R}_{i,j} \leftarrow \sum_{k=1}^n \mathbb{1}\{U_{i,j} \leq U_{k,j}\}$ 
        end for
    end for
     $\hat{\mathbf{U}} \leftarrow \frac{\mathbf{R}}{n+1}$  ▷ As  $\mathbf{U}$  is unknown, pseudo-observations  $\hat{\mathbf{U}}$  are created
     $\mathbf{u}^* \leftarrow \mathbf{1} - \hat{\mathbf{u}}$  ▷  $\hat{\mathbf{u}}$  are componentwise reflections of  $\mathbf{u}$ 
     $\Omega \leftarrow \text{PowerSet}(1, \dots, d)$ 
     $\bar{\mathbf{H}} \leftarrow \mathbf{0}$ 
    for each subset  $\Omega^*$  of  $\Omega$  do
        if  $\text{length}(\Omega^*) = 0$  then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + \mathbf{1}$ 
        else if  $\text{length}(\Omega^*) = 1$  then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} - (\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*)$  ▷ deduct  $\mathbf{1} - \mathbf{u}_j$  (1-dimensional) from  $\bar{\mathbf{H}}$ 
        else
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + (-1)^{\text{length}(\Omega^*)} \text{CumulativeFunctionEBB4}(\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*, \hat{\mathbf{U}}_{j \in \Omega^*}^*)$ 
        end if
    end for
    return  $\bar{\mathbf{H}}$ 
end procedure

```

Algorithm 6 Survival function of EBC-adapted empirical copula with beta survival margins of $\rho = 4$

```

procedure CUMULATIVEFUNCTIONEBETA4( $\mathbf{u}$ ,  $\hat{\mathbf{U}}$ ) ▷  $\hat{\mathbf{U}}$  instead of  $\mathbf{U}$ 
     $\mathbf{R} \leftarrow \hat{\mathbf{U}} \times (n + 1)$ 
     $C_n^v(\mathbf{u}) \leftarrow \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[ P \left( \mathcal{B}(n, u_{i,1}, \rho = 4) \geq \frac{R_{i,1}-0.5}{n} \right), \dots, P \left( \mathcal{B}(n, u_{i,d}, \rho = 4) \geq \frac{R_{i,d}-0.5}{n} \right) \right]$ 
    return  $C_n^v(\mathbf{u})$ 
end procedure

procedure SURVIVALFUNCTIONEBETA4( $\mathbf{u}$ ,  $\mathbf{U}$ )
     $n \leftarrow \text{nrow}(\mathbf{U})$ 
     $d \leftarrow \text{ncol}(\mathbf{U})$ 
    for i in 1:n do
        for j in 1:d do
             $\mathbf{R}_{i,j} \leftarrow \sum_{k=1}^n \mathbb{1}\{U_{i,j} \leq U_{k,j}\}$ 
        end for
    end for
     $\hat{\mathbf{U}} \leftarrow \frac{\mathbf{R}}{n+1}$  ▷ As  $\mathbf{U}$  is unknown, pseudo-observations  $\hat{\mathbf{U}}$  are created
     $\mathbf{u}^* \leftarrow \mathbf{1} - \hat{\mathbf{u}}$  ▷  $\hat{\mathbf{u}}$  are componentwise reflections of  $\mathbf{u}$ 
     $\Omega \leftarrow \text{PowerSet}(1, \dots, d)$ 
     $\bar{\mathbf{H}} \leftarrow \mathbf{0}$ 
    for each subset  $\Omega^*$  of  $\Omega$  do
        if  $\text{length}(\Omega^*) = 0$  then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + \mathbf{1}$ 
        else if  $\text{length}(\Omega^*) = 1$  then
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} - (\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*)$  ▷ deduct  $\mathbf{1} - \mathbf{u}_j$  (1-dimensional) from  $\bar{\mathbf{H}}$ 
        else
             $\bar{\mathbf{H}} \leftarrow \bar{\mathbf{H}} + (-1)^{\text{length}(\Omega^*)} \text{CumulativeFunctionEBeta4}(\mathbf{1} - \mathbf{u}_{j \in \Omega^*}^*, \hat{\mathbf{U}}_{j \in \Omega^*}^*)$ 
        end if
    end for
    return  $\bar{\mathbf{H}}$ 
end procedure

```

The algorithm for exceedance and cumulative probabilities for empirical copulas are shown below:

Algorithm 7 Exceedance probabilities of empirical copulas

```

procedure SURVIVALPROBABILITY(copula)            $\triangleright$  copula is exogenously inputted.

     $n \leftarrow 125$ 
     $B \leftarrow 30$ 
     $d \leftarrow \text{dim}(\text{copula})$ 
     $\mathbf{u\_mat} \leftarrow \mathbf{u} = (\mathbf{u}_{\cdot,1}, \mathbf{u}_{\cdot,2}, \dots, \mathbf{u}_{\cdot,d})$        $\triangleright \mathbf{u}_{\cdot,i}$ : identical evaluation vector  $(0.95, \dots, 1)$ 
    true_res_mat_s  $\leftarrow \bar{C}(\mathbf{1} - \mathbf{u\_mat})$        $\triangleright$  obtain exceedance probability for true copula

    for  $b$  in  $1:B$  do

         $\mathbf{U}^{n,d} \leftarrow$  sampling from copula           $\triangleright$  See Algorithm 3.1, 3.2, 3.3
        ec_res_mat_s( $b$ )  $\leftarrow$  SurvivalFunctionEC( $\mathbf{u\_mat}$ ,  $\mathbf{U}^{n,d}$ )
        ebc_res_mat_s( $b$ )  $\leftarrow$  SurvivalFunctionEBC( $\mathbf{u\_mat}$ ,  $\mathbf{U}^{n,d}$ )
        ecc_res_mat_s( $b$ )  $\leftarrow$  SurvivalFunctionECC( $\mathbf{u\_mat}$ ,  $\mathbf{U}^{n,d}$ )
        ebinom_res_mat_s( $b$ )  $\leftarrow$  SurvivalFunctionEBinom( $\mathbf{u\_mat}$ ,  $\mathbf{U}^{n,d}$ )
        ebb4_res_mat_s( $b$ )  $\leftarrow$  SurvivalFunctionEBB4( $\mathbf{u\_mat}$ ,  $\mathbf{U}^{n,d}$ )
        ebta4_res_mat_s( $b$ )  $\leftarrow$  SurvivalFunctionEBeta4( $\mathbf{u\_mat}$ ,  $\mathbf{U}^{n,d}$ )

    end for

    mean_ec_res_mat_s  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ec\_res\_mat\_s}(i)$ 
    mean_ebc_res_mat_s  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebc\_res\_mat\_s}(i)$ 
    mean_ecc_res_mat_s  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ecc\_res\_mat\_s}(i)$ 
    mean_ebinom_res_mat_s  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebinom\_res\_mat\_s}(i)$ 
    mean_ebb4_res_mat_s  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebb4\_res\_mat\_s}(i)$ 
    mean_ebta4_res_mat_s  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebeta4\_res\_mat\_s}(i)$ 

    CI_ec_res_mat_s  $\leftarrow$  2.5% and 97.5% quantile of ec_res_mat_s
    CI_ebc_res_mat_s  $\leftarrow$  2.5% and 97.5% quantile of ebc_res_mat_s
    CI_ecc_res_mat_s  $\leftarrow$  2.5% and 97.5% quantile of ecc_res_mat_s
    CI_lebinom_res_mat_s  $\leftarrow$  2.5% and 97.5% quantile of ebinom_res_mat_s
    CI_lebb4_res_mat_s  $\leftarrow$  2.5% and 97.5% quantile of ebb4_res_mat_s
    CI_lebta4_res_mat_s  $\leftarrow$  2.5% and 97.5% quantile of ebta4_res_mat_s

    return true_res_mat_s, mean_ec_res_mat_s, mean_ebc_res_mat_s, mean_ecc_res_mat_s,
    mean_ebb4_res_mat_s, mean_ebta4_res_mat_s, mean_ebinom_res_mat_s, CI_ec_res_mat_s,
    CI_ebc_res_mat_s, CI_ecc_res_mat_s, CI_lebb4_res_mat_s, CI_lebta4_res_mat_s,
    CI_lebinom_res_mat_s

end procedure

```

Algorithm 8 Cumulative probabilities of empirical copulas

```

procedure CUMULATIVEPROBABILITY(copula)       $\triangleright$  copula is exogenously inputted.

     $n \leftarrow 125$ 
     $B \leftarrow 30$ 
     $d \leftarrow \text{dim}(\text{copula})$ 
     $\mathbf{u\_mat} \leftarrow \mathbf{u} = (\mathbf{u}_{\cdot,1}, \mathbf{u}_{\cdot,2}, \dots, \mathbf{u}_{\cdot,d})$        $\triangleright \mathbf{u}_i$ : identical evaluation vector  $(0, \dots, 0.05)$ 
    true_res_mat_c  $\leftarrow C(\mathbf{u\_mat})$        $\triangleright$  obtain cumulative probability for true copula

    for  $b$  in  $1:B$  do

         $\mathbf{U}^{n,d} \leftarrow$  sampling from copula           $\triangleright$  See Algorithm 3.1, 3.2, 3.3
        ec_res_mat_c( $b$ )  $\leftarrow \frac{1}{n} \sum_{i=1}^j \mathbf{1}\{\hat{\mathbf{U}}_i \leq \mathbf{u\_mat}_{i,\cdot}\}$        $\triangleright \hat{\mathbf{U}}$  and  $\mathbf{R}$  defined previously
        ebc_res_mat_c( $b$ )  $\leftarrow \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d P(\mathbb{B}(R_{i,j}, n+1 - R_{i,j}) \leq u_{mat_{i,j}}))$ 
        ecc_res_mat_c( $b$ )  $\leftarrow \text{CumulativeFunctionECC}(\mathbf{u\_mat}, \mathbf{U}^{n,d})$ 
        ebinom_res_mat_c( $b$ )  $\leftarrow \text{CumulativeFunctionEBinom}(\mathbf{u\_mat}, \mathbf{U}^{n,d})$ 
        ebb4_res_mat_c( $b$ )  $\leftarrow \text{CumulativeFunctionEBB4}(\mathbf{u\_mat}, \mathbf{U}^{n,d})$ 
        ebta4_res_mat_c( $b$ )  $\leftarrow \text{CumulativeFunctionEBeta4}(\mathbf{u\_mat}, \mathbf{U}^{n,d})$ 

    end for

    mean_ec_res_mat_c  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ec\_res\_mat\_c}(i)$ 
    mean_ebc_res_mat_c  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebc\_res\_mat\_c}(i)$ 
    mean_ecc_res_mat_c  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ecc\_res\_mat\_c}(i)$ 
    mean_ebinom_res_mat_c  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebinom\_res\_mat\_c}(i)$ 
    mean_ebb4_res_mat_c  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebb4\_res\_mat\_c}(i)$ 
    mean_ebta4_res_mat_c  $\leftarrow \frac{1}{b} \sum_{i=1}^b \text{ebeta4\_res\_mat\_c}(i)$ 
    CI_ec_res_mat_c  $\leftarrow 2.5\%$  and  $97.5\%$  quantile of ec_res_mat_c
    CI_ebc_res_mat_c  $\leftarrow 2.5\%$  and  $97.5\%$  quantile of ebc_res_mat_c
    CI_ecc_res_mat_c  $\leftarrow 2.5\%$  and  $97.5\%$  quantile of ecc_res_mat_c
    CI_ebinom_res_mat_c  $\leftarrow 2.5\%$  and  $97.5\%$  quantile of ebinom_res_mat_c
    CI_ebb4_res_mat_c  $\leftarrow 2.5\%$  and  $97.5\%$  quantile of ebb4_res_mat_c
    CI_ebta4_res_mat_c  $\leftarrow 2.5\%$  and  $97.5\%$  quantile of ebta4_res_mat_c

    return true_res_mat_c, mean_ec_res_mat_c, mean_ebc_res_mat_c, mean_ecc_res_mat_c,
    mean_ebb4_res_mat_c, mean_ebta4_res_mat_c, mean_ebinom_res_mat_c, CI_ec_res_mat_c,
    CI_ebc_res_mat_c, CI_ecc_res_mat_c, CI_ebb4_res_mat_c, CI_ebta4_res_mat_c,
    CI_ebinom_res_mat_c

end procedure

```

3.4 Model Evaluation: Cramer-von-Mises (CvM) Statistic

To understand how close the nonparametric empirical copulas are to the true copula, the Cramer-von-Mises goodness-of-fit statistic S_n^{CvM} is used as it has higher performance on empirical processes compared to alternative goodness-of-fit statistics such as the Kolmogorov-Smirnov statistics (Genest et al., 2009).

The Cramer-von-Mises statistic is defined as below:

Definition 3.4.1 (Hofert et al., 2018, Genest et al., 2009) **Cramer-von-Mises statistic**

The Cramer-von-Mises statistic is:

$$S_n^{CvM} = \int_{[0,1]^d} n(C_n(\mathbf{u}) - C(\mathbf{u}))^2 dC_n(\mathbf{u}).$$

It can also be rewritten as a sum, and this form is extensively used in this work.

Definition 3.4.2 (Hofert et al., 2018, Genest et al., 2009) **Discretized Cramer-von-Mises statistic**

The Cramer-von-Mises statistic is:

$$S_n^{CvM} = \sum_{i=1}^n (C_n(\mathbf{u}_i) - C(\mathbf{u}_i))^2$$

Hence the algorithm for computing the Cramer-von-Mises Statistic is proposed below:

Algorithm 9 Discretized Cramer-von-Mises Statistic

```

procedure CvM(true_res_mat, mean_ec_res_mat, n)
    CvM  $\leftarrow \sum_{i=1}^n (\text{mean\_ec\_res\_mat}_i - \text{true\_res\_mat}_i)^2$ 
    return CvM
end procedure

```

4 Results

4.1 Preparation

Before plotting, a colour code table is given with reference to the large number of lines of each graph, to prevent convolution.

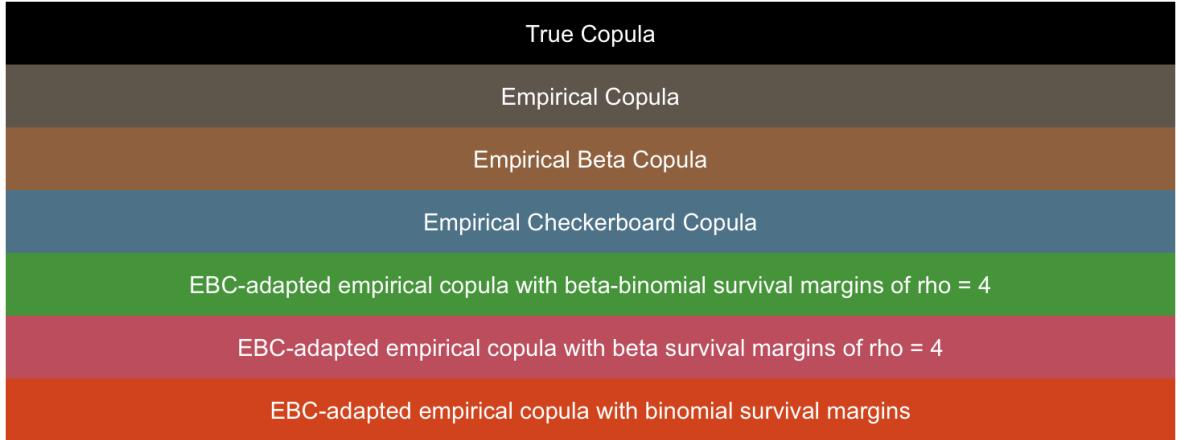


Figure 1: Colour code mapping

Notice that some combinations of copula, Kendall's tau and dimensions are not plotted.

For instance:

Recall that the domain of the parameter of Clayton and Gumbel-Hougaard copula is $(0, \infty)$ and $[1, \infty)$ respectively. Consider the Kendall's tau for Gumbel-Hougaard and Clayton copulas:

$$\begin{aligned}\tau_C &= \frac{\theta}{\theta + 2}, \quad \theta \in (0, \infty) \implies \tau_C \in (0, \infty) \\ \tau_G &= \frac{\theta - 1}{\theta}, \quad \theta \in [1, \infty) \implies \tau_G \in [0, \infty)\end{aligned}$$

Hence, we will only consider $\tau_C \in 0.25, 0.5, 0.75$ for Clayton and Gumbel copulas.

Moreover, the rotated copula for normal and student- t copulas faces Monte Carlo errors for $d \geq 3$ ([Hofert et al., 2018](#)), hence we will only consider normal and student- t copulas at $d = 2$.

4.2 Plots: exceedance probabilities vs u (evaluation points)

We first plot the (upper-tailed) evaluation points against the exceedance probabilities for each empirical copula. The evaluation points are on the x-axis and the exceedance probabilities are on the y-axis. The solid line is the mean of the exceedance probabilities corresponding to the evaluation point, and the dotted line is the confidence intervals of the exceedance probabilities corresponding to the evaluation point.

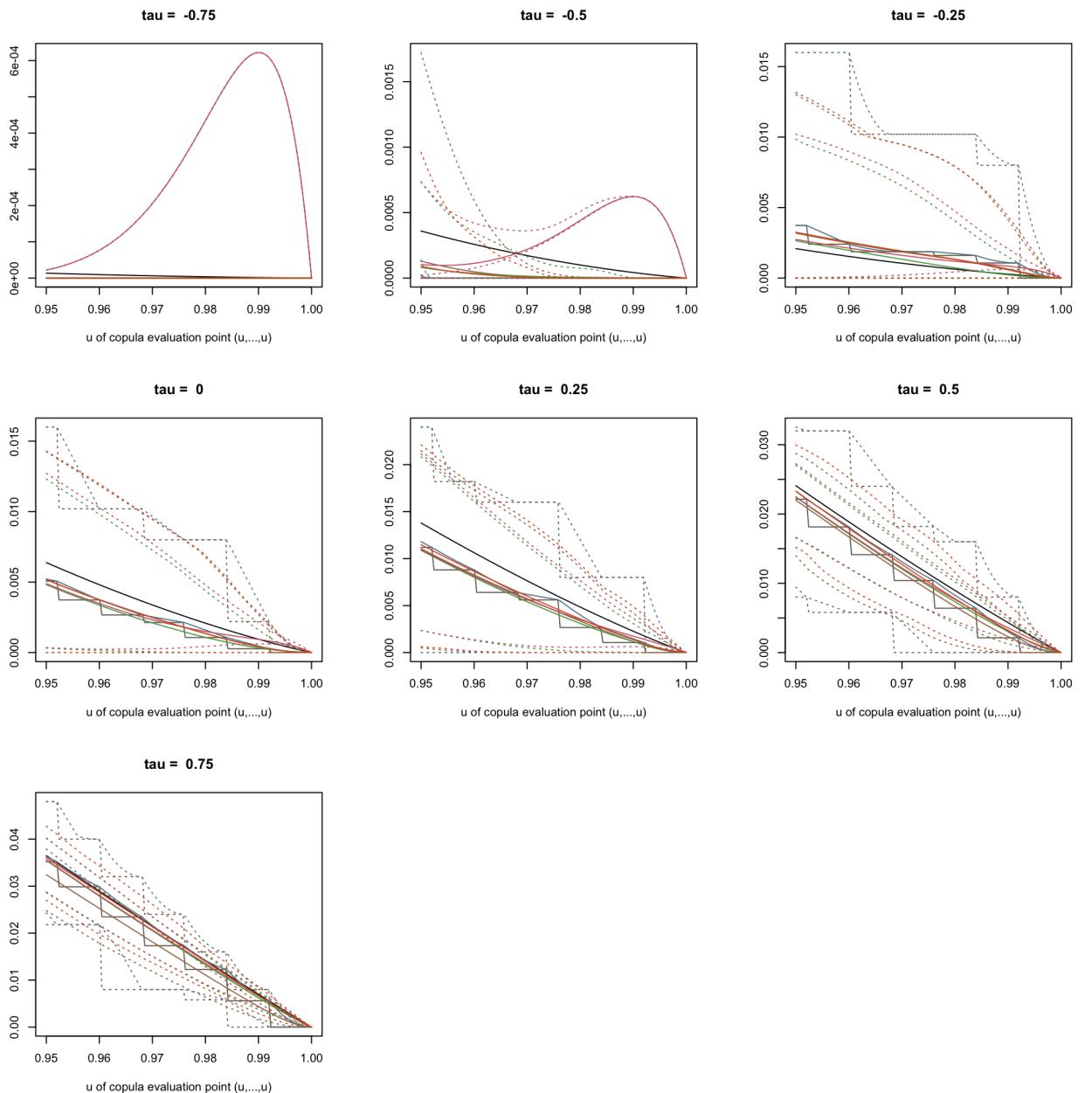


Figure 2: Student- t copula with $d = 2$

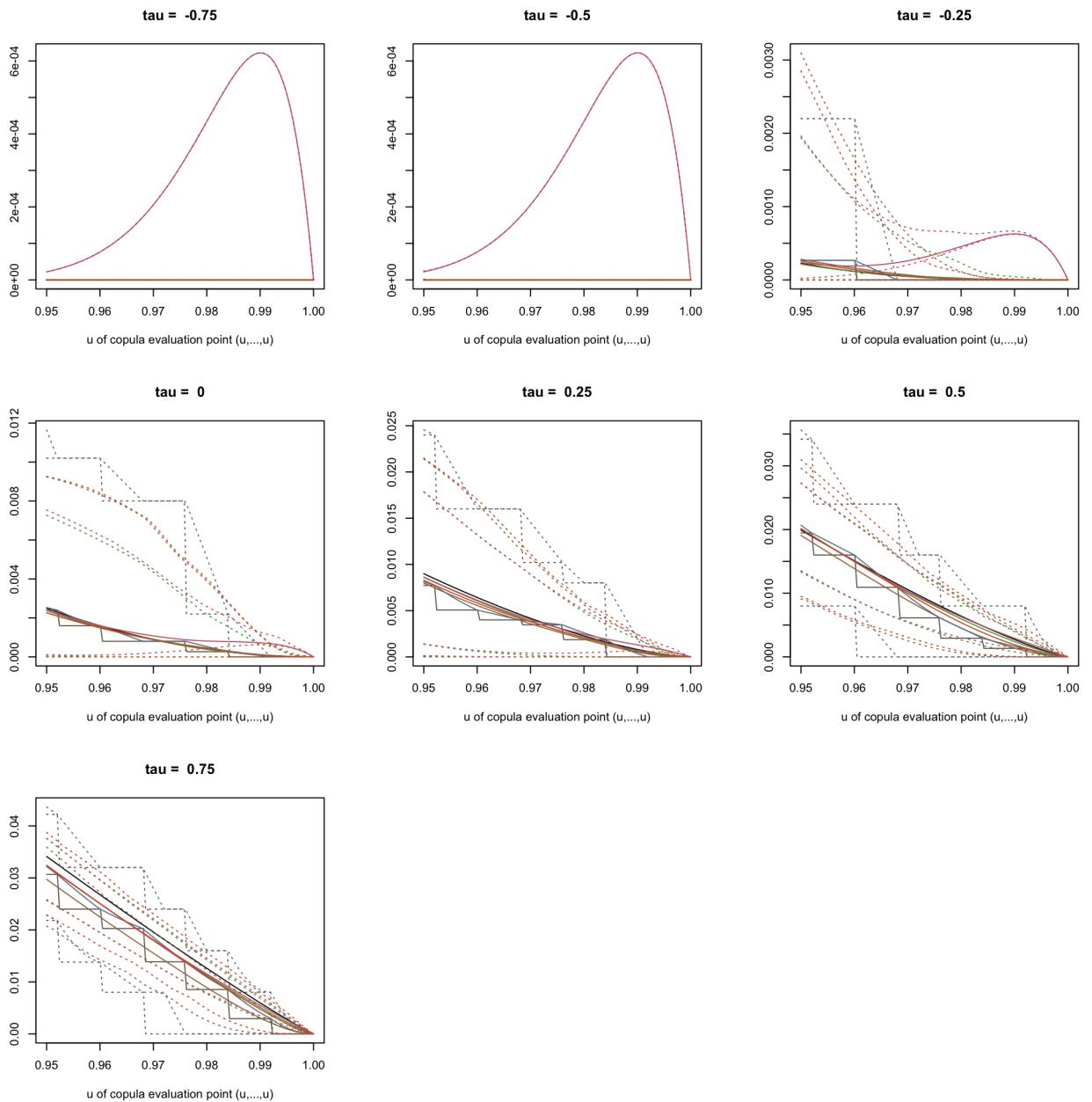


Figure 3: Gaussian copula with $d = 2$

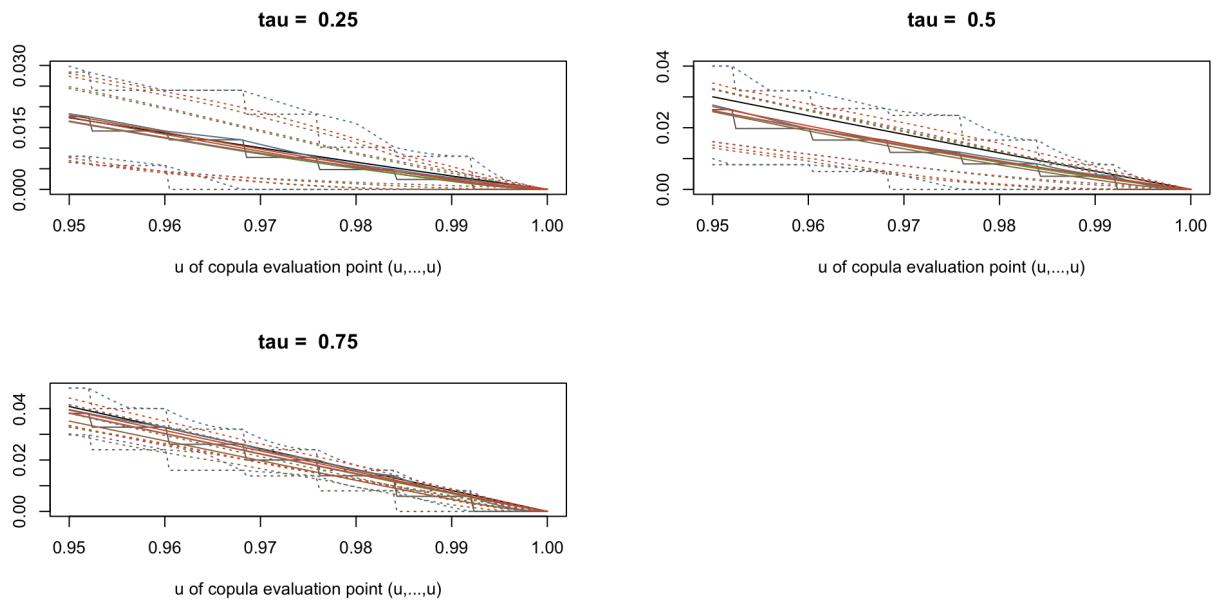


Figure 4: Gumbel-Hougaard copula with $d = 2$

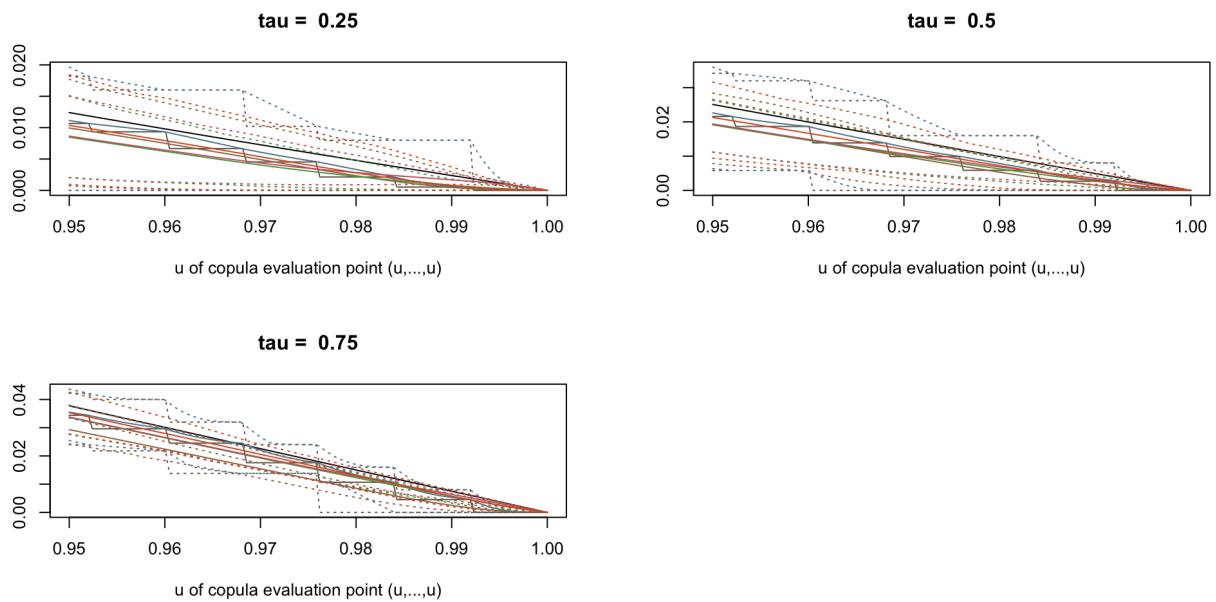


Figure 5: Gumbel-Hougaard copula with $d = 3$

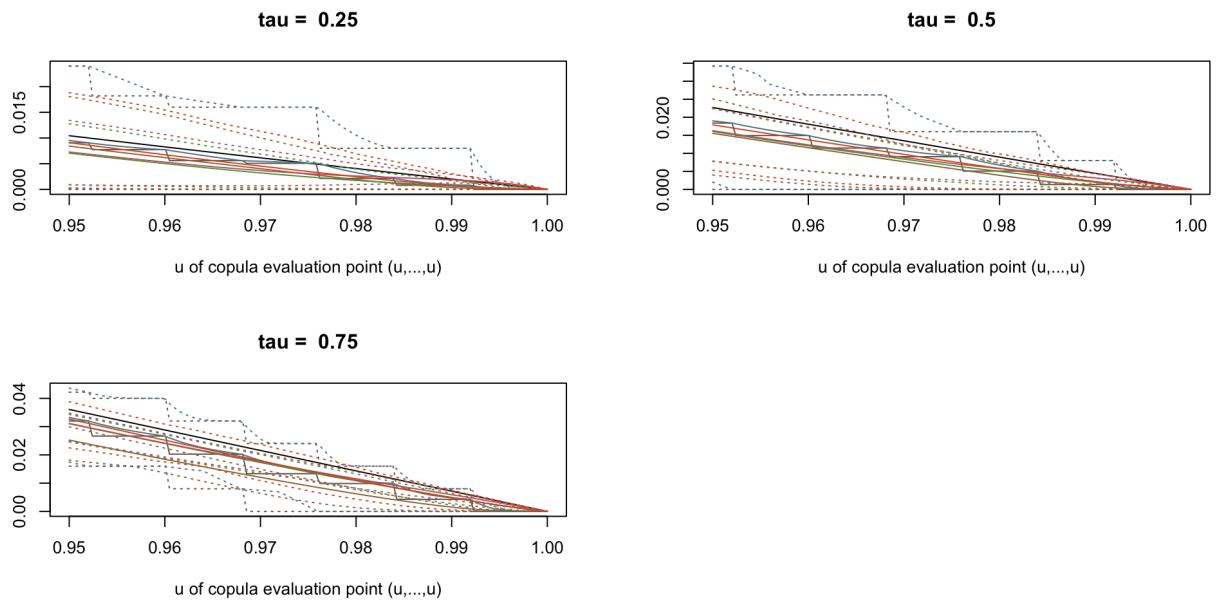


Figure 6: Gumbel-Hougaard copula with $d = 4$

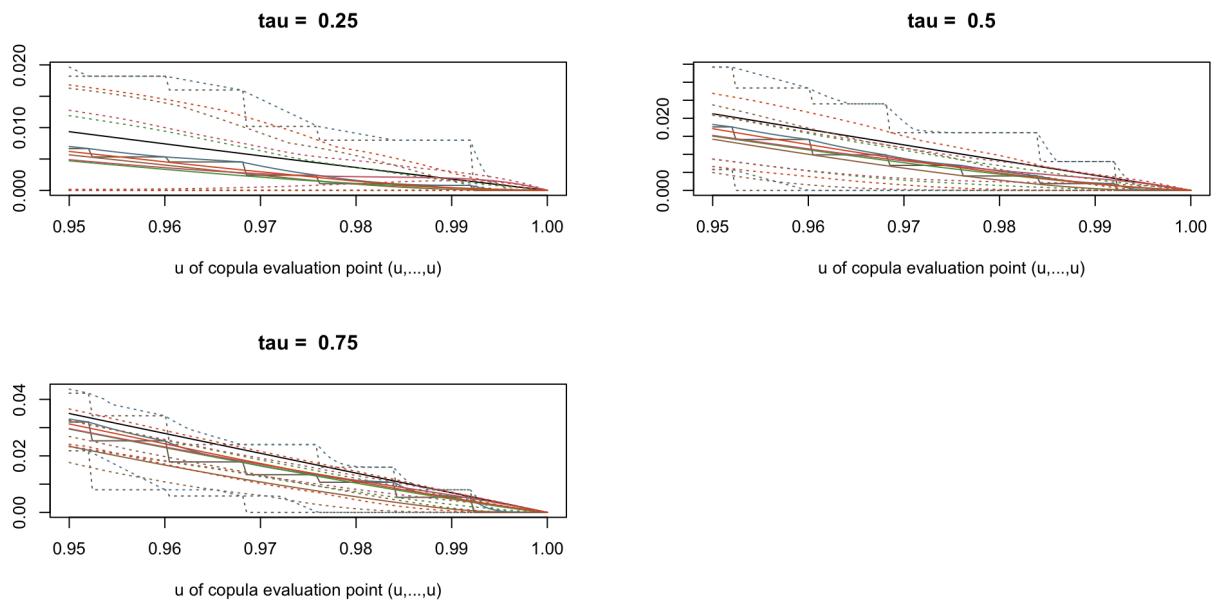


Figure 7: Gumbel-Hougaard copula with $d = 5$

4.3 CvM: Exceedance probabilities

To better understand the proximity of the exceedance probabilities from the class of empirical copula, with respect to the true copula, we then compute the Cramer-von-Mises statistics for all combinations of empirical copulas, Kendall's Tau and dimensions.

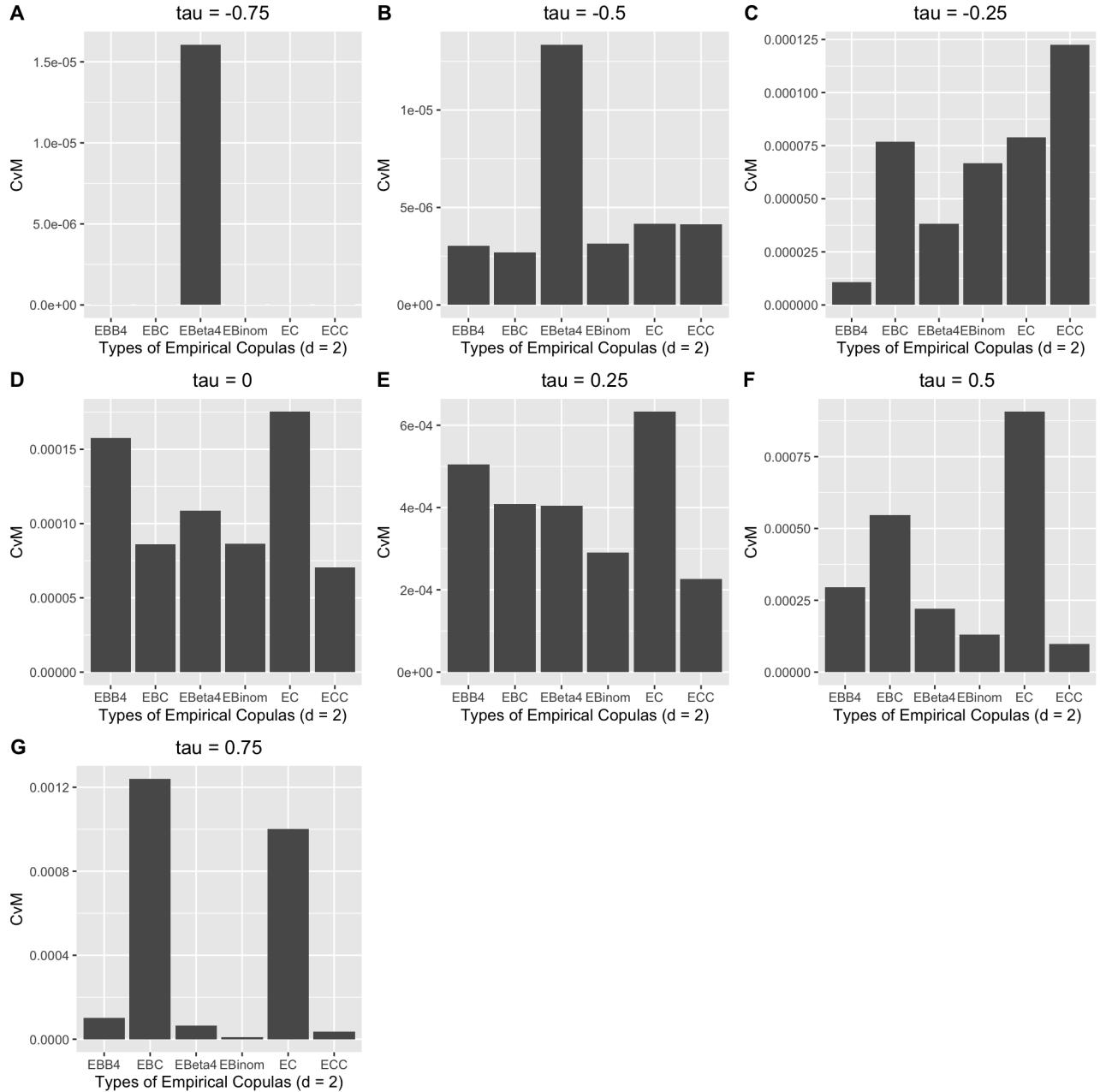


Figure 8: Student- t copula with $d = 2$

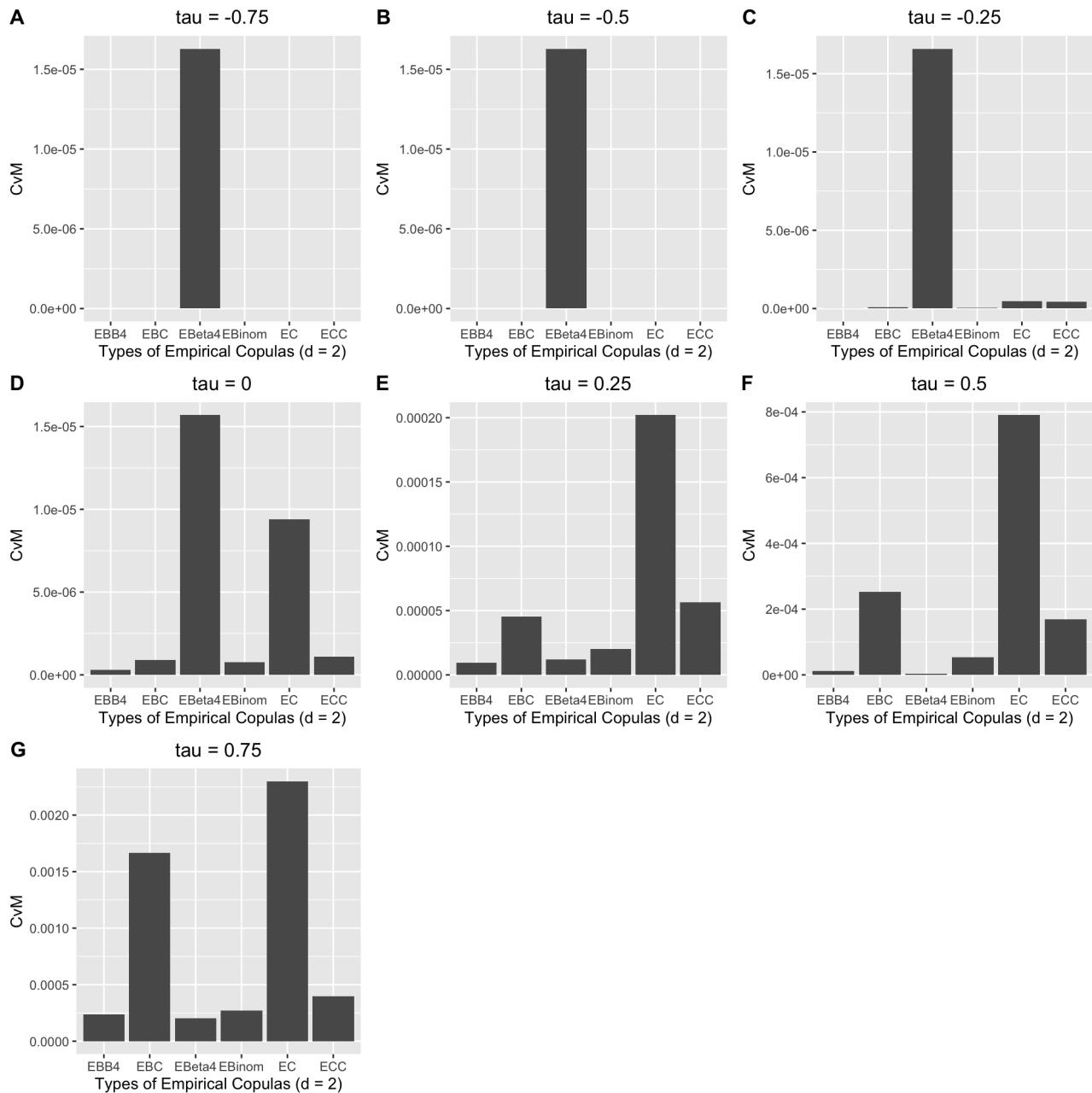


Figure 9: Gaussian copula with $d = 2$

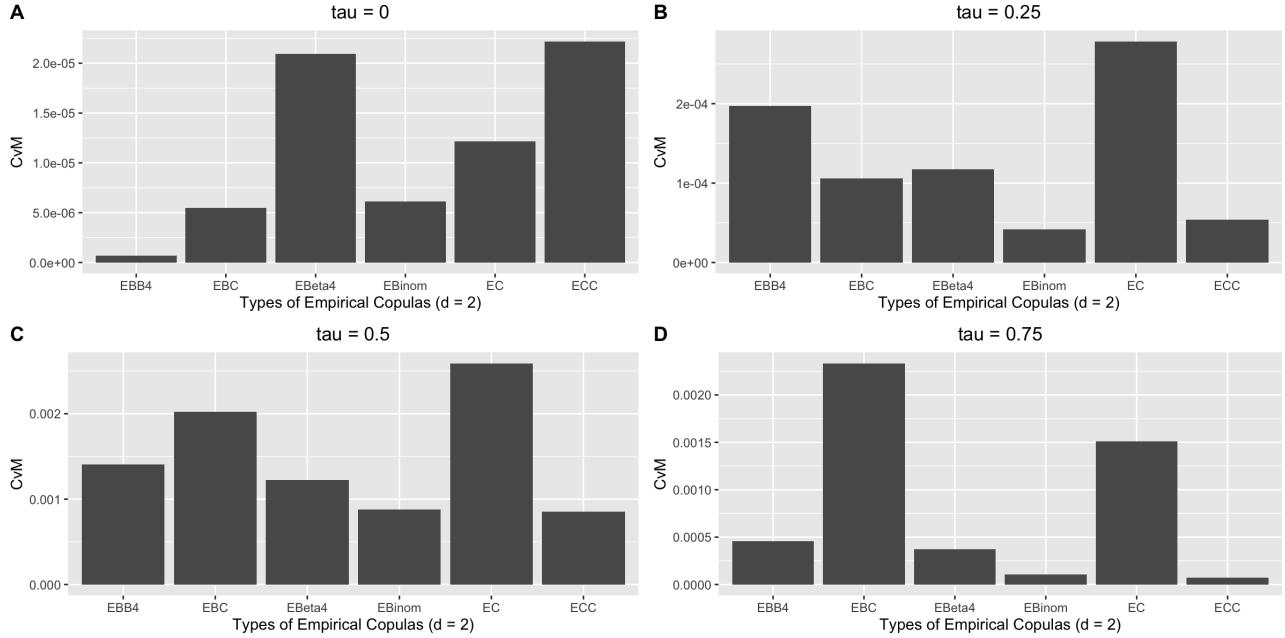


Figure 10: Gumbel-Hougaard copula with $d = 2$

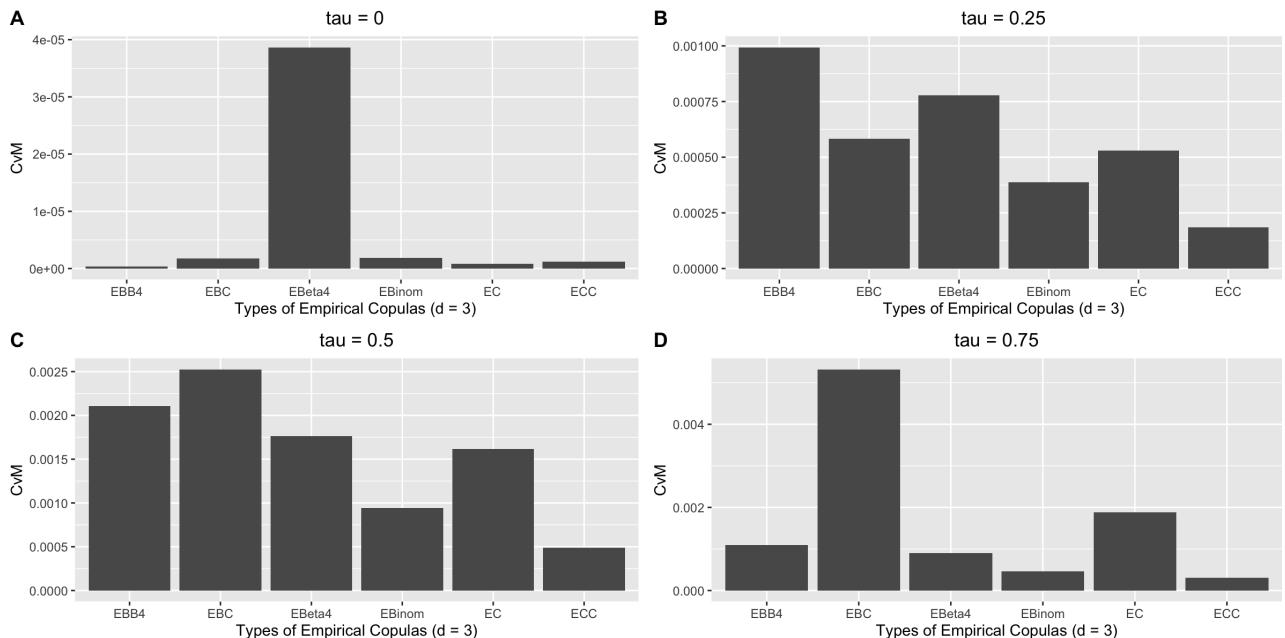


Figure 11: Gumbel-Hougaard copula with $d = 3$

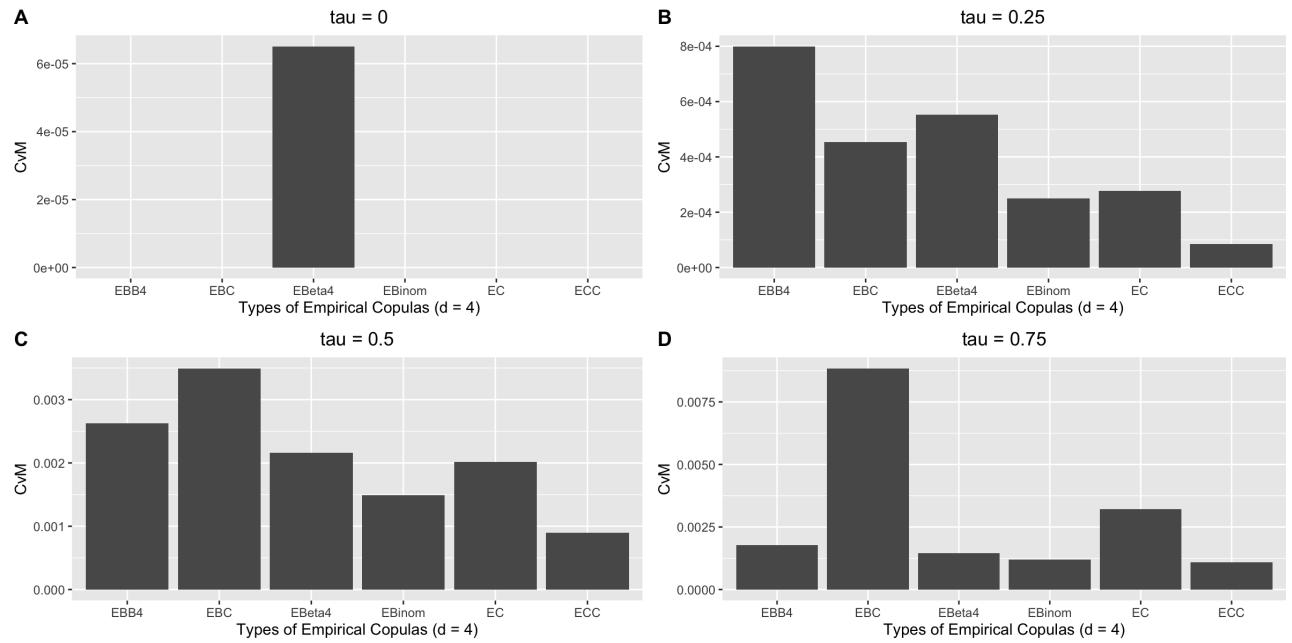


Figure 12: Gumbel-Hougaard copula with $d = 4$

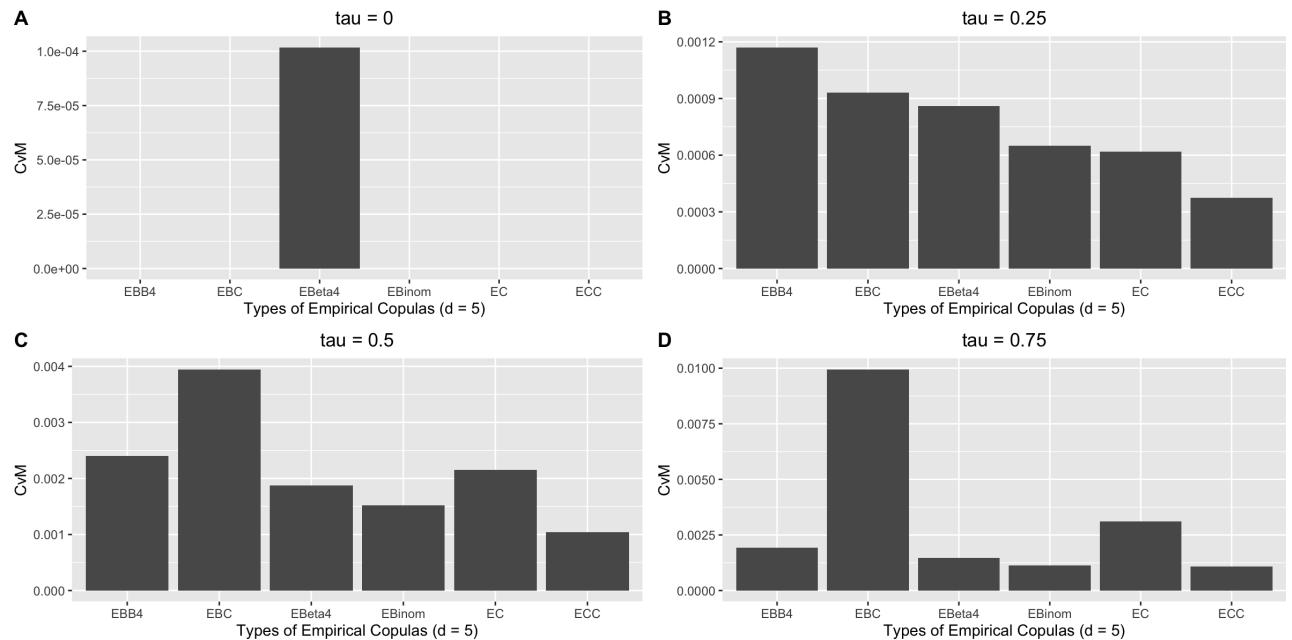


Figure 13: Gumbel-Hougaard copula with $d = 5$

4.4 Analysis: Exceedance probabilities

There are a few conclusions that can be reached by analysing the aforementioned plots.

4.4.1 Lower Kendall's tau values

It can be observed that there are an overall worse predictive performance at the lower Kendall's tau (notably negative Kendall's tau). From intuition, a negative Kendall's tau represents that, in the bivariate case, the margins are negatively related (points are concentrated on the top-left or bottom-right corner). Similar intuition holds for high-dimensional cases. Hence, there is little-to-no weight on the lower-tail or upper-tail (i.e. in the bivariate case, bottom-left and top-right corner). In the current analysis, the size of the dataset is relatively small (i.e. $n = 125$). Hence, even when sampling 125 samples from the target true copula, it is unlikely that the sample created will have any weight at the upper tail. Hence, for a smaller tau, most copulas (except beta-survival margins smoothed empirical copula) vastly underestimate the upper tail. It is noted that some graphs show a systematic overestimation with respect to the true copula, which is intuitive; as the probability of having samples at the extreme tails (for small Kendall's tau values) is low, sampling differences might bring the estimated copula to underestimate or overestimate the true copula.

It is notable that the smoothed EBC-adapted empirical copula with beta-binomial survival margins seems to best estimate the true-copula at lower values of Kendall's tau, compared to the other two types of smoothed empirical copulas.

To more accurately estimate the true-copula by using empirical copulas, a larger n is needed, as the class of empirical copula asymptotically converges to the true copula for large n . However, this becomes prohibitive when available data is not sufficient.

4.4.2 Erratic behaviour from smoothed EBC-adapted empirical copula with beta survival margins

It can be seen from the plots that the smoothed EBC-adapted empirical copula with beta survival margins has extremely volatile behaviour in the tail. It is noted that it is uncertain why the “survival function” exhibits a non-monotone pattern, as it is likely to be caused by rounding errors at the edge, as the order of the exceedance probabilities are from negative 5 to 10.

However, it is still interesting to understand the reason why, unlike the other types of empirical copulas, there appears to be a hump near $u = 0.99$.

Recall that the smoothed EBC-adapted empirical copula with beta survival margins is defined as follows:

$$C_n^v(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n C_n^\beta \left[\bar{\mathcal{B}}_{u_1}^{\mathbf{X}} \left(\frac{R_{i,1} - 0.5}{n} \right), \dots, \bar{\mathcal{B}}_{u_d}^{\mathbf{X}} \left(\frac{R_{i,d} - 0.5}{n} \right) \right]$$

where $\bar{\mathcal{B}}_{u_j}^{\mathbf{X}}$ is beta distributed. The survival function corresponding to this type of smoothed empirical copula is computed as follows:

$$\bar{H}_n^v(\mathbf{u}) = \bar{C}_n^v(\mathbf{1} - \mathbf{u}) = \sum_{J \subseteq \{1, \dots, d\}} (-1)^{|J|} C_n^v \left(u_1^{1(1 \in J)}, \dots, u_d^{1(d \in J)} \right), \quad \mathbf{u} \in [0, 1]^d$$

Hence we can see the main influencing factor on the size of the survival function is the smoothed EBC-adapted empirical copula with dimensions $j \in \{1, \dots, d\}$. Note that the beta distribution defined above is distributed with scale parameters α and β : (proof by algebraic manipulation, and can be found in [Kojadinovic and Yi, 2024](#)):

$$\begin{aligned} \alpha &= \frac{n - \rho}{\rho} \times u_j, \quad j \in \{1, \dots, d\} \\ \beta &= \frac{n - \rho}{\rho} \times (1 - u_j), \quad j \in \{1, \dots, d\} \end{aligned}$$

In our simulation, we used $n = 125$ and $\rho = 4$, and $u_j \in \{0.95, \dots, 1\}$. The resulting range of α can be calculated as $\alpha \in [28.7375, 30.25]$, and the resulting range of β can be calculated as $\beta \in [0, 1.5125]$. Hence, this is equivalent (componentwise), in evaluating the probability of a beta distribution with $\alpha \in [28.7375, 30.25]$ and $\beta \in [0, 1.5125]$ being larger than $\frac{R_{i,j}-0.5}{n} \in (0, 1)$. As the resulting α is extremely large and resulting β is extremely small, this results in a extremely heavy-right-tailed beta distribution. Substituting into the

smoothed empirical EBC-adapted empirical copula, it will result in a very abnormally large copula $C_n^v \left(u_1^{1(1 \in J)}, \dots, u_d^{1(d \in J)} \right)$, and hence a peak in the exceedance probability. As the appearance of an “up-tail” data point is rare during sampling, this sudden peak in the exceedance probability (in one data row) will skew the average exceedance probability. In further analysis, it is ideal to increase the sample size to see if this observations is retained.

With the following revelation, the smoothed EBC-adapted empirical copula with beta survival margins are very suitable (under a larger sample size n) to model extremal events (such as joint default scenarios).

4.4.3 Systematic underestimation of the upper-tail at positive Kendall’s tau

By closely observing the plots for higher Kendall’s tau (especially those with positive tau), we can observe that, albeit the estimators very closely resembles the true copula, it is often (systematically) lower than the true copula.

The problem of underestimation is rampant when the Kendall’s tau is close to 0. Due to the relatively small sample size ($n = 125$) and the discretized nature of the empirical copula, it places 0 probability mass if all rows in \mathbf{U} are smaller than the evaluation vector \mathbf{u} (when computing the survival function of the empirical copula). In reality, the true copula still has weight on the upper tail (albeit negligible), hence leading to the underestimation of the copula. Similarly to the other types of empirical copulas, they are in a sense, just continuous versions (EBC / EBB4 / EBinom / EBeta4) / interpolations (ECC) of the empirical copula. Hence, this underestimation property is unfortunately unavoidable, and could be overcame by increasing the sample size (as they are asymptotically convergent). As mentioned in the previous section, it is often unrealistic.

4.4.4 Bias and variance of different empirical estimators

From a interpretation standpoint, the empirical copula and the empirical beta copula are most “interpretable” in the sense that these two empirical copulas’ sampling algorithms are most simplistic, hence by interpreting the plots, it is easy to understand the result, in whether they are reasonable and intuitive. However, this is not the main focus of this work. We try to understand the bias and variance of the estimator with respect to the true

copula in the following paragraphs.

From a bias standpoint, we can try to interpret the Cramer-von-Mises statistic. It is possible, under a parametric setup, to estimate the p-value associated with the Cramer-von-Mises statistic. However, as the use of the class of empirical copulas is nonparametric, it is not possible to estimate the p-value of the Cramer-von-Mises statistic. Moreover, as we are utilising pseudo-observations, it is not suitable to devise a hypothesis test (and compute p-values). Hence, we will understand the proximity of the estimator to the true copula by comparing the magnitude of the CvM statistic between the different estimators.

For low values of Kendall's tau (non-positive tau), we can observe that the overall best performing estimator is the smoothed EBC-adapted empirical copula with beta-binomial survival margins, with the smallest S_n^{CvM} for most Kendall's tau ≤ 0 (with the exception of student- t $d = 2$ and $\tau = \{-0.5, 0\}$). However, the smoothed EBC-adapted empirical copula with beta survival margins is the worse performing estimator, as it has a much higher CvM compared to the other estimators (due to the erratic tail mentioned in previous sections).

For higher values of Kendall's tau (especially positive tau), we can observe that the empirical checkerboard copula and the 3 smoothed empirical copulas proposed by [Kojadinovic and Yi, 2024](#) overperforms the empirical copula and the empirical beta copula, as the CvM statistic is vastly smaller than EBC and EC (with some exceptions). Remarkably, the EBC-adapted empirical copula with binomial margins is the best performing estimator at high values of Kendall's tau with only few exceptions.

For the variance of the empirical copulas with respect to the real copula, we have also shown the confidence intervals implied by each empirical copula (in dotted lines). It can be remarked that the smoothed EBC-adapted empirical copula with beta-binomial and binomial survival margins respectively, have the lowest variance compared to the other estimators. This shows that beta-binomial and binomial survival margins smoothed EBC-adapted copula can achieve similar stability compared to other types of empirical copula even with a smaller sample size, which makes it a good candidate for tail modelling. On the other hand, despite the empirical checkerboard copula's exceedingly low bias (due to its interpolative nature), the variance is higher than the other empirical copulas. It can

be understood as the fact that empirical checkerboard copulas is an interpolation of the empirical copula, it is extremely flexible (in the bias-variance trade-off sense) and hence is prone to high variation, hence the higher variance.

It is important to note that there is no adequate estimators for negative values of Kendall's tau for this sample size.

4.5 Plots: u (evaluation points) vs cumulative probabilities

We then plot the (lower-tailed) evaluation points against the cumulative probabilities for each empirical copula. The evaluation points are on the x-axis and the cumulative probabilities are on the y-axis. The solid line is the mean of the exceedance probabilities corresponding to the evaluation point, and the dotted line is the confidence intervals of the exceedance probabilities corresponding to the evaluation point.

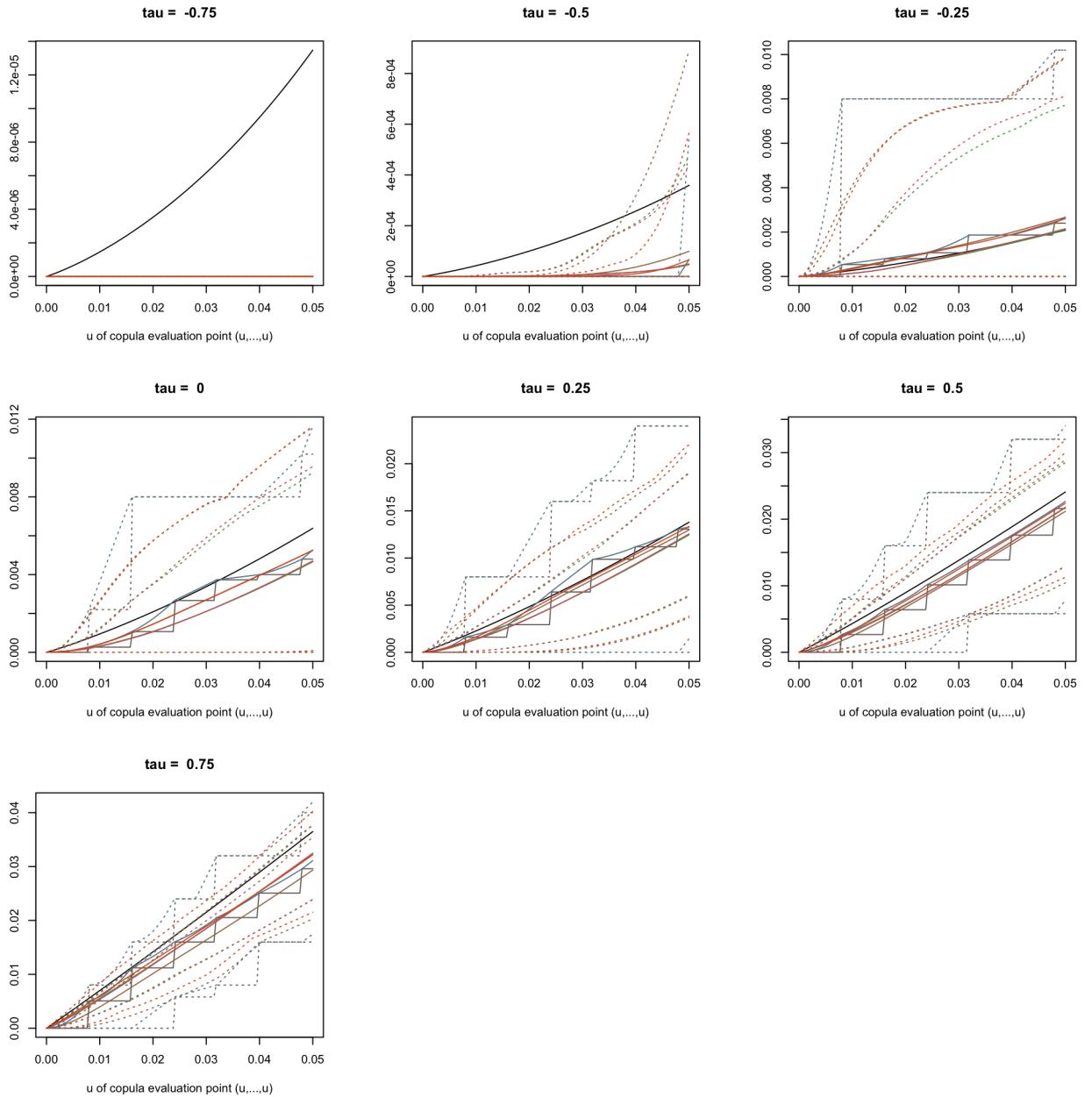


Figure 14: Student- t copula with $d = 2$

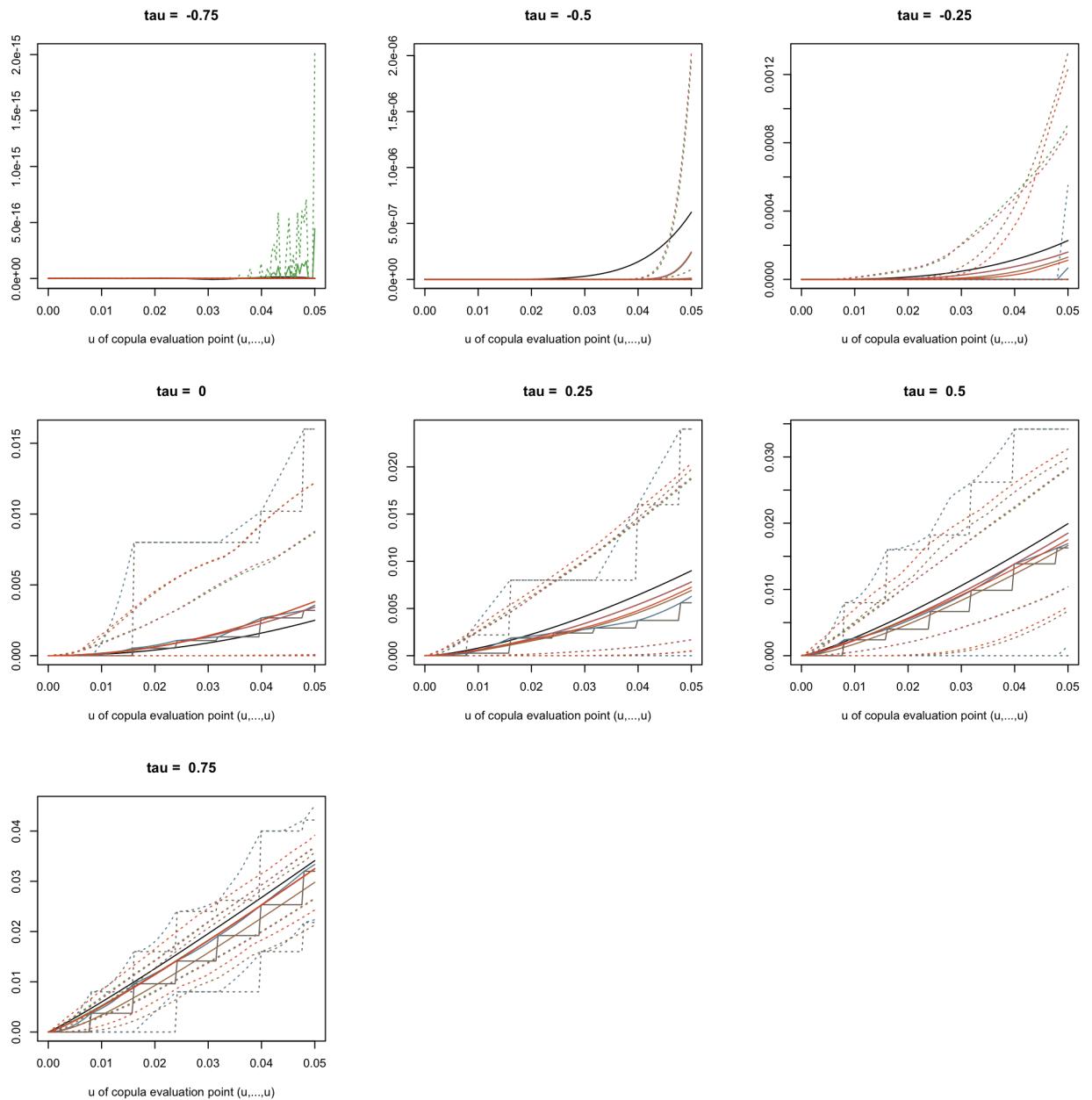


Figure 15: Gaussian copula with $d = 2$

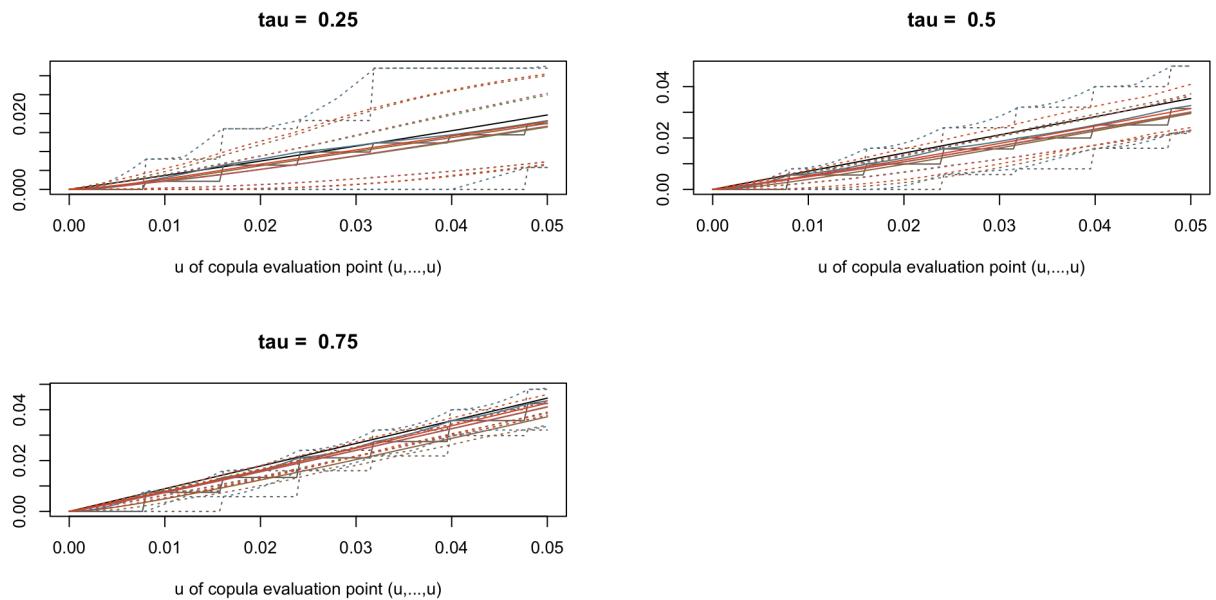


Figure 16: Clayton copula with $d = 2$

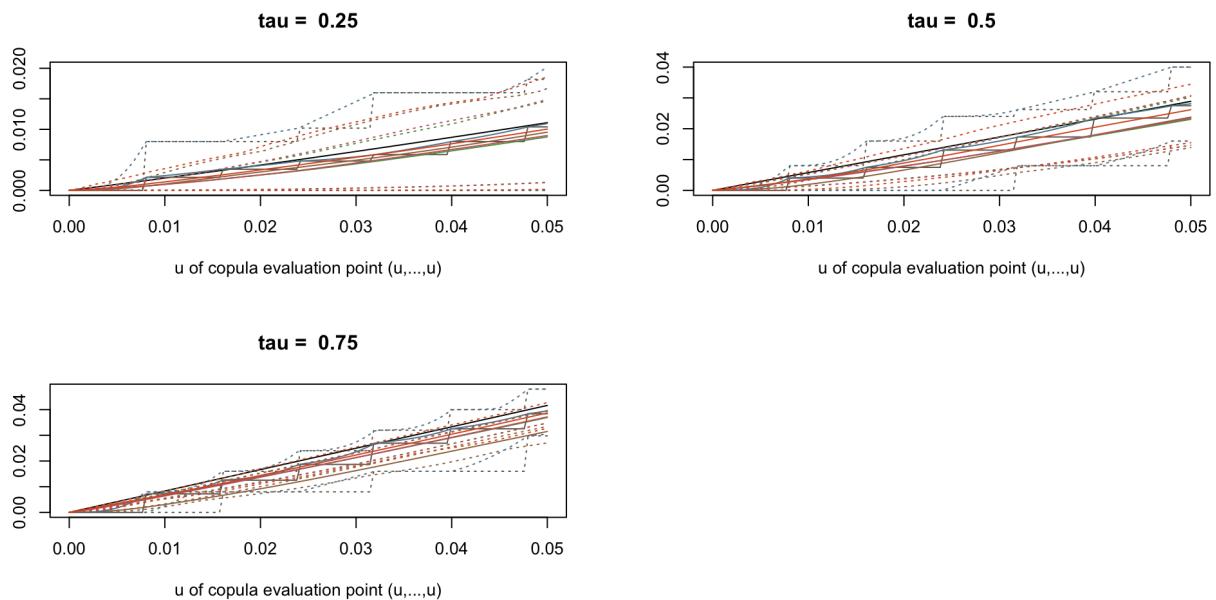


Figure 17: Clayton copula with $d = 3$

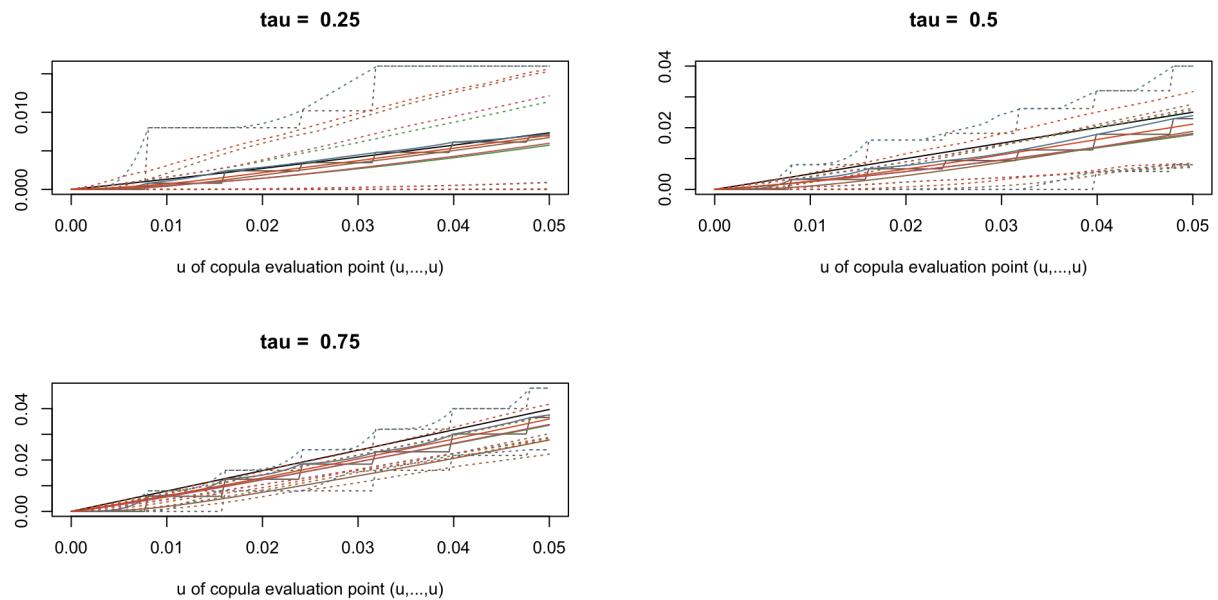


Figure 18: Clayton copula with $d = 4$

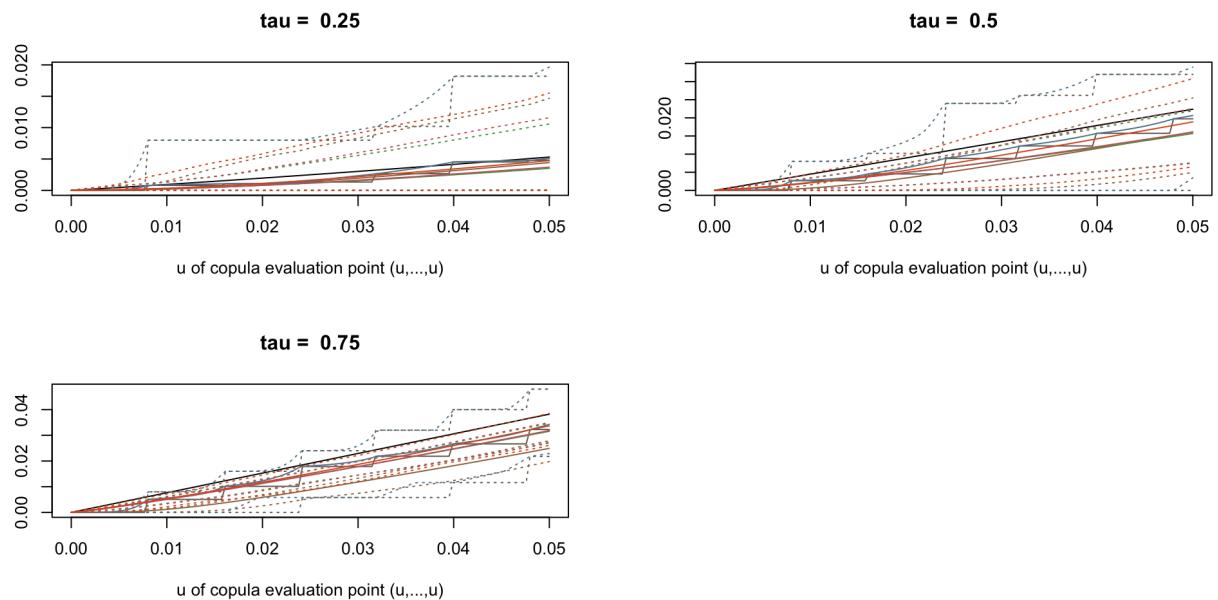


Figure 19: Clayton copula with $d = 5$

4.6 CvM: Cumulative probabilities

To better understand the proximity of the cumulative probabilities for the class of empirical copula, with respect to the true copula, we then compute the Cramer-von-Mises statistics for all combinations of empirical copulas, Kendall's tau and dimensions.

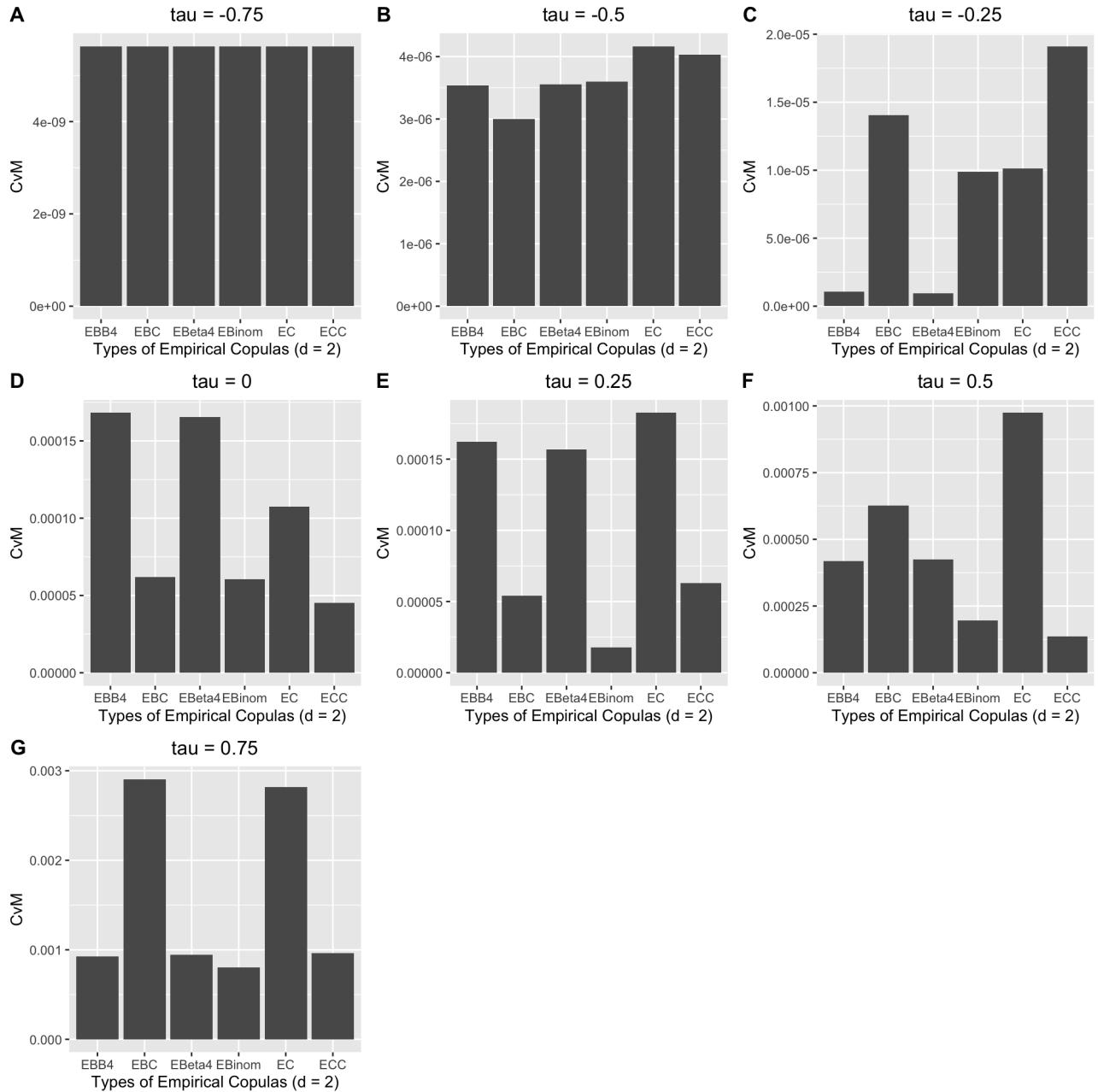


Figure 20: Student-*t* copula with $d = 2$

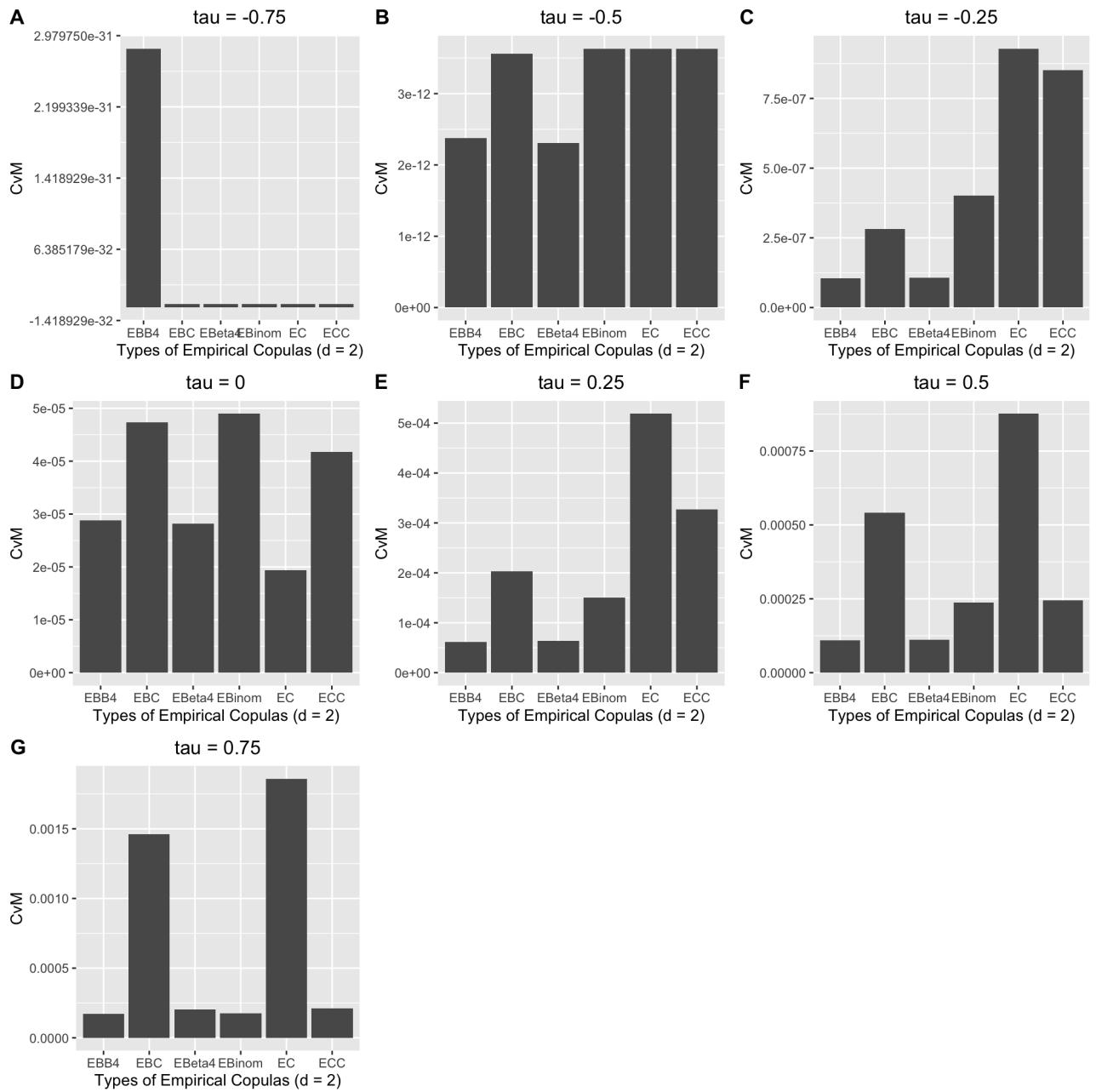


Figure 21: Gaussian copula with $d = 2$

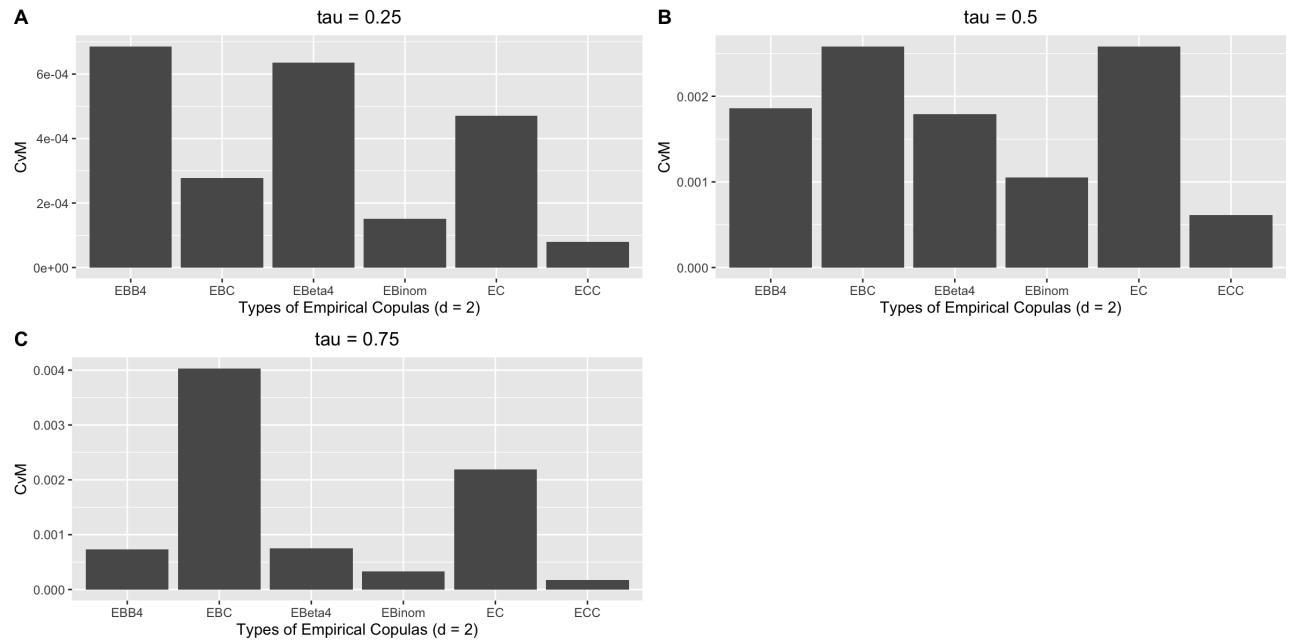


Figure 22: Clayton copula with $d = 2$

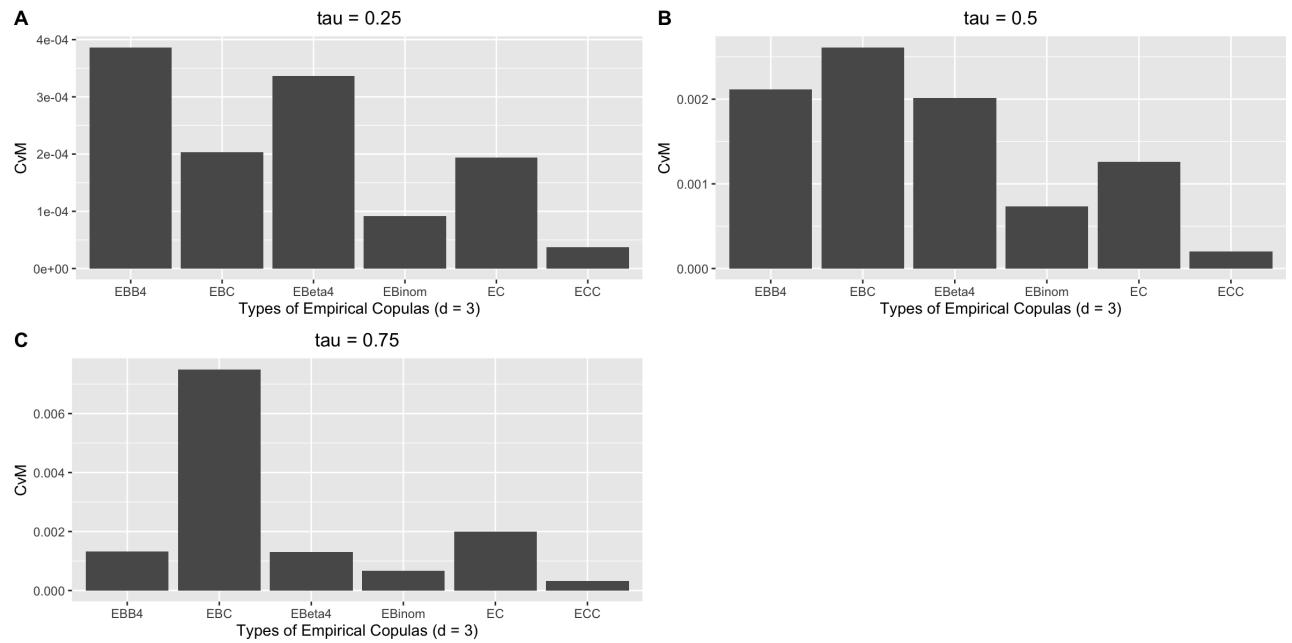


Figure 23: Clayton copula with $d = 3$

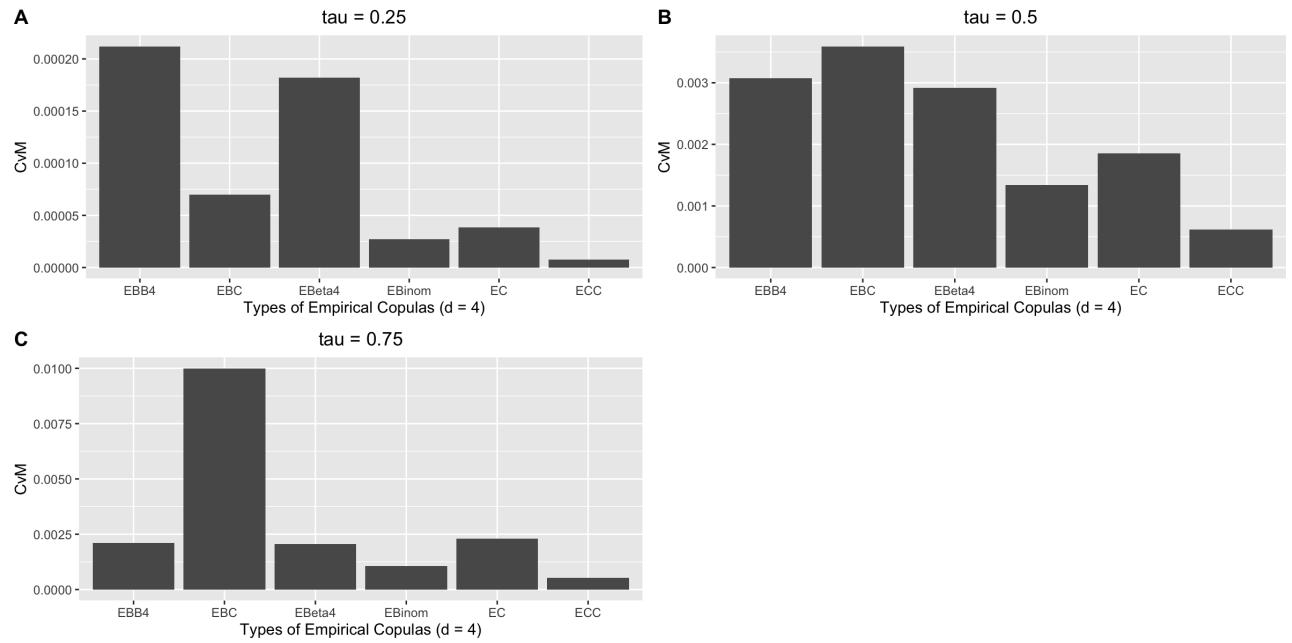


Figure 24: Clayton copula with $d = 4$

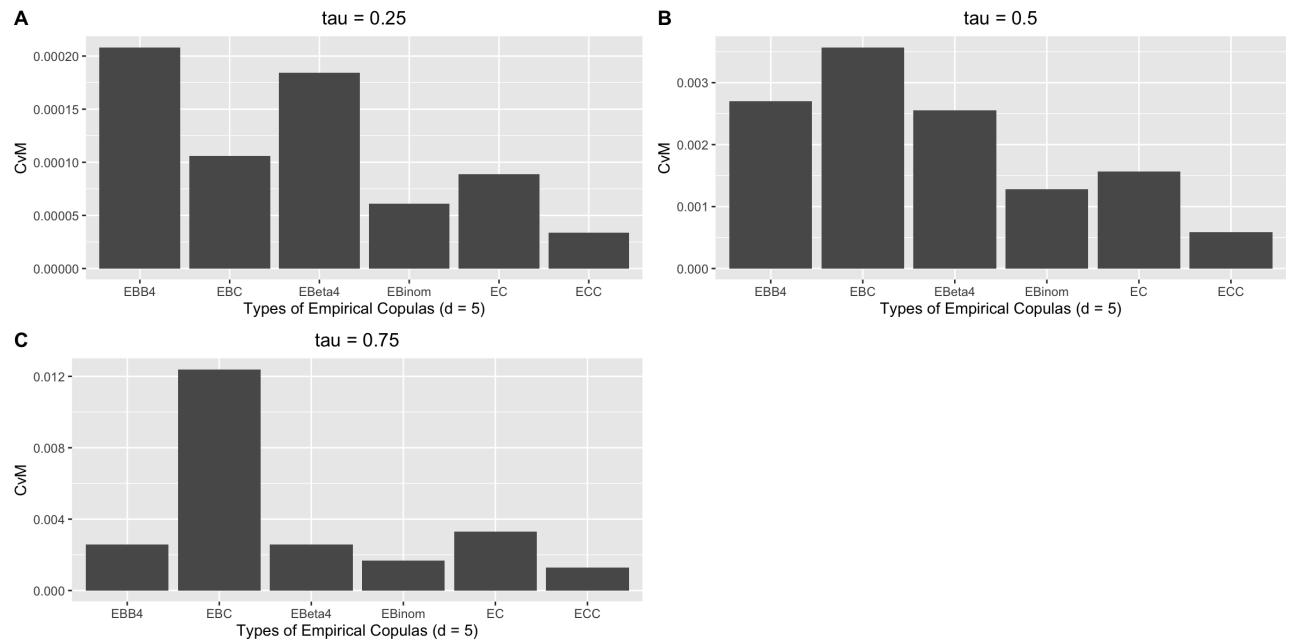


Figure 25: Clayton copula with $d = 5$

4.7 Analysis: Cumulative probabilities

Similar to the analysis for exceedance probabilities, there are a few conclusions that can be reached by analysing the aforementioned plots.

4.7.1 Lower Kendall's tau values

Similar to exceedance probabilities, the predictive performance of empirical copulas at the lower tail at low (negative) Kendall's tau values is poor. Similar to the intuition of the exceedance probabilities, a negative Kendall's tau represents that the points are concentrated on the top-left to bottom-right corner, and there is less probability mass in the bottom-left and top-right corner, for a bivariate case. Similarly, sampling with a sample size of $n = 125$ is unlikely to create any empirical mass at the lower-tail. For instance, by observing the plots corresponding to Student- t copulas ($d = 2$ and $\tau = \{-0.75, -0.5\}$) and Gaussian copulas ($d = 2$ and $\tau = \{-0.75, -0.5, -0.25\}$) reflects that some empirical copulas have empirical mass 0 at the lower tail. As empirical copulas asymptotically (for large n) converge to the true copula, having a higher sample size can mitigate the failure to model the lower tail. However, as mentioned in the exceedance probability analysis, the availability of large amounts of joint data are often unrealistic.

It is to note that for Gaussian copulas ($d = 2$ and $\tau = \{-0.75, -0.5\}$), it exhibits a rather peculiar sine-like / oscillating curve at the lower tail. This is caused as the cumulative probability is below the double precision machine accuracy, which led to numerical errors or rounding errors in the R environment, as it is at negative order 17.

4.7.2 Systematic underestimation of the lower-tail

Similar to the exceedance probability, the estimators at higher values of Kendall's tau (usually positively valued) are extremely close to the true copula, but exhibit a systematic underestimation of the lower-tail probability. The reasoning of this observation is virtually identical to the exceedance probability counterpart. For brevity, the original empirical copula (which is not a copula due to its discrete uniform margins) place 0 empirical mass on $[0, \mathbf{u}]$ if all rows in \mathbf{U} are larger than the evaluation point \mathbf{u} . Empirical copulas therefore systematically underestimate the true copula.

4.7.3 Bias and variance of different empirical estimators

From a bias standpoint, the best performing empirical copulas are the EBC-adapted empirical copula with binomial survival margins, and the empirical checkerboard copula. From the graph, it can be observed that the empirical checkerboard copula is extremely flexible around the true copula. A downside of this flexibility property is the high number of non-differentiable points relative to the other empirical estimators (except the empirical copula). On the other hand, the EBC-adapted empirical copula with binomial survival margins is differentiable nearly everywhere, yet retains the low bias with respect to the true copula, which slightly edges over empirical checkerboard copula despite its slightly inferior performance. The low bias for these two empirical copulas can also be observed via the CvM statistic plots (especially high Kendall's tau values). Notable exceptions include Student- t and Gaussian copula with $d = 2$ and Kendall's $\tau = -0.25$, where both the EBC-adapted empirical copula with beta and beta-binomial survival margins strictly outperform the other empirical estimators. It is well-noted in the exceedance probability section that the EBC-adapted empirical copula with beta-binomial survival margins also outperforms its competition for Student-t and Gaussian copula with $d = 2$ and $\tau = -0.25$. On the other hand, the empirical copula and the empirical beta copula are the worst two performing estimators with respect to bias of the true copula, which is consistent with the exceedance probability section.

From a variance standpoint, the empirical checkerboard copula and empirical copula have the largest variance. It is intuitive as they are an interpolation of the empirical copula, and the discretized nature of empirical copulas leads to the high variance. The EBC-adapted empirical copula with beta-binomial and beta survival margins have the lowest variance.

Similar to the exceedance probability modelling, no adequate estimators for extremely negative Kendall's tau are available due to the small sample size. It is only possible if the sample size is raised, which is limited by the restrictions mentioned earlier.

4.8 Recommendations

In light of the analysis done for exceedance probabilities and cumulative probabilities at upper-tail and lower-tail respectively, we propose estimators that are best in replicating the true copula.

4.8.1 Modelling exceedance probabilities at upper-tail

$\tau \leq -0.5$: N/A

$\tau = -0.25$: EBC-adapted empirical copula with beta-binomial survival margins

$\tau \geq 0$: EBC-adapted empirical copula with binomial survival margins

4.8.2 Modelling cumulative probabilities at lower-tail

$\tau \leq -0.5$: N/A

$\tau = -0.25$: EBC-adapted empirical copula with beta or beta-binomial survival margins

$\tau \geq 0$: EBC-adapted empirical copula with binomial survival margins

5 Conclusion

In this report, we have provided an overview of the important definitions and properties of copulas, introduced common copulas and rank correlation measures, and utilised 6 different types of empirical copulas. After that we have proposed algorithms to compute the exceedance probability and cumulative probability for a set of evaluation points (up-tail and low-tail respectively). Lastly, we plotted the exceedance probability and cumulative probability of the true copula and the empirical copulas proposed, and evaluated the underestimation and overestimation of the true copula by a goodness-of-fit test: Cramer-von-Mises test statistic.

Upon evaluation, we observe that despite the erratical behaviour of the EBC-adapted empirical copula with beta survival margins at negative Kendall's tau, it is the best estimator of the true copula for modelling cumulative probabilities at negative Kendall's tau together with the EBC-adapted empirical copula with beta-binomial survival margins. EBC-adapted empirical copula with beta-binomial survival margins is also the best performing estimator for modelling exceedance probabilities at negative Kendall's tau. For positive Kendall's tau, the EBC-adapted empirical copula with binomial survival margins is the best estimator for the true copula in modelling both cumulative and exceedance probabilities. This is in-line with the observations noted in [Kojadinovic and Yi, 2024](#), where the MSE of the smoothed empirical copulas overperform the empirical copula, empirical beta copula and empirical checkerboard copula. Specifically, the empirical checkerboard copula's variance is too large despite its low bias.

To improve the procedure of modelling exceedance probabilities and cumulative probabilities, it is best to have a more efficient algorithm in computing the survival function of empirical copulas. For instance, the survival copula is calculated using Appendix Theorem 1.4.2, which is extremely prohibitive as the dimension increases. The computation of survival function in dimension d requires the computation of 2^d empirical copulas. Future research could look at how to derive the survival function of the empirical copula directly, instead of obtaining 2^d empirical copulas.

Moreover, this research is grounded on trying to recover the true copula, and we did not apply it into a real-life dataset (such as modelling joint default probabilities). Future

research could entail providing a case study with a train-test split to evaluate the effectiveness of the class of empirical copulas outlined in this report. Recall that the financial crisis is inadvertently caused by a so-called “misspecification” of the copula. To be more exact, the Gaussian copula is widely used in the pricing of complex financial products, but the event of joint defaulting is highly dependent, whereas the Gaussian copula is not, hence underestimating the probability of joint default causing huge financial firms to go under (such as the Lehman Brothers). Empirical copulas opened the door to a realm of possibilities due to the property that it will never be misspecified: empirical copulaS always converge to the true copula upon a high enough sample size. In this report, it is shown that at high Kendall’s tau, empirical copulas closely resembles the true copula, hence showing its effectiveness potentially in financial modelling. It is though noted that the underestimation problem cannot be averted completely, hence it is recommended that further research can investigate suitable adjustment on the probability to minimize the difference between the estimator and the copula.

It is also remarked that, considering the small sample size, empirical copulas are rendered useless in the estimation of the exceedance and cumulative probability of the true copula under negative Kendall’s tau, which is not ideal. In the real world, there are often indices that are negatively correlated / discordant. For example, fixed income securities thrive under a high interest rate environment, while equity stocks plummet due to the increased borrowing costs. If variables with negative Kendall’s tau can be accurately modelled, it can allow more complex financial products / options to be created.

It could also have further applications in the insurance sector. Insurance products involving multiple-persons (couple insurance) is virtually unseen. Modelling bivariate mortality rates / morbidity rates (which is often extremely low) using the class of empirical copulas could also be one avenue of research.

All-in-all, Empirical copulas are extremely powerful and tractable under the correct assumptions due to their high flexibility and proximity to the true copula. By applying this technique to different settings, it can potentially revolutionise the financial sector, without revitalising past financial failures.

Bibliography

- Carley, H., & Taylor, M. D. (2002). A new proof of sklar's theorem. *Distributions with given marginals and statistical modelling*, 29–34.
- Clayton, D. (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.
- 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, 274–292.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer Berlin, Heidelberg.
- Fréchet, M. (1951). Sur les tableaux dont les marges et des bornes sont données. *Annales de l'Université de Lyon*, 9, 53–77.
- Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2), 199–213.
- Gumbel, E. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association*, 56, 335–349.
- Haan, L., & Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer New York, NY.
- Höoeffding, W. (1940). *Maßstabinvariante korrelationstheorie*. Teubner.
- Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2018). *Elements of Copula Modeling with R*. Springer Use R! Series. <https://www.springer.com/de/book/9783319896342>
- Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2023). *Copula: Multivariate dependence with copulas* [R package version 1.1-3]. <https://CRAN.R-project.org/package=copula>
- Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, 73, 671–678.

(<https://math.stackexchange.com/users/156814/jan>), J. (2023). Joint distribution given marginals and correlation [URL:<https://math.stackexchange.com/q/4665041> (version: 2023-03-23)].

Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30(1-2).

Kojadinovic, I., & Yi, B. (2024). A class of smooth, possibly data-adaptive nonparametric copula estimators containing the empirical beta copula. *Journal of Multivariate Analysis*, 201.

Marshall, A. W., & Olkin, I. (1988). Families of multivariate distributions. *American Statistical Association*, 83.

McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques, and tools*. Journal of The American Statistical Association - J AMER STATIST ASSN.

Nelson, R. (2006). *An introduction to copulas*. Springer.

Neuwirth, E. (2022). *Rcolorbrewer: Colorbrewer palettes* [R package version 1.1-3]. <https://CRAN.R-project.org/package=RColorBrewer>

Segers, J., Sibuya, M., & Tsukahara, H. (2017). The empirical beta copula. *Journal of Multivariate Analysis*, 155, 35–51.

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

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101.

Wolodzko, T. (2023). *Extradistr: Additional univariate and multivariate distributions* [R package version 1.10.0]. <https://CRAN.R-project.org/package=extraDistr>

1 Appendix: Copula Prerequisites

1.1 Definition and Construction

We first establish two important (and well-known) results on probability transformation. When F is a continuous distribution, F^{-1} is equivalent to the inverse of F .

Lemma 1.1.1 *Integral Transform*

For a r.v. $X \sim F$ where F is a continuous distribution function $\Rightarrow F(X) \sim \mathcal{U}(0, 1)$.

Proof: (Hofert et al., 2018) $F_{F_X}(x) = \mathbb{P}(F_X(x) \leq y) = \mathbb{P}(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y \Rightarrow F(X) \sim \mathcal{U}(0, 1)$

Lemma 1.1.2 *Quantile Transform*

For a r.v. $U \sim \mathcal{U}(0, 1)$, F and F^{-1} being any distribution function and its corresponding quantile function respectively $\Rightarrow F_X^{-1}(U) \sim F$.

Proof: Similar to Lemma 1.1, $\mathbb{P}(F_X^{-1}(U) \leq x) = \mathbb{P}(U \leq F_X(x)) = F_X(x)$

These two transformation implies that it is trivial to transform any continuous distribution into a standard uniform distribution, and subsequently to another continuous distribution of interest. This allows us to relate random variables with different underlying distributions easily. We can generalise dependence by creating a continuous multivariate distribution function that links all the marginal standard uniform distributions.

The definition of Copula is illustrated below, with reference to Hofert et al., 2018.

Definition 1.1.1 *Copula*

A copula is a continuous d -dimensional distribution function $C(u_1, u_2, \dots, u_d)$ that satisfies the following properties:

1. *Groundedness:* $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_d) = 0$ if any $u_i = 0$.

2. $\mathcal{U}(0, 1)$ margins: $\forall i \in \{1, \dots, d\}$, $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$

3. *d-increasing:* $\mathbf{a}, \mathbf{b} \in [0, 1]^d$, $\Delta_{(\mathbf{a}, \mathbf{b})} C = \sum_{i \in \{0, 1\}^d} (-1)^{\sum_{j=1}^d i_j} C(a_1^{i_1} b_1^{1-i_1}, \dots, a_d^{i_d} b_d^{1-i_d}) \geq 0$,
where $\Delta_{(\mathbf{a}, \mathbf{b})} C$ is the *C-volume* defined w.r.t hyperrectangle (\mathbf{a}, \mathbf{b}) .

What makes copulas so popular is its uniqueness under a continuous framework. This is brought about by the famous results of Sklar's Theorem.

Theorem 1.1.1 (Sklar, 1959) *Sklar's Theorem*

1. *Existence:* For univariate margins $F_1, F_2, \dots, F_d \sim \mathcal{U}(0, 1)$, a *d-dimensional distribution function* F , $\exists C : [0, 1]^d \rightarrow [0, 1]$ such that $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$, $\mathbf{x} \in \mathbb{R}^d$, where C is uniquely defined on $\prod_{j=1}^d \text{ran } F_j$.
2. *Converse:* Given any copula $C : [0, 1]^d \rightarrow [0, 1]$ and univariate margins F_1, F_2, \dots, F_d , $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$ is a *d-dimensional distribution function*, $\mathbf{x} \in \mathbb{R}^d$.

Notice that if the univariate margins F_1, F_2, \dots, F_d are continuous (and $\mathcal{U}(0, 1)$ distributed), C is uniquely defined.

This shows that there is a one-to-one correspondence between the univariate margins to the observable *d-dimensional distribution function*, linked by a unique copula function.

Its' significance can be illustrated by the following analogy provided in [Hofert et al., 2018](#). Consider two *d-dimensional random vectors* \mathbf{X} and \mathbf{Y} , where its components (marginals) (i.e. (X_1, \dots, X_d) and (Y_1, \dots, Y_d)) are distributed differently (e.g. let's propose that X_1, \dots, X_i are gamma distributed, X_{i+1}, \dots, X_d are normally distributed; Y_1, \dots, Y_i are beta distributed, Y_{i+1}, \dots, Y_d are pareto distributed). However, it is unrealistic to compare their dependence as their marginal distribution are different. However, utilising Appendix Lemma 1.1.1 can allow us to transform all marginal distributions to standard uniform margins, then applying a Copula function to model dependence. If it is provided that the random vectors are continuous, Having the same joint distribution (or the same copulas)

implies that they have the same dependence structure.

It's ability to model dependence between multivariate vectors can be amplified by the invariance principle. The theorem presented below is obtained from [Hofert et al., 2018](#):

Theorem 1.1.2 Invariance Property of Copulas

Let $\mathbf{X} \sim F$ with continuous univariate margins F_1, F_2, \dots, F_d and copula C . It follows that $T_1(X_1), T_2(X_2), \dots, T_d(X_d)$ also has copula C , where $\forall j \in \{1, \dots, d\}$, T_j is a strictly increasing function on $\text{ran } X_j$.

Proof:

$$\begin{aligned} P(T_1(X_1) \leq x_1, \dots, T_d(X_d) \leq x_d) &= P(X_1 \leq T_1^{-1}(x_1), \dots, X_d \leq T_d^{-1}(x_d)) \\ &= C(F_1(T_1^{-1}(x_1)), F_2(T_2^{-1}(x_2)), \dots, F_d(T_d^{-1}(x_d))) \end{aligned}$$

We can also see that $\forall j \in \{1, \dots, d\}$, $P(T_j(X_j) \leq x_j) = P(X_j \leq T_j^{-1}(x_j)) = F_j(T_j^{-1}(x_j))$ and the result follows.

Copulas can provide a way to understand the dependence without knowledge of the original (before-transformed) marginal distribution, and the fact that imposing an increasing transformation does not distort dependence allows flexibility in transforming the variable to standard uniform marginals without the worry of over or underestimating dependence.

Before introducing some common copula families and distributions, it is best to define a bound for the possible values of a copula. The definitions are obtained from [Hofert et al., 2018](#).

Theorem 1.1.3 (Höeffding, 1940; Fréchet, 1951) Frechét-Hoeffding Bounds

$$\begin{aligned} C^-(\mathbf{u}) &= \max\left\{\sum_{j=1}^d u_j - d + 1, 0\right\}, \quad C^+(\mathbf{u}) = \min_{1 \leq j \leq d} \{u_j\}, \quad \mathbf{u} \in [0, 1]^d \\ \Rightarrow \forall C(\mathbf{u}), \quad C^-(\mathbf{u}) &\leq C(\mathbf{u}) \leq C^+(\mathbf{u}), \quad \mathbf{u} \in [0, 1]^d \end{aligned}$$

This establishes the maximal and minimal value of a copula (which is analogically similar to +1 and -1 in pearson's correlation coefficient). Note that $C^-(\mathbf{u})$ at $d = 2$ is the countermonotonic copula and $C^+(\mathbf{u})$ is the comonotonic copula.

1.2 Common Copula distributions

Distributions of Copulas have been well-researched in the past decade. There are three main copula classes: Archimedean copulas, Elliptical copulas and Extreme-value copulas. In this report, we focus on the first two classes. Definition of these copulas are obtained from [Hofert et al., 2018](#).

1.2.1 Archimedean Copulas

Definition 1.2.1 ([Hofert et al., 2018](#)) *Archimedean copula*

$C(\mathbf{u})$ is an archimedean copula if it exhibits the following form:

$$C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \cdots + \psi^{-1}(u_d)), \quad \mathbf{u} \in [0, 1]^d$$

where ψ is a generator function.

Definition 1.2.2 ([Hofert et al., 2018](#)) *Generator Function*

ψ is a generator function for an archimedean copula if it exhibits the following properties:

1. $\psi : [0, \infty] \rightarrow [0, 1]$ is continuous, strictly decreasing; $\psi(0) = 1$, $\psi(\infty) \rightarrow 0$
2. $\psi \in \Psi$, where Ψ is the class of all generators
3. $\forall k \in \{0, \dots, d-2\}$, $(-1)^k \psi^{(k)}(t) \geq 0$ and $(-1)^{d-2} \psi^{(d-2)}(t)$ is nonincreasing and convex on $t \in (0, \infty)$ (d-monotonicity)

Note that if $\forall k \in \{0, \dots, d\}$, $(-1)^k \psi^{(k)}(t) \geq 0$, ψ is completely monotone.

For the purpose of this report, only heavy-tailed copulas are presented below:

1. **Clayton copula** (Clayton, 1978)

$$C_\theta^C(\mathbf{u}) = \begin{cases} \max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\}^{-1/\theta}, & d = 2, \theta \in [-1, \infty) \setminus \{0\}, \mathbf{u} \in [0, 1]^2 \\ (1 - d + \sum_{j=1}^d u_j^{-\theta})^{-1/\theta}, & d \geq 3, \theta \in (0, \infty), \mathbf{u} \in [0, 1]^d \end{cases}$$

where $\psi(t) = (1+t)^{-1/\theta}$

2. **Gumbel-Hougaard copula** (Gumbel, 1961; Hougaard, 1986)

$$C_\theta^{GH}(\mathbf{u}) = \exp\left(-\left(\sum_{j=1}^d (-\ln u_j)^\theta\right)^{1/\theta}\right), \theta \in [1, \infty), \mathbf{u} \in [0, 1]^d$$

where $\psi(t) = \exp(-t^{1/\theta})$

It is noted that Clayton copulas have unusually high weight on the left (low) tail, and Gumbel-Hougaard have unusually high weight on the right (upper) tail, hence these copulas are used to evaluate the effectiveness of evaluating low and upper tail risk by the class of empirical copulas.

1.2.2 Elliptical Copulas

Definition 1.2.3 (Hofert et al., 2018) *Elliptical copula*

$C(\mathbf{u})$ is an elliptical copula if it exhibits the following form:

$$C(\mathbf{u}) = H(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)), \mathbf{u} \in \prod_{j=1}^d \text{ran } F_j$$

where H is a multivariate elliptical distribution function, F_1, \dots, F_d are the univariate margins.

Definition 1.2.4 (Hofert et al., 2018) *Elliptical distribution*

\mathbf{X} is an elliptical distribution if it exhibits the following form:

$$\mathbf{X} = \boldsymbol{\mu} + R\mathbf{A}\mathbf{S}$$

where $\mathbf{S} \sim \mathcal{U}(\{\mathbf{x} \in \mathbb{R}^k : \|\mathbf{x}\| = 1\})$ is uniformly distributed across the unit sphere of order k , \mathbf{A} is the lower triangular square matrix with nonnegative entries such that $\mathbf{A}\mathbf{A}^T = \Sigma$, where Σ is the covariance matrix (Cholesky decomposition of Σ), and $R \geq 0$ having a density:

$$f_R(r) = \frac{2\pi^{\frac{d}{2}} r^{d-1} g(r^2)}{\Gamma(\frac{d}{2})}$$

It is remarked that if $\text{rank}(\Sigma) = d$ (full rank), then the density function of \mathbf{X} exhibits the form:

$$h(\mathbf{x}) = \frac{1}{\sqrt{\det \Sigma}} g((\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}))$$

where $g : [0, \infty) \rightarrow [0, \infty)$ is the density generator

If we supply a specific g , a different elliptical copula can be created. The two main copulas in the elliptical copula class are t copulas and normal copulas. Although only t copulas have heavy tail on both tails, considering the significance of the Gaussian copulas, both copulas are listed below:

1. Student t copula (Hofert et al., 2018)

The density generator g exhibits the form:

$$g(t) = \frac{\Gamma(\frac{v+d}{2})}{(\pi v)^{\frac{d}{2}} \Gamma(\frac{v}{2})} \left(1 + \frac{t}{v}\right)^{-\frac{v+d}{2}}, \quad t \in [0, \infty)$$

then the density of the copula is:

$$c^t(\mathbf{u}) = \frac{\Gamma(\frac{v+d}{2})}{\Gamma(\frac{v}{2}) \sqrt{\det P}} \left(\frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} \right)^d \frac{(1 + \frac{\mathbf{x}^T \Sigma^{-1} \mathbf{x}}{v})^{-\frac{v+d}{2}}}{\prod_{j=1}^d (1 + \frac{x_j^2}{v})^{-\frac{v+1}{2}}}, \quad \mathbf{u} \in (0, 1)^d$$

where $\mathbf{x} = (t_v^{-1}(u_1), \dots, t_v^{-1}(u_d))$.

2. Gaussian copula (Hofert et al., 2018)

The density generator g exhibits the form:

$$g(t) = \frac{\exp(-\frac{t}{2})}{(2\pi)^{\frac{d}{2}}}, t \in [0, \infty)$$

then the density of the copula is:

$$c^n(\mathbf{u}) = \frac{1}{\sqrt{\det P}} \exp\left(-\frac{1}{2}\mathbf{x}^T(\Sigma^{-1} - I_d)\mathbf{x}\right), \mathbf{u} \in (0, 1)^d$$

where $\mathbf{x} = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$.

It is noted that Student t copulas have high tail dependence on both sides of the tail, while Gaussian copulas do not have tail dependence on both sides of the tail (truly both sides of the spectrum!).

1.2.3 Extreme-value Copulas

Extreme value copulas are only slightly mentioned due to the fact that the Gumbel-Hougaard Copula is also a extreme-value copula.

Considering the one dimensional case, distributional characteristics in the tail are often underestimated, hence it is optimal to fit a distribution that only considers the maximal value of a set of observations. As discussed in [Embrechts et al., 1997](#), By choosing a suitable block size and obtaining the maxima, it can be seen that, if suitably normalised and let block size $n \rightarrow \infty$, it will converge to the generalised extreme-value distribution.

Generalised in multiple literatures (eg. [Haan and Ferreira, 2006](#), [Hofert et al., 2018](#)), we can extend this condition to the multivariate case; i.e., if the i.i.d. of block maxima random vectors are suitably normalised, they will converge to the multivariate generalised extreme-value distribution. Note that the marginal distributions are also univariate generalised extreme-value distributions.

The definition of the extreme-value copula are obtained from [Hofert et al., 2018](#).

Definition 1.2.5 (Hofert et al., 2018) *Extreme-value copula*

$C(u)$ is an extreme-value copula if for $\mathbf{u} \in [0, 1]^d$, there exists a C^* that exhibits the following properties:

$$\lim_{n \rightarrow \infty} C^*(\mathbf{u}^{\frac{1}{n}})^n = C(\mathbf{u})$$

and $C^* \in MDA(G_\xi)$ (Maximum domain of attraction), where G_ξ is the generalised extreme-value distribution with shape parameter ξ .

An important result of extreme-value copulas can be obtained by algebraic manipulation. Proof is omitted but could be found in Hofert et al., 2018.

Lemma 1.2.1 (Hofert et al., 2018) *Max stability characterisation of Extreme-value copulas*

$$\begin{aligned} & A \text{ copula } C \text{ is a extreme-value copula} \\ \iff & \forall \mathbf{u} \in [0, 1]^d \text{ and } r \in \mathbb{N}, C(\mathbf{u}^{\frac{1}{r}})^r = C(\mathbf{u}) \end{aligned}$$

One notable example of the extreme-value copula is the gumbel-hougaard copula, which is also an archimedean copula.

1.3 A Primer in Correlation Measures

In this section, several rank correlation measures will be introduced as ways to quantify dependence other than the most commonly used pearson's correlation coefficient.

1.3.1 Pitfalls of Pearson's Correlation Coefficient

Hofert et al., 2018 has provided a very detailed analysis on how Pearson's correlation coefficient is deficient in measuring dependence. This report will provide a brief overview on the pitfalls of pearson's correlation coefficient.

1. Nonexistence of correlation coefficient

Recall the definition: $\rho_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}}$. If either X_i or X_j does not have a

second moment, then ρ_{ij} will be undefined. A possible counterexample is student t distribution with a degree of freedom $v \in (1, 2]$, then the second moment is unbounded.

2. Invariance property violated

Notice that correlation coefficient measures the magnitude of change of a variable with respect to the opposing variable, hence a linear increasing transformation fulfils the invariance requirement. However, a non-linear transformation (even if increasing) on the random variable will lead to a possible multiplicative / non-linear distortion of the magnitude, which leads to a change in the correlation coefficient as it is only a measure of linear dependence.

3. One-to-one correspondence between marginal distributions and correlation coefficient on the joint distribution violated

A elegant counterexample is provided, originated from stackexchange: (<https://math.stackexchange.com/questions/371333/can-two-random-variables-with-the-same-marginal-distributions-be-un-correlated> 2023)

$X \in \{-1, 0, 1\}$, $Y \in \{0, 1\}$, for any $p \in (\frac{1}{6}, \frac{1}{3})$,

$$\begin{aligned} P(X = -1, Y = 1) &= P(X = 1, Y = 1) = p \\ P(X = -1, Y = 0) &= P(X = 1, Y = 0) = \frac{1}{3} - p \\ P(X = 0, Y = 0) &= 2p - \frac{1}{3} \\ P(X = 0, Y = 1) &= \frac{2}{3} - 2p \end{aligned}$$

We observe that they have same marginals for all values of p , and X and Y are uncorrelated. This shows that there are an infinite amount of joint distributions corresponding to the marginal distribution and correlation.

4. Uncorrelatedness does not imply Independence

A good counterexample would be a transformation on a standard normal random variable: $X \sim \mathcal{N}(0, 1)$. Consider X^2 . $\rho_{X,X^2} \neq 0$ but $E[XX^2] - E[X]E[X^2] = E[X^3] -$

$E[X]E[X^2] = 0$ as the skewness of a standard normal random variable = 0 due to symmetry.

5. Attaining Correlation = -1 or 1 is not necessarily possible for some marginal distributions

A good counterexample would be two Bernoulli random variables. Let $X \sim Ber(p)$, $Y \sim Ber(q)$. By the Frechét bounds, the correlation coefficient have the following bounds:

$$\frac{\max(p + q - 1, 0) - pq}{\sqrt{pq(1-p)(1-q)}} \leq \rho \leq \frac{\min(p, q) - pq}{\sqrt{pq(1-p)(1-q)}}$$

There exists some values of p and q such that the lower and upper bound of the Frechét bounds are not equal to -1 and 1 respectively.

1.3.2 Spearman's Rho

It is worthy to mention that spearman's rho is equivalent to the pearson's correlation coefficient of the two random variables after applying integral transformation.

Definition 1.3.1 (Spearman, 1904) *Spearman's Rho*

For a bivariate random vector (X, Y) , where $X \sim F$ and $Y \sim G$,

$$\rho_s = \rho_s(X, Y) = \text{Corr}(F(X), G(Y))$$

Remark 1.1 *Copula Dependency*

Recall that the formula for Spearman's Rho is:

$$\rho_s = \frac{\text{Cov}(F(X), G(Y))}{\sqrt{\text{Var}(F(X))\text{Var}(G(Y))}}$$

From the Hoeffding Covariance lemma ([Hoeffding, 1940](#), [McNeil et al., 2005](#)) it provides an integral representation of the covariance:

$$\text{Cov}(X, Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} H(x, y) - F(x)G(y) dy dx$$

It can be shown that upon substitution of X and Y by $F(X)$ and $G(Y)$:

$$\begin{aligned}\text{Cov}(F(X), F(Y)) &= \int_0^1 \int_0^1 H(F^{-1}(x), G^{-1}(y)) - F(F^{-1}(x))G(G^{-1}(y)) \, dy \, dx \\ &= \int_0^1 \int_0^1 H(F^{-1}(x), G^{-1}(y)) - xy \, dy \, dx\end{aligned}$$

By Sklar's Theorem ([Sklar, 1959](#)), $H(F^{-1}(x), G^{-1}(y)) = C(x, y)$ and it follows that

$$\text{Cov}(F(X), F(Y)) = \int_0^1 \int_0^1 C(x, y) - xy \, dy \, dx$$

Which is independent of the marginal distributions $F(\cdot)$ and $G(\cdot)$. Note that by Appendix Lemma 1.1.1, $F(X), G(Y) \sim \mathcal{U}(0, 1)$ with $\mathbb{E}[\mathcal{U}(0, 1)] = \frac{1}{2}$ and $\text{Var}(\mathcal{U}(0, 1)) = \frac{1}{12}$, It follows that:

$$\rho_s = 12 \text{Cov}(F(X), G(Y)) = 12 \int_0^1 \int_0^1 C(x, y) - xy \, dy \, dx$$

This immediately implies that any increasing transformation on X and Y does not affect spearman's rho.

Also notice that the integral transform of X and Y are standard uniform, which implies that spearman's rho always exists.

As shown in [Hofert et al., 2018](#), the lower and upper bounds of bivariate copula defined by the Fréchet-Hoeffding bounds are essentially equivalent to almost surely decreasing and increasing function. It then follows by [Nelson, 2006](#) Theorem 5.1.8 that the countermonotonic and comonotonic bivariate copula have spearman's rho of -1 and +1 respectively. Hence, $\rho_s \in [0, 1]$.

1.3.3 Kendall's Tau

Before we dive into the definition of Kendall's Tau, we first establish the concepts of concordance and discordance.

Definition 1.3.2 *Concordance and Discordance*

For two points at (x, y) and (x', y') , it is concordant if

$$(x - x')(y - y') > 0$$

It is discordant if

$$(x - x')(y - y') < 0$$

Definition 1.3.3 (Kendall, 1938, Hofert et al., 2018) **Kendall's Tau**

For a bivariate random vector (X, Y) , where $X \sim F$ and $Y \sim G$, and a vector (X', Y') which is i.i.d. with (X, Y) ,

$$\tau = \tau(X, Y) = \mathbb{E}[\text{sign}((X - X')(Y - Y'))]$$

$$\text{Where } \text{sign}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

We can rewrite the definition of Kendall's Tau to this alternative representation which can provide us a better intuition to its properties.

Lemma 1.3.1 (Hofert et al., 2018) **Alternative Representation by Probability**

By the definition of expectation of indicator functions, Kendall's Tau can be rewritten in this form:

$$\begin{aligned} \tau &= \mathbb{E}[\mathbb{1}\{(X - X')(Y - Y') > 0\} - \mathbb{1}\{(X - X')(Y - Y') < 0\}] \\ &= \mathbb{P}((X - X')(Y - Y') > 0) - \mathbb{P}((X - X')(Y - Y') < 0) \\ &= 2 \mathbb{P}((X - X')(Y - Y') > 0) - 1 \end{aligned}$$

Note that the countermonotonic copula is discordant and the comonotonic copula is concordant. This follows that $\tau \in [-1, 1]$.

We can also rewrite Kendall's Tau based on another representation involving copulas.

Lemma 1.3.2 (Hofert et al., 2018) **Alternative Representation by Copulas**

By Sklar's Theorem, it can be represented in terms of (and only) copulas.

$$\begin{aligned} \tau &= 4 \mathbb{P}(X - X' < 0, Y - Y' < 0) - 1 \\ &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} H(x, y) dH(x, y) - 1 \\ &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} C(F(X), G(Y)) dC(F(X), G(Y)) - 1 \end{aligned}$$

This shows that Kendall's Tau only depends on the underlying copula, instead of the marginal distributions. This shows that Kendall's Tau is invariant under strictly increasing transformations.

Lastly, we can rewrite Kendall's Tau based on a representation involving the correlation coefficient.

Lemma 1.3.3 (Hofert et al., 2018) *Alternative Representation by Correlation*

We can rewrite Kendall's Tau as:

$$\begin{aligned}\tau &= 4 \mathbb{P}(X - X' < 0, Y - Y' < 0) - 1 \\ &= \text{Corr}(\mathbb{1}\{X - X'\}, \mathbb{1}\{Y - Y'\})\end{aligned}$$

This shows that the denominator of the correlation representation is a Bernoulli random variable and the second moment is defined. Hence, this ensures the existence of Kendall's Tau under any distribution.

1.3.4 Significance

These rank correlation measures are an upgrade compared to the Pearson's correlation coefficient: It is capable of handling non-linear dependence and at the same, allow more strict properties to be exhibited on the dataset to be evaluated, for example, existence, invariance and the attainment of $[-1, 1]$ to improve comparability.

Moreover, both measures only depend on the underlying copula. [Hofert et al., 2018](#) also

stated that a lot of these copula families have a one-to-one correspondence between these rank correlation measures and the parameter specifying the copula. Hence, we can deduce the parameter for a specific copula by simply measuring the rank correlation of the dataset.

This is instrumental in our further analysis as we would like to generate datasets (with different underlying copulas) with the same rho or tau to ensure standardisation and comparability across distributions.

1.4 Survival Copulas and Rotated Copulas

In this report, we are investigating both left (low) and right (high) tail dependence, hence generating survival copulas to model right tail is essential. For better modelling capabilities, having an analytical solution for a survival copula (corresponding to the original copula) is pivotal.

Definition 1.4.1 *Survival Copula*

The survival copula is a multivariate survival function: $\bar{F}(\mathbf{x}) = \mathbb{P}(\mathbf{X} > \mathbf{x})$, $\mathbf{x} \in \mathbb{R}$ with marginal distributions $F_j(\bar{x}_j) = \bar{H}(\infty, \dots, x_j, \dots, \infty) = \mathbb{P}(X_j > x_j)$.

Lets extend Sklar's theorem to survival copulas.

Theorem 1.4.1 ([Hofert et al., 2018](#)) *Sklar's Theorem for Survival Copulas*

1. *Existence:* For univariate margins $\bar{F}_1, \bar{F}_2, \dots, \bar{F}_d$, a d -dimensional survival function \bar{F} , $\exists \bar{C} : [0, 1]^d \rightarrow [0, 1]$ such that $\bar{F}(\mathbf{x}) = \bar{C}(\bar{F}_1(x_1), \dots, \bar{F}_d(x_d))$, $\mathbf{x} \in \mathbb{R}^d$, where C is uniquely defined on $\prod_{j=1}^d \text{ran } \bar{F}_j$.

2. *Converse:* Given any survival copula $\bar{C} : [0, 1]^d \rightarrow [0, 1]$ and univariate margins $\bar{F}_1, \bar{F}_2, \dots, \bar{F}_d$, $\bar{F}(\mathbf{x}) = \bar{C}(\bar{F}_1(x_1), \dots, \bar{F}_d(x_d))$ is a d -dimensional survival function, $\mathbf{x} \in \mathbb{R}^d$.

It is important to link survival copulas to their original copulas, which will be illustrated below: (Definition follows [Hofert et al., 2018](#))

Theorem 1.4.2 ([Hofert et al., 2018](#)) *Properties of Survival Copulas*

For $\mathbf{U} \sim C$:

$$1. \mathbf{1} - \mathbf{U} \sim \bar{C}$$

$$2. C \text{ can be recovered by } \bar{C} \text{ by:}$$

$$\bar{C}(\mathbf{u}) = \sum_{J \subseteq \{1, \dots, d\}} (-1)^{|J|} C((1 - u_1)^{\mathbf{1}(1 \in J)}, \dots, (1 - u_d)^{\mathbf{1}(d \in J)}), \quad \mathbf{u} \in [0, 1]^d$$

It can be seen that constructing survival copulas are rather computationally intensive. However, in this report, dimensions are restricted to a rather lower number. Using Appendix Theorem 1.4.2, we will implement an analytical form of survival copulas from the smoothed empirical copulas.

For convenience, we consider rotated copulas as a more generalised form of survival copulas.

Definition 1.4.2 ([Hofert et al., 2018](#)) *Rotated Copulas*

If $\mathbf{U} \sim C$, then we can define rotated copula of C $\text{rot}_{\mathbf{r}}(C)$ by:

$$((1 - r_1)U_1 + r_1(1 - U_1), \dots, (1 - r_d)U_d + r_d(1 - U_d)) \sim \text{rot}_{\mathbf{r}}(C)$$

It can be observed that $\bar{C} = \text{rot}_{\mathbf{1}}(C)$, where $\mathbf{1} \in \mathbb{1}^d$.

This provides an intuitive way to understand the definition of survival copulas, where the flip of C is determined by \mathbf{r} componentwise equal to 1. This implies the survival copula is basically a rotated copula that flips all dimensions. Rotated copulas provide some great properties in the measuring of datasets with opposite dependence. However, this is out of the scope of this report. Though, for optimizing computation, the computation of true survival copulas will be done through rotated copulas.

By understanding and computing an analytical solution for the various types of copulas

(and empirical copulas), this allows us to efficiently and (relatively) more accurately measure tail risk, without the variation induced by the numerical estimation due to observations generated by the underlying / empirical copulas.

2 Appendix: Empirical Copulas

2.1 Empirical beta copula

2.1.1 Empirical Beta Copula: Formulation

Let $\mathbf{W}_i = (W_{i,1}, \dots, W_{i,d})$, $i \in \{1, \dots, n\}$, \mathbf{W} is independent of \mathbf{X} . First obtain \mathbf{R}_i from \mathbf{X}_i .

Let $V_{1,j}, \dots, V_{n,j}$ be the order statistics of $W_{1,j}, \dots, W_{n,j}$, for $j \in \{1, \dots, d\}$. Notice that \mathbf{W}

is standard uniform distributed in its column components, while \mathbf{R} is only asymptotically standard uniform distributed in its column components. We then obtain:

$$\hat{\mathbf{V}}_i = (V_{R_{i,1},1}, \dots, V_{R_{i,d},d})$$

This procedure is basically choosing the corresponding order statistics from a uniform distribution in its components, based on the rank of its corresponding asymptotic distribution.

Theorem 2.1.1 *The order statistics of a collection of i.i.d standard uniform distributed variables is Beta distributed*

Let $W_{1,j}, \dots, W_{n,j} \stackrel{i.i.d.}{\sim} \mathcal{U}(0, 1)$. Then, $V_{R_{k,j},j} \sim \mathcal{B}(R_{k,j}, n+1-R_{k,j})$.

Proof is omitted.

Hence, as each entry of $\hat{\mathbf{V}}$ is beta distributed, it suffices to construct the empirical beta copula (hence the name) similar to the construction of the unsmoothed empirical copula.

$$C_m^\beta(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{P}(V_{R_{i,j},j} \leq u_j) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d F_{n,R_{i,j}}(u_j), \quad \mathbf{u} \in [0, 1]^d$$

2.1.2 The empirical beta copula as a special case of the bernstein copula

It is also shown by [Segers et al., 2017](#) that the empirical beta copula is actually a special case of the empirical bernstein copula $B_m(C_n)$, where $\mathbf{m} = (n, \dots, n)$.

Definition 2.1.1 ([Segers et al., 2017](#)) *Bernstein Polynomial for a d-variate function*

The Bernstein polynomial for a d -variate function, with order \mathbf{m} is denoted as:

$$B_{\mathbf{m}}(f)(\mathbf{u}) = \sum_{s_1=0}^{m_1} \cdots \sum_{s_d=0}^{m_d} f\left(\frac{s_1}{m_1}, \dots, \frac{s_d}{m_d}\right) \prod_{j=1}^d p_{m_j, s_j}(u_j)$$

Where $p_{m_j, s_j}(u_j) = \mathbb{P}(\mathbb{B}(m_j, u_j) = s_j)$ is binomial distributed.

It can be shown that when $f = C$ (copula) then the Bernstein polynomial is a copula.

Proposition 2.1 (Segers et al., 2017) **Bernstein Polynomials are copulas**

Substituting $f : [0, 1]^d$ with $C : [0, 1]^d \rightarrow [0, 1]$, it then follows that:

$$B_{\mathbf{m}}(C)(\mathbf{u}) = \sum_{s_1=0}^{m_1} \cdots \sum_{s_d=0}^{m_d} C\left(\frac{s_1}{m_1}, \dots, \frac{s_d}{m_d}\right) \prod_{j=1}^d p_{m_j, s_j}(u_j)$$

1. (groundedness) When any $u_j = 0$, $j \in \{1, \dots, d\}$, the product term leads to $B_{\mathbf{m}}(C)(u_1, \dots, u_{j-1}, 0, u_j) = 0$
2. (marginal distributions are standard uniformly distributed) When $\mathbf{u} = (1, \dots, 1, u_j, 1, \dots, 1)$, notice that:

$$p_{m_j, s_j}(1) = \mathbb{P}(\mathbb{B}(m_j, 1) = s_j) = \binom{m_j}{s_j} 1^{s_j} 0^{m_j - s_j} = \mathbb{1}\{m_j = s_j\}$$

Hence it follows that:

$$B_{\mathbf{m}}(C)(\mathbf{u}) = \sum_{s_j=0}^{m_j} C(1, \dots, 1, \frac{s_j}{m_j}, 1, \dots, 1) p_{m_j, s_j}(u_j) = \sum_{s_j=0}^{m_j} \frac{s_j}{m_j} p_{m_j, s_j}(u_j) = u_j$$

3. (d -increasing) The Bernstein polynomial is infinitely differentiable (Segers et al., 2017). It is also shown by Segers et al., 2017 that $\frac{\partial^d B_{\mathbf{m}}(C)(\mathbf{u})}{\partial u_1 \dots \partial u_d} > 0$. $B_{\mathbf{m}}(C)(\mathbf{u})$ is d -increasing.

It is shown by Segers et al., 2017 that, not only the Bernstein polynomial, the empirical Bernstein copula with degrees divisible by n is also a copula. This provides further validity that the empirical beta copula, which is a special case (for degrees being n) of the empirical Bernstein copula, is a copula.

Theorem 2.1.2 (Segers et al., 2017) **Empirical Bernstein Copula with orders equal to divisors of n are copulas**

The Empirical Bernstein Copula is

$$B_{\mathbf{m}}(C_n)(\mathbf{u}) = \sum_{s_1=0}^{m_1} \cdots \sum_{s_d=0}^{m_d} C_n\left(\frac{s_1}{m_1}, \dots, \frac{s_d}{m_d}\right) \prod_{j=1}^d p_{m_j, s_j}(u_j).$$

It is trivial that C_n is grounded and d -increasing. To show that the margins are (instead of asymptotically uniform) standard uniformly distributed, consider:

$$B_{\mathbf{m}}(C_n)(\mathbf{u}) = \sum_{s_j=0}^{m_j} C_n(1, \dots, 1, \frac{s_j}{m_j}, 1, \dots, 1) p_{m_j, s_j}(u_j)$$

Invoking the definition of the empirical copula and sufficient algebra yields:

$$\begin{aligned} B_{\mathbf{m}}(C_n)(\mathbf{u}) &= \sum_{s_j=0}^{m_j} \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left\{ \frac{R_{i,j}}{n} \leq \frac{s_j}{m_j} \right\} p_{m_j, s_j}(u_j) \equiv \frac{s_j}{m_j} \\ &\quad \sum_{i=1}^n \mathbb{1} \left\{ \frac{R_{i,j}}{n} \leq \frac{s_j}{m_j} \right\} \equiv \frac{s_j n}{m_j} \end{aligned}$$

Since $\frac{R_{i,j}}{n} \sim \mathcal{U}(\{\frac{1}{n}, \dots, 1\})$,

$$\sum_{i=1}^n \mathbb{1} \left\{ \frac{R_{i,j}}{n} \leq \frac{s_j}{m_j} \right\} = \left\lfloor \frac{s_j n}{m_j} \right\rfloor \rightarrow \left\lfloor \frac{s_j n}{m_j} \right\rfloor \equiv \frac{s_j n}{m_j}$$

The equivalence implies that m_j has to be a divisor of n .

2.1.3 Alternative formulation of the empirical beta copula

Note that the empirical Bernstein copula can be written as a mixture by Segers et al., 2017:

$$B_{\mathbf{m}}(C_n)(\mathbf{u}) = \int_{[0,1]^d} C_n(\mathbf{w}) d\mu_{\mathbf{m},\mathbf{u}}(\mathbf{w})$$

where $\mu_{\mathbf{m},\mathbf{u}}(\mathbf{w})$ is the joint distribution function: $D_{\mathbf{A}}(\mathbf{w})$ where $\mathbf{A} \sim \left(\frac{S_{m_1,u_1}}{n}, \dots, \frac{S_{m_d,u_d}}{n}\right)$ and $S_{m_j,u_j} \sim \mathbb{B}(m_j, u_j)$, $j \in \{1, \dots, d\}$

Let $\mathbf{m} = (n, \dots, n) = \mathbf{n}$. Hence, the empirical beta copula can be written as:

$$B_{\mathbf{n}}(C_n)(\mathbf{u}) = \int_{[0,1]^d} C_n(\mathbf{w}) d\mu_{\mathbf{n},\mathbf{u}}(\mathbf{w})$$

2.2 Smoothed and data-adaptive copulas proposed by Kojadinovic and Yi, 2024

2.2.1 Generalized smoothed empirical copula: Formulations

Notice that we can define a function $\mathbb{K}_{\mathbf{r}}^{\mathbf{x}}(\mathbf{u})$ with useful properties:

Definition 2.2.1 (Kojadinovic & Yi, 2024) *A d-variate “survival function”*

For a vector of non-negative integers \mathbf{r} and a matrix with non-negative entries \mathbf{x} , The function $\mathbb{K}_{\mathbf{r}}^{\mathbf{x}}(\mathbf{u})$ is defined by

$$\mathbb{K}_{\mathbf{r}}^{\mathbf{x}}(\mathbf{u}) = \int_{[0,1]^d} \mathbb{1}\left(\frac{\mathbf{r}}{n} \leq \mathbf{w}\right) dD_{\mathbf{W}_u^{\mathbf{x}}}(\mathbf{w}) = \mathbb{E}\left[\mathbb{1}\left(\frac{\mathbf{r}}{n} \leq \mathbf{W}_u^{\mathbf{x}}\right)\right], \quad \mathbf{u} \in [0,1]^d$$

We notice that, in terms of the random vector $\mathbf{W}_u^{\mathbf{x}}$, it is a d-variate survival function.

$$\mathbb{K}_{\mathbf{r}}^{\mathbf{x}}(\mathbf{u}) = \mathbb{P}\left(\mathbf{W}_u^{\mathbf{x}} \geq \frac{\mathbf{r}}{n}\right)$$

Using the previous section’s notation, we can rewrite $\mathbb{K}_{\mathbf{r}}^{\mathbf{x}}(\mathbf{u})$ as:

$$\mathbb{K}_{\mathbf{r}}^{\mathbf{x}}(\mathbf{u}) = \bar{D}_{\mathbf{W}_u^{\mathbf{x}}}\left(\frac{\mathbf{r}}{n}\right)$$

where \bar{D} is the survival function of the corresponding D if D is absolutely continuous.

Then we can rewrite the generalized smoothed empirical copula:

$$\begin{aligned} C_n^v(\mathbf{u}) &= \int_{[0,1]^d} C_n(\mathbf{w}) dD_{\mathbf{W}_u^{\mathbf{X}}}(\mathbf{w}) = \int_{[0,1]^d} \frac{1}{n} \sum_{i=1}^n \mathbb{1}\left(\frac{\mathbf{R}_i}{n} \leq \mathbf{w}\right) dD_{\mathbf{W}_u^{\mathbf{X}}}(\mathbf{w}) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{[0,1]^d} \mathbb{1}\left(\frac{\mathbf{R}_i}{n} \leq \mathbf{w}\right) dD_{\mathbf{W}_u^{\mathbf{X}}}(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^n \mathbb{K}_{\mathbf{R}_i}^{\mathbf{X}}(\mathbf{u}), \quad \mathbb{K}_{\mathbf{r}}^{\mathbf{X}}(\mathbf{u}) = \mathbb{P}\left(\mathbf{W}_u^{\mathbf{X}} \geq \frac{\mathbf{r}}{n} \mid \mathbf{X}\right). \end{aligned}$$

with a slight condition that the components of \mathbf{W}_u^X need to be discretized (Kojadinovic & Yi, 2024) with values $\{0, \frac{1}{n}, \dots, 1\}$. (If not, a slight adjustment of $-\frac{1}{2}$ has to be imposed on \mathbf{R}_i .)

Theorem 2.2.1 (Kojadinovic & Yi, 2024) *A generalised form of $\mathbb{K}_r^X(\mathbf{u})$*

Consider a d -variate survival function $\bar{F}_{\mathbf{u}}^{\mathbf{x}}(\mathbf{w}) = \mathbb{P}(\mathbf{W}_u^X > \mathbf{w})$, $\mathbf{w} \in [0, 1]^d$.

Note that if the condition of discretized \mathbf{W}_u^X is true, then it follows that

$$\bar{D}_{\mathbf{W}_u^X}\left(\frac{\mathbf{r}}{n}\right) = \bar{F}_{\mathbf{u}}^{\mathbf{x}}\left(\frac{\mathbf{r}-1}{n}\right)$$

By Sklar's Theorem for survival copulas (Hofert et al., 2018, Sklar, 1959), it also follows that there exists a copula $\tilde{C}_{\mathbf{u}}^{\mathbf{x}}$ such that:

$$\begin{aligned}\bar{F}_{\mathbf{u}}^{\mathbf{x}}\left(\frac{\mathbf{r}-1}{n}\right) &= \tilde{C}_{\mathbf{u}}^{\mathbf{x}}\left[\bar{F}_{u_1}^{\mathbf{x}}\left(\frac{r_1-1}{n}\right), \dots, \bar{F}_{u_d}^{\mathbf{x}}\left(\frac{r_d-1}{n}\right)\right], \\ \mathbb{K}_r^{\mathbf{x}}(\mathbf{u}) &= \tilde{C}_{\mathbf{u}}^{\mathbf{x}}\left[\bar{F}_{u_1}^{\mathbf{x}}\left(\frac{r_1-1}{n}\right), \dots, \bar{F}_{u_d}^{\mathbf{x}}\left(\frac{r_d-1}{n}\right)\right].\end{aligned}$$

Hence we can write the smoothed empirical copula as

$$C_n^v(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^{\mathbf{X}}\left[\bar{F}_{u_1}^{\mathbf{X}}\left(\frac{R_{i,1}-1}{n}\right), \dots, \bar{F}_{u_d}^{\mathbf{X}}\left(\frac{R_{i,d}-1}{n}\right)\right].$$

Notice that \mathbf{X} is associated on the copula \tilde{C} . This represents the fact that \tilde{C} might depend on the underlying observations \mathbf{X} (such as the empirical beta copula).

To see if these smoothed empirical copulas outperform the empirical beta copula, we will

include some of the specific smoothed empirical copulas suggested by Kojadinovic and Yi, 2024. For the case where \mathbf{W}_u^X is discrete at its components, we would like to choose the components that take on values in $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$, with $\mathbb{E}(W_{u_j}^X) = u_j$. Before diving into the survival margins for $\bar{F}_{\mathbf{u}}^{\mathbf{x}}$, specify $W_{u_j}^{\mathbf{x}} = \frac{S_{n,u_j}}{n}$. Kojadinovic and Yi, 2024 have specified 2 different discrete distributions that fulfil the property $\mathbb{E}(W_{u_j}^X) = u_j$:

1. Marginal distributions binomial distributed

$$\mathbb{P}(S_{n,u_j} = s) = \binom{n}{s} u_j^s (1-u_j)^{n-s}$$

$$\mathbb{E}[S_{n,u_j}] = n u_j, \quad \text{Var}(S_{n,u_j}) = n u_j (1-u_j)$$

2. Marginal distributions beta-binomial distributed

$$\begin{aligned}\mathbb{P}(S_{n,u_j} = s) &= \binom{n}{s} \frac{\mathcal{B}(s + \alpha, n - s + \beta)}{\mathcal{B}(\alpha, \beta)} \\ \mathbb{E}[W_{u_j}^{\mathbf{X}}] &= u_j = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(W_{u_j}^{\mathbf{X}}) = \rho \frac{u_j(1 - u_j)}{n} \\ \rho &= \frac{\alpha + u_j n}{\alpha + u_j}\end{aligned}$$

Kojadinovic and Yi, 2024 have also proposed a continuous distribution (for continuous $W_{u_j}^{\mathbf{X}}$).

Notice that if $\mathbf{W}_u^{\mathbf{X}}$ is continuous, it then follows that $\bar{D}_{\mathbf{W}_u^{\mathbf{X}}} = \bar{F}_{\mathbf{u}}^{\mathbf{x}}$.

However, the rewritten smoothed empirical copula for continuous distribution i.e.

$$C_n^v(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^{\mathbf{X}} \left[\bar{F}_{u_1}^{\mathbf{X}} \left(\frac{R_{i,1}}{n} \right), \dots, \bar{F}_{u_d}^{\mathbf{X}} \left(\frac{R_{i,d}}{n} \right) \right]$$

does not fulfil the requirement of uniform margins. i.e.

$$C_n^v(1, \dots, 1, u_j, 1, \dots, 1) = \mathbb{E} \left[\frac{\lfloor nW_{u_j}^{\mathbf{X}} \rfloor}{n} \middle| \mathbf{X} \right] < \mathbb{E} \left[W_{u_j}^{\mathbf{X}} \middle| \mathbf{X} \right] = u_j$$

A reasonable adjustment in this case is to add a modification to $R_{i,j}$:

$$C_n^v(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^{\mathbf{X}} \left[\bar{F}_{u_1}^{\mathbf{X}} \left(\frac{R_{i,1} - 0.5}{n} \right), \dots, \bar{F}_{u_d}^{\mathbf{X}} \left(\frac{R_{i,d} - 0.5}{n} \right) \right].$$

The proposed continuous distribution is shown below:

1. Marginal distributions beta distributed

$$\begin{aligned}\mathbb{E}[W_{u_j}^{\mathbf{X}}] &= u_j = \frac{\alpha}{\alpha + \beta} \\ \text{Var}(W_{u_j}^{\mathbf{X}}) &= \rho \frac{u_j(1 - u_j)}{n} \\ \rho &= \frac{n u_j}{\alpha + u_j}\end{aligned}$$

Furthermore, we need to propose a reasonable copula $\tilde{C}_{\mathbf{u}}^{\mathbf{X}}$.

In Appendix Theorem 2.2.1, the smoothed empirical copula is written as

$$\begin{aligned}C_n^v(\mathbf{u}) &= \frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^{\mathbf{X}} \left[\bar{F}_{u_1}^{\mathbf{X}} \left(\frac{R_{i,1}-1}{n} \right), \dots, \bar{F}_{u_d}^{\mathbf{X}} \left(\frac{R_{i,d}-1}{n} \right) \right], \\ \mathbb{E}[C_n^v(\mathbf{u})] &= E \left[\frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^{\mathbf{X}} \left[\bar{F}_{u_1}^{\mathbf{X}} \left(\frac{R_{i,1}-1}{n} \right), \dots, \bar{F}_{u_d}^{\mathbf{X}} \left(\frac{R_{i,d}-1}{n} \right) \right] \right].\end{aligned}$$

When we analyse the components of $\tilde{C}_{\mathbf{u}}^{\mathbf{X}}$ (Kojadinovic & Yi, 2024) we can observe that:

Lemma 2.2.1 (Kojadinovic & Yi, 2024) *Expectation of $\bar{F}_{u_j}^{\mathbf{X}} \left(\frac{R_{i,j}-1}{n} \right)$*

We can show that for all $j = \{1, \dots, d\}$:

$$\mathbb{E} \left[\bar{F}_{u_j}^{\mathbf{X}} \left(\frac{R_{i,j}-1}{n} \right) \right] = u_j, \quad u_j \in [0, 1]$$

Proof:

Taking expectation of $\bar{F}_{u_j}^{\mathbf{X}} \left(\frac{R_{i,j}-1}{n} \right)$:

$$\mathbb{E} \left[\bar{F}_{u_j}^{\mathbf{X}} \left(\frac{R_{i,j}-1}{n} \right) \right] = \mathbb{E} \left[\mathbb{E} \left[\bar{F}_{u_j}^{\mathbf{X}} \left(\frac{R_{i,j}-1}{n} \right) \middle| \mathbf{X} \right] \right]$$

Recall that $R_{i,j} \sim \mathcal{U}(\{ \frac{1}{n}, \dots, 1 \})$

$$\mathbb{E} \left[\mathbb{E} \left[\bar{F}_{u_j}^{\mathbf{X}} \left(\frac{R_{i,j}-1}{n} \right) \middle| \mathbf{X} \right] \right] = \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n \bar{F}_{u_j}^{\mathbf{X}} \left(\frac{k-1}{n} \right) \right] = \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n P \left(W_{u_j}^{\mathbf{X}} > \frac{k-1}{n} \middle| \mathbf{X} \right) \right]$$

As $W_{\mathbf{u}}^{\mathbf{X}}$ is also discretized componentwise:

$$\begin{aligned}\mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n P \left(W_{u_j}^{\mathbf{X}} > \frac{k-1}{n} \middle| \mathbf{X} \right) \right] &= \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n P \left(W_{u_j}^{\mathbf{X}} \geq \frac{k}{n} \middle| \mathbf{X} \right) \right] = \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[\mathbb{1}\{nW_{u_j}^{\mathbf{X}} \geq k\} \middle| \mathbf{X} \right] \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\frac{\lfloor nW_{u_j}^{\mathbf{X}} \rfloor}{n} \middle| \mathbf{X} \right] \right] = \mathbb{E} [u_j] = u_j\end{aligned}$$

Then it suffices to observe that:

$$\mathbb{E}[C_n^v(\mathbf{u})] = \frac{1}{n} \sum_{i=1}^n \tilde{C}_{\mathbf{u}}^{\mathbf{X}}(u_1, \dots, u_d) = \tilde{C}_{\mathbf{u}}^{\mathbf{X}}(u_1, \dots, u_d) = \tilde{C}_{\mathbf{u}}^{\mathbf{X}}(\mathbf{u})$$

We could let the survival copula $\tilde{C}_{\mathbf{u}}^{\mathbf{X}}$ to be similar to the true copula to obtain good performance. Therefore, a good proxy is $\tilde{C}_{\mathbf{u}}^{\mathbf{X}} = C_n^\beta$, and this is used in the models we test in this work.