Thomas Kuenzer BSc

Construction of New Multidimensional Copulas using the CM-Algorithm

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Financial and Actuarial Mathematics

submitted to

Graz University of Technology

Supervisor

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Eranda Dragoti-Çela

Institute of Discrete Mathematics

Graz, July 2016

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

Date

Signature

Konstruktion neuer mehrdimensionaler Copulas mithilfe des CM-Algorithmus

Kurzfassung

Copulas sind ein nützliches und weitverbreitetes Konzept, um Abhängigkeiten von Zufallsvariablen darzustellen. Die Copulamodelle, die derzeit in Standardanwendungen benutzt werden, sind im Hinblick auf Symmetrie und Parameteranzahl einigen Beschränkungen unterworfen. Deshalb ist die Entwicklung neuer Modelle ein zentrales Anliegen der Forschung. Bei Aktienkursen beobachtet man, dass sinkende Preise stärker miteinander korrelieren als steigende Preise. Diese Asymmetrie findet sich nicht in den viel verwendeten Copulas wieder, die oft symmetrisch sind. Wir verwenden den CM-Algorithmus, um eine neue Familie von Copulas zu entwickeln: die Panik-Copulas. Sie werden aus zwei unabhängigen Verteilungen für den ruhigen Markt und den panikgetriebenen, hochkorrelierten Markt konstruiert. Wenn die Panikverteilung eine gewisse Panikschwelle unterschreitet, wird ihr Wert verwendet, ansonsten der Wert des ruhigen Marktes. Die Copula dieser Verteilung ist asymmetrisch und bietet zusätzliche Flexibilität bei der Modellierung des unteren Teils der Verteilung. Unser Copulamodell kann in kleinen bis mittleren Dimensionen verwendet werden. Die praktische Anwendung der neuen Modelle zeigen wir anhand von realen Marktdaten. Die Anpassung des Modells wird numerisch durch Hill-Climbing-Optimierung einer modifizierten Likelihoodfunktion vorgenommen; die Ränder werden nichtparametrisch geschätzt. Das so angepasste Modell wird mit den gängigen Goodness-of-Fit-Statistiken untersucht und passt signifikant besser zu den Marktdaten als die t-Copula, obwohl es nur wenige zusätzliche Parameter enthält. Mit dem neuentwickelten Modell berechnen wir den Value-at-Risk und Expected Shortfall von Aktienportfolios. Die Ergebnisse zeigen die gewünschte höhere Abhängigkeit im unteren Teil der Verteilung.

Abstract

Copulas are a widespread and convenient concept of modelling dependence of random variables. Currently, the copula models used in standard applications have restrictions in terms of symmetry or number of parameters in the model. Therefore, developing new models of copulas is a central concern of contemporary research. In market data, it has been observed that price drops correlate more strongly with each other than increases do. This is not reflected in the commonly used copulas that are often symmetric. We use the CM-algorithm to develop a new family of copulas, the panic copulas. They are constructed by using two independent distributions for the calm market and the panic-stricken market where the variables are highly correlated. If the panic distribution reaches below a panic threshold, the random variable realises this value. If not, the calm market distribution is used. By extracting the copula of this distribution, we gain an asymmetric kind of copula with additional flexibility in modelling the lower end of the distribution. Our copula model can be applied in cases of small to medium high dimensions. We show a practical application of the new models by fitting the panic copula models to real market data. The fitting is done numerically with a hill-climbing routine using a modified maximum likelihood function, while the margins are estimated non-parametrically. The resulting model is evaluated using common goodness-of-fit statistics and it turns out to fit the data significantly better than the t-copula in many cases while involving only few additional parameters. Using the new copula model, the Value-at-Risk and expected shortfall of stock portfolios can be calculated by means of Monte Carlo simulation. The results show the desired property of higher dependence in the lower part of the distribution.

Contents

List of Figures

List of Tables

List of Algorithms

Introduction

The word *copula* is originally a Latin noun meaning link or tie. The term has found its way into the fields of logic and linguistics, describing a word that links the subject of a clause to the predicate. In a mathematical context, copulas are multivariate distribution functions on the unit cube. While this is the most easily understandable definition, the origin of their name becomes apparent only after we look at them from another point of view. Copulas are functions joining one-dimensional distribution functions together to form a multivariate distribution function. They describe the dependence structure between random variables, *linking* together the marginal distributions.

While copulas had appeared in numerous contexts of mathematical publications in the years before, it was only in 1959 that those linking functions actually received a name, when Abe Sklar used the name in a letter to Fréchet, apparently drawing inspiration from the usage of the noun as a grammatical term. The study of copulas started becoming popular beginning in 1981 and experienced growing interest in the 1990s, when the field underwent significant evolution and copulas were increasingly proposed for applications in statistics and probability. In 1999, Roger B. Nelsen published the book "An introduction to copulas" on this subject. It has become the standard work on copulas and a classical reference in this field. Apart from their theoretical impact in studying forms of dependence between random variables that are scale-independent, copulas are very useful in constructing multivariate probability distributions by specifying the marginal distribution and their dependence structure separately. When applying copulas to practical problems, there are several computational difficulties to overcome. Therefore, the first copula models to find their way into practical application were the simpler ones, such as the Gaussian copula.

In the aftermath of the global financial crisis 2008-2009, the usage of copulas was partly blamed for the financial models of that time not accurately reflecting reality, favouring excessive risk-taking in the financial market and thus facilitating the emergence of a crisis. In particular, the Gaussian copula used for correlation of defaults in the ambits of credit risk management had to take the blame. In the early 2000s, its use surged in popularity following a paper (see [33]) on default correlation by David X. Li, because it was an easy and quick approach to depict correlation of different default times in-

side the model. Li's model came to be used widely in the credit derivative market and was one of the forces to enable the rapid growth of this market in the years before the crisis. The Gaussian copula and its contribution to the financial crisis were extensively discussed in the media and popular press, see [41] and [26]. Forbes went as far as to decry it as "formula from hell" in their column [31]. The criticism on copula models certainly went too far in some aspects, but it became evident that the models implemented during the run-up to the financial crisis were faulty and relied on too much simplification of the reality. In fact, the limitations of the model had been well-known before the crisis and had even been pointed to by the Wall Street Journal in [48]. However, these claims that were coming mainly from the academic world were largely ignored or dismissed as being of exclusively academic nature.

It has since become clear that while the usage of copulas is not to be indiscriminately condemned, one needs to pay more attention to the details. The classical approach in finance relies on the static copula theory, which does not pay special attention to the dynamic aspect of dependence. Additionally, the widespread Gaussian copula also underestimates the frequency of correlated extreme events. Therefore, the field of copula theory is currently branching out in different aspects: dynamic copulas are being studied as a way to model dependence of stochastic processes. On the other hand, the static copula models in use need to be refined to map the real data better into the mathematical model. The models used need to pay more attention to outliers in the empirical data. Thus, the search for new copula types has been recognised as an important task of statistics and probability theory.

In this thesis, we will take a look at copula theory in general, fitting procedures for copulas and model evaluation. Furthermore, we will specifically look at the computational approach to copula implementation proposed by Meucci in [35]. We will then go on and construct a new type of asymmetrical copula, the *panic copula*, that places importance on the lower tail of the distribution in order to more accurately reflect the empirical observation of higher correlation of losses compared to profits. We develop a general model of the panic copula and propose methods to fit the model to real data, including the medium-dimensional case, i.e. up to 17 dimensions. We apply the developed model to real market data, calculating the Value-at-Risk and the expected shortfall for portfolios of several stocks. We compare the results with the currently widespread and well-known models and evaluate their additional benefits. The implementation is done in Matlab, using the publicly available implementation of the CM-algorithm by Meucci.

Chapter 1

Preliminaries

1.1 Elliptical Distributions

The theory for elliptical distributions is based on [13].

Definition 1.1. A random vector $X \in \mathbb{R}^d$ has a spherical distribution if for all orthogonal $U \in \mathbb{R}^{d\times d}$,

$$UX \overset{d}{=} X.$$

That is, the distribution of X is distributionally indifferent under rotations. We then write $X \sim S_d(\psi)$, where ψ is the characteristic generator of the distribution.

Theorem 1.2. *These characterisations are equivalent:*

i) X has a spherical distribution

ii) The characteristic function ϕ_X of X is given by a function $\psi : \mathbb{R} \to \mathbb{R}$ and

$$\phi_X(t) = \psi(t^\top t) = \psi(t_1^2 + \cdots + t_d^2),$$

for all $t \in \mathbb{R}^d$.

iii) For all $a \in \mathbb{R}^d$,

$$a^\top X \overset{d}{=} \|a\| \cdot X_1,$$

where $\|a\|^2 = a^\top a$.

Proof. i) $\Rightarrow$ ii). Because for all orthogonal U, we have

$$\phi_X(t) = \phi_{UX}(t) = \phi_X(U^\top t),$$

this means that ϕ_X can only depend on the length of t, i.e. $\|t\| = \sqrt{t^\top t}$.
ii) $\Rightarrow$ iii). For $t \in \mathbb{R}$,

$$\begin{aligned}\phi_{a^\top X}(t) &= \phi_X(t\ a) = \psi(t^2 a^\top a) \\ &= \psi((t\|a\|e_1)^\top (t\|a\|e_1)) = \phi_{X_1}(t\|a\|) = \phi_{\|a\|X_1}(t).\end{aligned}$$

iii) $\Rightarrow$ i). Let U be an orthogonal matrix and $t \in \mathbb{R}^d$, then

$$\begin{aligned}\phi_{UX}(t) &= \mathbb{E}[\exp(i(U^\top t)^\top X)] = \mathbb{E}[\exp(i\|U^\top t\|X_1)] \\ &= \mathbb{E}[\exp(i\|t\|X_1)] = \mathbb{E}[\exp(it^\top X)] = \phi_X(t).\end{aligned}$$

□

Theorem 1.3. *X has a spherical distribution if and only if*

$$X \stackrel{d}{=} RS, \tag{1.1}$$

where $R \geq 0$ and S are independent random variables and S is uniformly distributed on the unit sphere $\{s \in \mathbb{R}^d : s^\top s = 1\}$.

Proof. If $X \stackrel{d}{=} RS$ with the given random vector S and scalar random variable R, then we have for all orthogonal U

$$\phi_{UX}(t) = \mathbb{E}[\exp(it^\top USR)] = \mathbb{E}[\exp(it^\top SR)] = \phi_X(t),$$

therefore X is spherical.

The other direction can be seen in [13]. □

Definition 1.4. $X \in \mathbb{R}^d$ has an elliptical distribution if

$$X \stackrel{d}{=} \mu + A\,Y, \tag{1.2}$$

where $\mu \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d\times k}$ are constant and $Y \sim S_k(\psi)$. We write $X \sim E_d(\mu, \Sigma, \psi)$, with $\Sigma = A^\top A$. μ is called the location vector, Σ is the dispersion matrix and ψ is the characteristic generator of the distribution.

Provided that the variances are finite, it is possible to find a representation $X \sim E_d(\mu, \Sigma, \psi)$ such that Σ is the covariance matrix of X.

Proposition 1.5. *The characteristic function of elliptical $X = \mu + AY$ with $Y \sim S_k(\psi)$ is*

$$\phi_X(t) = e^{it^\top \mu}\psi(t^\top \Sigma t).$$

Corollary 1.6. *X has an elliptical distribution if and only if*

$$X \stackrel{d}{=} \mu + RAS, \tag{1.3}$$

where $R \geq 0$ is a random variable and S is independent from R and uniformly distributed on the unit sphere $\{s \in \mathbb{R}^d : s^\top s = 1\}$. $A \in \mathbb{R}^{d\times k}$ with $A^\top A = \Sigma$.

Proposition 1.7. *For $X \sim E_d(\mu_1, \Sigma, \psi_1)$ and $Y \sim E_d(\mu_2, \Sigma, \psi_2)$,*

$$X + Y \sim E_d(\mu_1 + \mu_2, \Sigma, \psi_1 \cdot \psi_2).$$

Proof. For the characteristic function of $X + Y$, we can see that

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t) = e^{it^\top(\mu_1+\mu_2)}\ \psi_1(t^\top \Sigma t)\psi_2(t^\top \Sigma t).$$

□

From any covariance matrix Σ, we can calculate the *correlation matrix* R by the following normalisation:

$$R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_i \cdot \Sigma_j}}$$

The entries of this matrix are often also referred to using the notation ρ_{ij}.

Proposition 1.8. *Let the random vector $X \sim E_2(0, \Sigma, \psi)$ and ρ be the (linear) correlation between the two components. If $\mathbb{P}(X = 0) = 0$, then*

$$\mathbb{P}(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{\arcsin \rho}{2\pi}.$$

Proof. By standardisation, we get $Y \sim E_2(0, P, \psi)$ with $P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\mathbb{P}(X_1 > 0, X_2 > 0) = \mathbb{P}(Y_1 > 0, Y_2 > 0)$ with Y being spheric. For $Z \sim S_2(\psi)$, we have

$$\begin{aligned}(Y_1, Y_2) &\overset{d}{=} (Z_1, \rho Z_1 + \sqrt{1-\rho^2} Z_2) \\ &\overset{d}{=} R\ (\cos\Theta, \rho\cos\Theta + \sqrt{1-\rho^2}\sin\Theta),\end{aligned}$$

with random variable $R \geq 0$ and Θ uniformly distributed on $[-\pi, \pi)$. If we write $\rho = \sin\phi$, then

$$\begin{aligned}\mathbb{P}(X_1 > 0, X_2 > 0) &= \mathbb{P}(\cos\Theta > 0,\ \sin\phi\cos\Theta + \cos\phi\sin\Theta > 0) \\ &= \mathbb{P}(\cos\Theta > 0, \sin(\Theta + \phi) > 0).\end{aligned}$$

Therefore, Θ must be in the intersection of $(-\pi/2, \pi/2)$ and $(-\phi, \pi - \phi)$. We see that

$$\mathbb{P}\big(\Theta \in (-\pi/2, \pi/2) \cap (-\phi, \pi - \phi)\big) = \frac{\frac{1}{2}\pi + \phi}{2\pi}.$$

□

1.2 Multivariate t-Distribution

A multivariate t-distributed vector $X \in \mathbb{R}^d$ is constructed by

$$X = \mu + \frac{Y}{\sqrt{V/\nu}}, \quad Y \sim N_d(0, \Sigma), \quad V \sim \chi^2_\nu, \quad \mu \in \mathbb{R}^d,\ \Sigma \in \mathbb{R}^{d\times d},\ \nu > 0,$$

where Y and V are independent and Σ is a positive definite covariation matrix. The density function of $X \sim St_d(\mu, \Sigma, \nu)$ is then for $x \in \mathbb{R}^d$

$$f(x|\mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\nu\pi)^{d/2}\sqrt{\det(\Sigma)}} \left(1 + \frac{1}{\nu}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)^{-\frac{d+\nu}{2}}.$$

Similar to the one-dimensional case, we have $\mathbb{E}(X) = \mu$ for $\nu > 1$ and $\mathrm{Var}(X) = \frac{\nu}{\nu-2}\Sigma$ for $\nu > 2$.

If we define the random vector $Z = AX + b$ with $A \in \mathbb{R}^{k\times d}$ and $b \in \mathbb{R}^k$, then $Z \sim St_k(A\mu + b, A\Sigma A^\top, \nu)$. Thus, if we partition the random vector into two random variables $X^\top = (X_1^\top, X_2^\top)$ with dimensions d_1 and d_2, then the two parts are again t-distributed with $X_i \sim St_{d_i}(\mu_i, \Sigma_{ii}, \nu),\ i = 1, 2$. The conditional distribution of $X_1|X_2$ is:

$$\begin{aligned} X_1|X_2 \ &\sim \ St_{d_1}(\mu_{1|2}, \Sigma_{1|2}, \nu + d_2), \\ \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \\ \Sigma_{1|2} &= \frac{\nu + (x_2-\mu_2)^\top \Sigma_{22}^{-1}(x_2-\mu_2)}{\nu + d_2}\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^\top\right). \end{aligned} \tag{1.4}$$

The proof for this can be seen in [39]. Taking the limit for $\nu \to \infty$, we get the analogous results for the multivariate normal distribution.

Chapter 2

Copulas

It is intuitive to think of random vectors as having two separate traits. Firstly, we see how every component behaves *on its own*. Then, we observe how they *depend on each other*. The former finds its mathematical counterpart in the marginal distribution functions. The latter has long been only reflected in statistical measures such as the covariance or the correlation. Copulas are a very useful way to think of the two fundamental properties of random vectors in a concise way. Copulas permit splitting up the joint distribution function into its margins and a function that describes exhaustively their dependence, so to speak.

The theory presented in this chapter is based on the books [13] and [37].

2.1 Definition

A d-dimensional copula is a cumulative distribution function C of a random vector on $[0,1]^d$ with uniform margins. An equivalent definition will follow, but first we have to make some technical definitions. Cumulative distribution functions (cdf) will be referred to as distribution functions; probability density functions (pdf) will be referred to as density functions.

Definition 2.1. Given a function $F : S_1 \times \cdots \times S_d \to \mathbb{R}$,

1. F is called grounded if for all $1 \leq k \leq d$, with a_k being the least element in S_k,

$$F(x_1, \ldots, x_{k-1}, a_k, x_{k+1}, \ldots, x_d) = 0,$$

   for all $x \in S_1 \times \cdots \times S_d$ where $x_k = a_k$.

2. F is called d-increasing if for $a_k, b_k \in S_k$ with $a_k \leq b_k$, $\forall 1 \leq k \leq d$,

$$\sum_{x \in \bigotimes_{k=1}^{d} \{a_k, b_k\}} (-1)^{s(x)} F(x) \geq 0, \qquad \text{with } s(x) = \#\{k : x_k = a_k\}.$$

Before we can actually define the concept of a copula, it is advisable to define the concept of a *sub-copula*. Sub-copulas already have *some* properties a copula must have, but allow for more flexibility in the sense that they do not have to be defined on the whole unit cube.

Definition 2.2. A d-dimensional subcopula is a function $C' : S_1 \times \cdots \times S_d \to \mathbb{R}$ which fulfils

i) $\{0,1\} \subseteq S_1, S_2 \subseteq [0,1]$,

ii) C' is grounded and d-increasing,

iii) For $u_k \in S_k,\ 1 \leq k \leq d, \quad C'(1,\ldots,1,u_k,1,\ldots,1) = u_k$.

In this definition, the property of groundedness can be replaced by monotonicity in every component. From the definition, it is immediately clear that all subcopulas will have values in $[0,1]$ only.

Proposition 2.3. *Let C' have margins $C'(1,\ldots,1,u_k,1,\ldots,1) = u_k$ and be d-increasing, then the following holds: C' is grounded if and only if C' is non-decreasing in every variable.*

Proof. "$\Rightarrow$" Let C' be grounded. Let $u \in S$, $1 \leq k \leq d$ and $0 \leq x \leq y \leq 1$. We show that C' is non-decreasing in the k-th variable. Define

$$a = (0,\ldots,0,x,0,\ldots,0)^\top, \qquad b = (u_1,\ldots,u_{k-1},y,u_{k+1},\ldots,u_d)^\top.$$

Using these points a and b in the inequality of C' being d-increasing, we can see that all but two terms are zero because C' is grounded. It follows that

$$C'(u_1,\ldots,u_{k-1},y,u_{k+1},\ldots,u_d) - C'(u_1,\ldots,u_{k-1},x,u_{k+1},\ldots,u_d) \geq 0.$$

"$\Leftarrow$" Let C' be non-decreasing in every variable. Let $u \in S$ such that $u_k = 0$. Because of the monotonicity, we have

$$C'(u) \leq C'(1,u_2,\ldots,u_d) \leq \cdots \leq C'(1,\ldots,1,u_k,1,\ldots,1) = u_k = 0$$

□

The following lemma will provide us with important results for Sklar's theorem in the next section.

Lemma 2.4. *For $1 \leq k \leq d$, let S_k be a non-empty subset of $\bar{\mathbb{R}} = \mathbb{R} \uplus \{\pm\infty\}$ and $S = \otimes_{k=1}^d S_k$. Let H be a grounded and d-increasing function on S with margins denoted by $F_k(x_k) = H(\infty,\ldots,\infty,x_k,\infty,\ldots,\infty)$. Then for every two points $x,\ y \in S$,*

$$|H(x) - H(y)| \leq \sum_{k=1}^{d} |F_k(x_k) - F_k(y_k)|\,. \tag{2.1}$$

Proof.

$$
\begin{aligned}
|H(x) - H(y)| &\leq \sum_{k=1}^{d} |H(x_1, \ldots, x_k, y_{k+1}, \ldots, y_d) - H(x_1, \ldots, x_{k-1}, y_k, \ldots, y_d)| \\
&\leq \sum_{k=1}^{d} |F_k(x_k) - F_k(y_k)| \,,
\end{aligned}
$$

where the first inequality is an application of the triangle inequality over a telescopic sum. The second line follows from noting that since H is grounded and d-increasing, H is non-decreasing in every component. Assuming w.l.o.g. that $x_k \geq y_k$, it follows that

$$
0 \leq H(x_1, \ldots, x_k, y_{k+1}, \ldots, y_d) - H(x_1, \ldots, x_{k-1}, y_k, \ldots, y_d) \leq F_k(x_k) - F_k(y_k),
$$

where the right hand side of the inequality stems from iteratively applying the property of H being d-increasing as follows:

$$
a = (0, \ldots, 0, y_k, 0, \ldots, 0, y_d)^\top, \qquad b = (x_1, \ldots, x_k, y_{k+1}, \ldots, y_{d-1}, 1)^\top
$$

This then leads to

$$
\begin{aligned}
&H(x_1, \ldots, x_k, y_{k+1}, \ldots, y_d) - H(x_1, \ldots, x_{k-1}, y_k, \ldots, y_d) \\
&\quad \leq H(x_1, \ldots, x_k, y_{k+1}, \ldots, y_{d-1}, 1) - H(x_1, \ldots, x_{k-1}, y_k, \ldots, y_{d-1}, 1).
\end{aligned}
$$

This iteration is then continued analogously for the remaining components. □

Because all sub-copulas satisfy the assumptions of this lemma, we can conclude that every sub-copula is a Lipschitz-continuous function satisfying

$$
|C'(u) - C'(v)| \leq \|u - v\|_1.
$$

Finally, we can define the copula as a special case of a sub-copula.

Definition 2.5. A copula is a subcopula which is defined on $[0,1]^d$. That is, a copula is a function $C : [0,1]^d \to [0,1]$, such that

(i) C is grounded, i.e. for all $u \in [0,1]^d$ and all $1 \leq k \leq d$

$$
C(u_1, \ldots, u_{k-1}, 0, u_{k+1}, \ldots, u_d) = 0, \tag{2.2}
$$

(ii) the one-dimensional margins are the identity function, i.e.

$$
C(1, \ldots, 1, u_k, 1, \ldots, 1) = u_k, \ \forall 1 \leq k \leq d, \tag{2.3}
$$

(iii) C is d-increasing, i.e. for all $a, b \in [0,1]^d$ such that $a \leq b$ componentwise

$$\sum_{u \in \bigotimes_{k=1}^{d} \{a_k, b_k\}} (-1)^{s(u)} C(u) \geq 0, \qquad \text{with } s(u) = \#\{k : u_k = a_k\}. \tag{2.4}$$

The last property (2.4) is also called the rectangle inequality. All distribution functions on $\mathbb{R}^d$ are d-increasing. This property ensures that for a random vector $U \sim C$ and over every d-dimensional "rectangle" $R = [a_1, b_1] \times \cdots \times [a_d, b_d]$, the probability $\mathbb{P}(U \in R)$ is non-negative, see [7].

Definition 2.6. If C is an n-dimensional copula and we have the subset $\{k_1, \ldots, k_d\} \subset \{1, \ldots, n\}$, then the d-dimensional margin along the indices $\{k_1, \ldots, k_d\}$ is defined as

$$C_{k_1,\ldots,k_d}(u_1, \ldots, u_d) = C(v), \qquad \text{with } v_j = \begin{cases} u_i & \text{if } j = k_i, \\ 1 & \text{otherwise,} \end{cases}$$

which is again a copula according to its definition.

There are many different families of copula functions. Some of them and their properties are described in Section 2.5. It is inevitable that some types of copulas will be used in examples before they are clearly defined in a later section.

2.2 Properties

Additionally to the properties included in the definition, we already know from Lemma 2.4 that all copulas are (Lipschitz-) continuous functions. Furthermore, we can establish pointwise upper and lower bounds on the values of a copula C.

Theorem 2.7 (Fréchet Bounds)**.** *Every copula C satisfies*

$$\max(u_1 + \cdots + u_d - d + 1, 0) \leq C(u) \leq \min(u_1, \ldots, u_d), \quad \forall u \in [0,1]^d.$$

While the lower Fréchet bound is no copula for $d > 2$, the upper bound is. The bounds are the best possible because for every $u \in [0,1]^d$, there exists a copula that reaches the lower bound in the point u.

Proposition 2.8. *The upper Fréchet bound*

$$C^+(u) = \min(u_1, \ldots, u_d)$$

is a copula for all $d \in \mathbb{N}$.

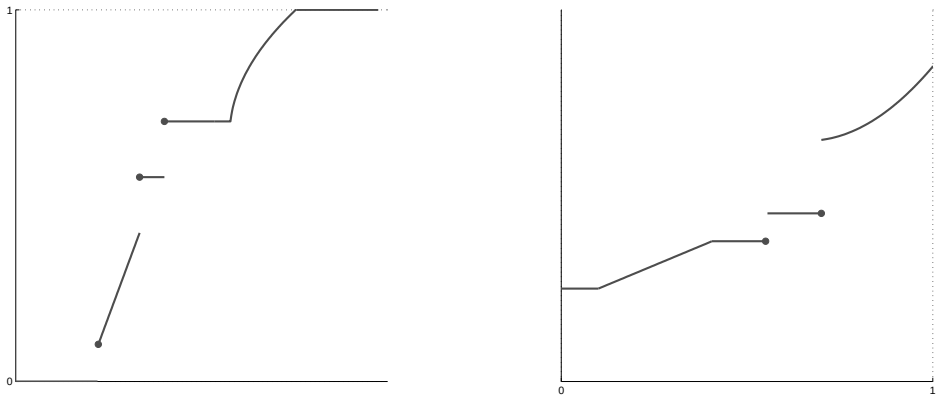

Figure 2.1: An example of a distribution function (left) and its general inverse (right). Note that the regular inverse function does not exist in this case.

Proof. This follows easily from the following construction: Let $U_0 \in [0,1]$ be uniformly distributed. Define the mutually completely dependent random vector $U = (U_0, U_0, \ldots, U_0)^\top \in [0,1]^d$. Since U has uniform margins, its distribution function C is a copula. Trivially, it follows that $\mathbb{P}(U \leq u) = \min(u_1, \ldots, u_d)$. □

In order to get an understanding of how copulas and joint distribution functions are related, we have to define the generalised inverse of a distribution function. This is needed because not every distribution function is *strictly* increasing, which rules out the existence of the classical inverse function.

Definition 2.9. The generalised inverse of a non-decreasing function G is defined by

$$G^{\leftarrow}(y) = \inf\{x : G(x) \geq y\}.$$

If G is strictly increasing, then the generalised inverse coincides with the regular inverse function G^{-1} on the domain of G^{-1}. For continuous G, $G(G^{\leftarrow}(u)) = u$ holds. For further properties of the generalised inverse, see [13].

Proposition 2.10. *Let G be a distribution function, $X \sim G$ and $U \sim U(0,1)$ a standard uniform distributed random variable. Then*

i) $\mathbb{P}(G^{\leftarrow}(G(X)) = X) = 1$.

ii) Quantile transformation: $\mathbb{P}(G^{\leftarrow}(U) \leq x) = G(x)$

iii) Probability transformation: If G is a continuous distribution function, then $G(X) \sim U(0,1)$.

Proof. The proof follows from the properties of the generalised inverse:

i) As a distribution function, G is non-decreasing. It is clear that

$$G^{\leftarrow}(G(z)) = \inf\{w : G(w) \geq G(z)\} \leq z$$

If $G^{\leftarrow}(G(z)) < z$, then $G(w) = G(z)$ holds on the whole interval $[G^{\leftarrow}(G(z)), z]$. The probability that X falls on such an interval of constant G is zero. Therefore, $G^{\leftarrow}(G(X)) = X$ almost surely.

ii) Because $G(y) \geq u \Leftrightarrow y \geq G^{\leftarrow}(u)$, we see that

$$\mathbb{P}(G^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq G(x)) = G(x)$$

iii) Because G is continuous, $G^{\leftarrow}$ is strictly increasing. Therefore,

$$\begin{aligned}\mathbb{P}(G(X) \leq u) &= \mathbb{P}(G^{\leftarrow}(G(X)) \leq G^{\leftarrow}(u)) \overset{\text{i)}}{=} \mathbb{P}(X \leq G^{\leftarrow}(u)) \\ &= G(G^{\leftarrow}(u)) \overset{G \text{ cont.}}{=} u.\end{aligned}$$

□

We come now to the heart of copula theory. The following theorem by Abe Sklar explains concisely how joint distribution functions can be decomposed into marginal distributions and copula. It also marks the first time that the word *copula* was used to describe a function linking one-dimensional marginal distribution functions together to form a multivariate distribution function.

Theorem 2.11 (Sklar 1959)**.** *Let F be a d-dimensional joint distribution function with its margins denoted by $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \tag{2.5}$$

for all $x \in [-\infty, \infty]^d$. If the margins are continuous, C is unique. In any case, C is uniquely determined on $\otimes_{k=1}^{d} F_k(\bar{\mathbb{R}})$.

Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, equation (2.5) *defines a joint distribution function with the given margins $F_1, \ldots, F_d$.*

Proof. "⇒" The proof consists in two parts. First, we show that there exists a unique subcopula fulfilling the equality, and then we extend the subcopula to a copula.

F matches the conditions of Lemma 2.4, so for any two points $x, y \in \bar{\mathbb{R}}^d$, we see that if $F_k(x_k) = F_k(y_k)$ for all $1 \leq k \leq d$, then $F(x) = F(y)$. Thus, we can uniquely define a function C' from the set of ordered pairs

$$\left\{ \Big(\big(F_1(x_1), \ldots, F_d(x_d)\big), F(x_1, \ldots, x_d) \Big) | x \in \bar{\mathbb{R}}^d \right\}.$$

C' is defined on $\otimes_{k=1}^{d} F_k(\bar{\mathbb{R}})$. It is indeed a subcopula, as the conditions follow directly from the properties of the distribution function F. Subsequently, C' fulfils by construction

$$C'(F_1(x_1), \ldots, F_d(x_d)) = F(x_1, \ldots, x_d).$$

If the margins are continuous, then the range of the marginal distribution functions is $[0, 1]$ and C' is itself a copula.

In the contrary case, the subcopula C' can be extended to a copula by linear interpolation in each component. For two dimensions, this can be read in [37]. As the extension is not unique, C is not unique in the case of non-continuous margins.

"$\Leftarrow$" If C is a copula and $F_1, \ldots, F_d$ are given univariate distribution functions, it is easy to see that the resulting function F has the margins F_k. The conditions for F being indeed a multivariate distribution function follow directly from the properties of C. □

The function C defined in Theorem 2.11 will be called the copula of the distribution F from now on. It is defined uniquely for continuous margins by

$$C(u_1, \ldots, u_d) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)), \tag{2.6}$$

which is the distribution function of $(F_1(X_1), \ldots, F_d(X_d))$ if $X \sim F$, as proven in the probability transformation of Proposition 2.10.

As easy as the statement of Sklar's theorem looks, the proof shows that many things are to be considered. In fact, the proof was not provided in the original publication in 1959. This resulted in years of people working in this field having to reconstruct the proof on their own. In the case of non-continuous distributions, the problem of non-uniqueness arises. It is still possible to give pointwise bounds on the values of C for discrete distribution cases. But effectively, the problem of the lack of uniqueness runs so deep that fundamental problems emerge. For example, the independence of the components of a random vector is no longer equivalent to the independence copula $C^{\perp}$ being the copula of the vector. The independence copula (or product copula) is defined as

$$C^{\perp}(u_1, \ldots, u_d) = \prod_{k=1}^{d} u_k.$$

This absence of equivalence is of course in at its very core a cosmetic problem that stems from the ambiguity of the values of C in points where it will not be evaluated because of the discrete margins. Therefore, the copula behind the distribution of X is essentially not identifiable any more.

Example 1. Let X be two-dimensional and $X \sim F$. If F has continuous margins, the following holds: X_1 and X_2 are independent if and only if their copula is $C^{\perp}(u_1, u_2) = u_1 u_2$.

"$\Leftarrow$" is trivial. "$\Rightarrow$", if X_1 and X_2 are independent, we have $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$. Therefore by Sklar's theorem, $C^\perp$ is a possible copula. Because the margins of X are continuous, we additionally see that $C^\perp$ is indeed the *unique* copula of X.

If the margins of X are not continuous, we may provide a counterexample: let X_1 and X_2 be independent Bernoulli variables with $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 0) = \frac{1}{2}$. Then the copula $C(u_1, u_2) = \frac{1}{2}(\max(u_1 + u_2 - 1, 0) + \min(u_1, u_2))$ satisfies $C(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4}$, as does the independence copula.

But the complications that discrete margins cause don't stay on this superficial level. No property of the continuous case should be blindly taken to be valid in the case of discrete margins. Further problems of the discrete case will be discussed later on. An in-depth investigation of this vast field of problems that arise in conjunction with discrete margins can be found in [20].

We come now to a basic copula property that is the indifference to monotonous transformations. Since we see the copula as the sole dependence part of a joint distribution, it shouldn't be surprising that transforming each component via a strictly monotonous function will not alter the copula.

Proposition 2.12. *If $X \in \mathbb{R}^d$ is a random vector with continuous margins and $T_1, \dots, T_d$ are strictly increasing functions, then X and its transformed $(T_1(X_1), \dots, T_d(X_d))$ have the same copula C.*

Proof. Because X_k has a continuous distribution function, X_k has no point masses. Therefore, the strictly increasing transformation $Y_k = T_k(X_k)$ does not, either. Furthermore, the distribution function of Y_k is $\tilde{F}_k = F_k \circ T_k^{-1}$. Thus,

$$\begin{aligned} C_X(u_1, \dots, u_d) &= \mathbb{P}(F_1(X_1) \le u_1, \dots, F_d(X_d) \le u_d) \\ &= \mathbb{P}(F_1 \circ T_1^{-1}(T_1(X_1)) \le u_1, \dots, F_d \circ T_d^{-1}(T_d(X_d)) \le u_d) \\ &= \mathbb{P}(\tilde{F}_1(Y_1) \le u_1, \dots, \tilde{F}_d(Y_d) \le u_d) \\ &= C_Y(u_1, \dots, u_d). \end{aligned}$$

□

Now that we have seen that strictly increasing transformations of the margins have no effect on the copula, it is only natural to ask what happens with strictly decreasing transformations. The concept that arises in this context is the *survival copula*.

Definition 2.13. If C is a copula, i.e. it is the distribution function of $(U_1, \dots, U_d)$ with standard uniform margins, its survival copula $\tilde{C}$ is the distribution function of $(1 - U_1, \dots, 1 - U_d)$. If $\tilde{C} = C$, we call C (radially) symmetric.

Survival copulas $\tilde{C}$ are not to be confused with survival functions of copulas $\bar{C}$. The difference lies in

$$\bar{C}(u_1, \ldots, u_d) = \mathbb{P}(U_1 > u_1, \ldots, U_d > u_d),$$
$$\tilde{C}(u_1, \ldots, u_d) = \mathbb{P}(1 - U_1 \leq u_1, \ldots, 1 - U_d \leq u_d) = \bar{C}(1 - u_1, \ldots, 1 - u_d).$$

In contrast to the survival copula $\tilde{C}$, the survival function of a copula $\bar{C}$ itself is not a copula. Please note that the notation $\tilde{C}$ is different from the more common way of referring to the survival copula. We are using a tilde instead of the usual hat to avoid notational conflicts with the empirical copula estimator defined in Definition 2.41.

Example 2. If C is a three-dimensional copula, then its survival function $\bar{C}$ and its survival copula $\tilde{C}$ are given as follows:

$$\begin{aligned}\bar{C}(u_1, u_2, u_3) &= 1 - u_1 - u_2 - u_3 \\ &\quad + C_{12}(u_1, u_2) + C_{13}(u_1, u_3) + C_{23}(u_2, u_3) - C(u_1, u_2, u_3), \\ \tilde{C}(u_1, u_2, u_3) &= -2 + u_1 + u_2 + u_3 + C_{12}(1 - u_1, 1 - u_2) + C_{13}(1 - u_1, 1 - u_3) \\ &\quad + C_{23}(1 - u_2, 1 - u_3) - C(1 - u_1, 1 - u_2, 1 - u_3).\end{aligned}$$

The generalisation in higher dimensions is obvious.

Because of Proposition 2.12, it is clear that if we want to extract the copula of a multivariate normal distribution $N_d(\mu, \Sigma)$, the parameter μ is irrelevant and Σ can be taken as a correlation matrix. Furthermore, it follows from Definition 2.13 that $\tilde{C} = C$ in the case of the Gauss copula.

Proposition 2.14. *For the survival copula $\tilde{C}$ of a random vector X with survival function $\bar{F}$, an analogous version of Sklar's theorem applies:*

$$\bar{F}(x_1, \ldots, x_d) = \tilde{C}(\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)) \tag{2.7}$$

Proof. If the margins are continuous, the equation is easily derived:

$$\begin{aligned}\bar{F}(x_1, \ldots, x_d) &= \mathbb{P}(X_1 > x_1, \ldots, X_d > x_d) \\ &= \mathbb{P}(F_1(X_1) \geq F_1(x_1), \ldots, F_d(X_d) \geq F_d(x_d)) \\ &= \mathbb{P}(1 - F_1(X_1) \leq \bar{F}_1(x_1), \ldots, 1 - F_d(X_d) \leq \bar{F}_d(x_d)) \\ &= \tilde{C}(\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)).\end{aligned}$$

□

Proposition 2.15. *If a random vector X with continuous margins has a copula C, then $-X$ has the copula $\tilde{C}$.*

Proof. This can be seen from:

$$\begin{aligned} C_{-X}(u_1,\dots,u_d) &= F_{-X}(F^{\leftarrow}_{-X_1}(u_1),\dots,F^{\leftarrow}_{-X_d}(u_d)) \\ &= F_{-X}(-F^{\leftarrow}_{X_1}(1-u_1),\dots,-F^{\leftarrow}_{X_d}(1-u_d)) \\ &= \bar{F}_X(F^{\leftarrow}_{X_1}(1-u_1),\dots,F^{\leftarrow}_{X_d}(1-u_d)) \\ &\overset{(2.7)}{=} \tilde{C}_X(u_1,\dots,u_d). \end{aligned}$$

□

In fact, if T_k are strictly decreasing transformations, then Y obtained by $Y_k = T_k(X_k)$ has the copula $\tilde{C}$. This follows easily from Proposition 2.12.

A property which is important in practical applications of copulas is *exchangeability*, i.e. whether or not the copula treats all components the same. It is also possible that some components have the same dependence structure within their group and to components outside of it; they then form an exchangeable group.

Definition 2.16. We call a copula C exchangeable if

$$C(u_1,\dots,u_d) = C(u_{\pi(1)},\dots,u_{\pi(d)}), \tag{2.8}$$

for all permutations $\pi : \{1,\dots,d\} \to \{1,\dots,d\}$. We call $I \subseteq \{1,\dots,d\}$ an exchangeable group of copula C if (2.8) holds for all permutations π that only permutate elements of I and have $\{1,\dots,d\}\backslash I$ as fixed points.

As is usual with probability distributions, many copulas also have a density function.

Definition 2.17. The density of a copula C is given by

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

under the assumption that C is sufficiently differentiable.

Proposition 2.18. *The density of the survival copula $\tilde{C}$ is*

$$\tilde{c}(u_1,\dots,u_d) = \frac{\partial^d C(1-u_1,\dots,1-u_d)}{\partial u_1 \dots \partial u_d} = c(1-u_1,\dots,1-u_d),$$

under the assumption that C is sufficiently differentiable.

Proposition 2.19. *If X is a d-dimensional random vector with distribution function F and density function f, F_i and $f_i > 0$ respectively for the margins $1 \le i \le d$, then the density c of its copula C is computed directly by*

$$c(u_1,\dots,u_d) = \frac{f(F_1^{-1}(u_1),\dots,F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}. \tag{2.9}$$

Proof. By differentiating equation (2.6) and taking the ordinary inverse of the marginal distribution functions. This can be done because the marginal densities are strictly positive and therefore all F_i are strictly increasing. □

Using this equation, we can theoretically compute the copula density of every joint distribution we have the densities of.

Example 3. The Gaussian copula is defined as the copula of the multivariate normal distribution. Using (2.9), we see that the copula density of a two-dimensional Gaussian copula with correlation ρ is

$$c(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2\Phi^{-1}(u_1)^2 - 2\rho\Phi^{-1}(u_1)\Phi^{-1}(u_2) + \rho^2\Phi^{-1}(u_2)^2}{2(1-\rho^2)}\right),$$

where Φ^{-1} is the inverse distribution function of a standard normal distribution.

2.3 Measures of Dependence

2.3.1 Kendall's Tau

One of the best known measures of dependence (or more precisely: measures of concordance) is Kendall's tau. If (X_1, X_2) and (Y_1, Y_2) are independent and identically distributed with distribution function F, Kendall's tau is defined as

$$\tau(X_1, X_2) = \mathbb{P}((X_1 - Y_1)(X_2 - Y_2) > 0) - \mathbb{P}((X_1 - Y_1)(X_2 - Y_2) < 0). \quad (2.10)$$

This is the probability of concordance minus the probability of discordance. It is obvious that this measure does not depend on the margins of the joint distribution but only on the copula.

Proposition 2.20. *If X_1 and X_2 have continuous marginal distributions, τ can also be calculated directly from their copula.*

$$\tau(C) = 4\int_{[0,1]^2} C(u, v)dC(u, v) - 1 \quad (2.11)$$

Proof. From the definition and noting that the distribution is continuous, we can write

$$\begin{aligned}
\tau(X_1, X_2) &= 2\mathbb{P}((X_1 - Y_1)(X_2 - Y_2) > 0) - 1 \\
&= 4\mathbb{P}(X_1 < Y_1, X_2 < Y_2) - 1 \\
&= 4\mathbb{E}[\mathbb{P}(X_1 < Y_1, X_2 < Y_2|Y_1, Y_2)] - 1 \\
&= 4\int_{\mathbb{R}^2} \mathbb{P}(X_1 < y_1, X_2 < y_2)dF(y_1, y_2) - 1 \\
&= 4\int_{\mathbb{R}^2} C(F_1(y_1), F_2(y_2))dC(F_1(y_1), F_2(y_2)) - 1,
\end{aligned}$$

and substitute the marginal distribution functions $F_i(y_i) = u_i$. □

In the case of an Archimedean copula, τ can be obtained from the generator ϕ_C. To measure pairwise dependence, τ of a d-dimensional random vector X (with its independent copy Y) can be written in short form as

$$\tau(X) = \text{Cov}(\text{sgn}(X - Y)),$$

which is then a $d \times d$ matrix.

The sample version of Kendall's tau between two sample vectors X and Y is:

$$\hat{\tau}(X, Y) = \frac{2}{N(N-1)} \sum_{n=1}^{N} \sum_{m=n+1}^{N} \text{sgn}((X_n - X_m)(Y_n - Y_m)) \tag{2.12}$$

Theorem 2.21. *Kendall's tau of a copula and its survival copula are the same.* $\tau(C) = \tau(\tilde{C})$.

Proof. This follows trivially from Definition 2.13. □

2.3.2 Spearman's Rho

Another measure of concordance is the rank correlation coefficient Spearman's rho. If (X_1, X_2), (Y_1, Y_2) and (Z_1, Z_2) are independent and identically distributed with distribution function F, Spearman's rho is defined as

$$\rho_S(X_1, X_2) = 3\left(\mathbb{P}((X_1 - Y_1)(X_2 - Z_2) > 0) - \mathbb{P}((X_1 - Y_1)(X_2 - Z_2) < 0)\right). \tag{2.13}$$

Proposition 2.22. *If X_1 and X_2 have continuous marginal distributions, ρ_S can be calculated from the copula C of X_1, X_2:*

$$\rho_S(C) = 12 \int_{[0,1]^2} C(u, v) du dv - 3, \tag{2.14}$$

which is the linear correlation between U_1 and U_2 of the copula, i.e.

$$\rho_S(X_1, X_2) = Corr(F_1(X_1), F_2(X_2)).$$

Proof. From the definition and noting that the distribution is continuous, we can write

$$\begin{aligned}
\rho_S(X_1, X_2) &= 3 \cdot \Big(2\mathbb{P}((X_1 - Y_1)(X_2 - Z_2) > 0) - 1\Big) \\
&= 12\, \mathbb{P}(X_1 < Y_1, X_2 < Z_2) - 3 \\
&= 12\, \mathbb{E}[\mathbb{P}(X_1 < Y_1, X_2 < Z_2 | Y_1, Z_2)] - 3 \\
&= 12 \int_{\mathbb{R}^2} F(y_1, z_2) dF_1(y_1) dF_2(z_2) - 3 \\
&= 12 \int_{[0,1]^2} C(u, v) du dv - 3,
\end{aligned}$$

which proves (2.14).

For two random variables Z_1, Z_2 with joint distribution function G and margins $G_1,\ G_2$, the covariance is given by Höffding's formula

$$\mathrm{Cov}(Z_1, Z_2) = \int_{[0,1]^2} (G(z_1, z_2) - G_1(z_1)G_2(z_2))\, dz_1 dz_2.$$

A proof of this equation can be found in [13]. Plugging $F_1(X_1)$ and $F_2(X_2)$ into this formula yields

$$\mathrm{Cov}(F_1(X_1), F_2(X_2)) = \int_{[0,1]^2} (C(u, v) - uv)\, dudv.$$

Formula (2.14) follows then by dividing by the variance of the uniformly distributed margins. □

Theorem 2.23. *Spearman's rho of a copula and its survival copula are the same.* $\rho_S(C) = \rho_S(\tilde{C})$.

The sample version is obtained by either replacing the marginal distribution function above with the empirical estimate

$$\hat{\rho}_S(X, Y) = \frac{\widehat{\mathrm{Cov}}(\hat{F}_X(X), \hat{F}_Y(Y))}{\sqrt{\widehat{\mathrm{Var}}(\hat{F}_X(X)) \cdot \widehat{\mathrm{Var}}(\hat{F}_Y(Y))}},$$

or using the ascending rank statistics of the sample:

$$R_i = \mathrm{rank}(X_i), \qquad S_i = \mathrm{rank}(Y_i).$$

Then we get:

$$\begin{aligned} \hat{\rho}_S(X, Y) &= \frac{12}{N(N^2-1)} \left(\sum_{n=1}^{N} R_n \cdot S_n - \frac{N(N+1)^2}{4} \right) \\ &= 1 - 6\frac{\sum_{n=1}^{N}(R_n - S_n)^2}{N(N^2-1)} \end{aligned} \tag{2.15}$$

2.3.3 Relation between Kendall's Tau and Spearman's Rho

Since both τ and ρ_S aim to measure similar things, it is only natural that they have some connection between them. In fact, the relationship between them can be summed up by the following bounds:

$$\begin{cases} \frac{3}{2}\tau - \frac{1}{2} \leq \rho_S \leq \frac{1}{2} + \tau - \frac{1}{2}\tau^2, & \text{if } \tau \geq 0, \\ -\frac{1}{2} + \tau + \frac{1}{2}\tau^2 \leq \rho_S \leq \frac{3}{2}\tau + \frac{1}{2}, & \text{if } \tau < 0. \end{cases} \tag{2.16}$$

The region within these bounds is shown in Figure 2.2. The inequalities are not always sharp, i.e. not all points (τ, ρ_S) on the bounds can be reached.

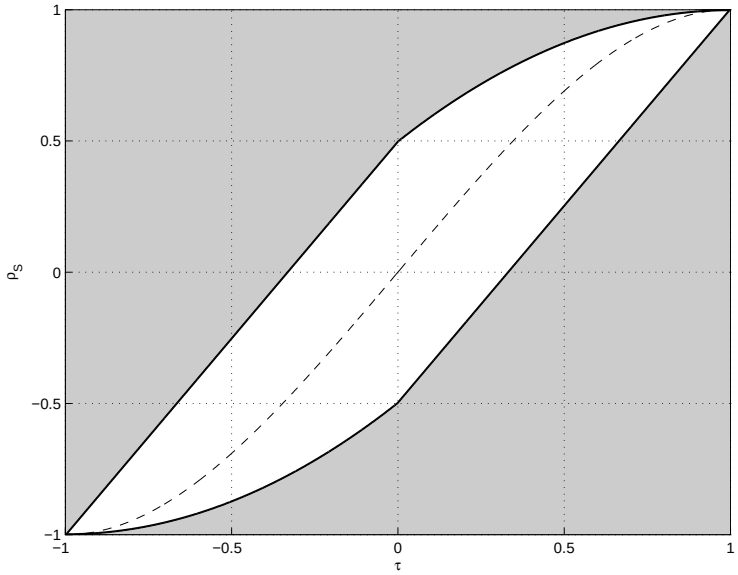

Figure 2.2: The bounds of the τ-ρ_S-region expressed in inequality (2.16). The dashed line shows the points attainable by a Gaussian copula.

For positive τ, the lower bound of ρ_S can always be reached. The upper bound can only be reached on the points $(\tau, \rho_S) = (1 - \frac{2}{n}, 1 - \frac{2}{n^2}),\ n \geq 2$. Therefore, we can presume that there exists a function Φ that interpolates those points and represents the sharp bounds of the region. For each point on the sharp bounds of the τ-ρ_S-region, there exist a pair of mutually completely dependent random variables X and Y that match the exact values of τ and ρ_S. The proof of this along with the exact function Φ can be read in [43]. In essence, Y is completely dependent on X if Y almost surely takes only one value for every value of X. The formal definition is as follows:

Definition 2.24. Let X and Y be two random variables. Y is completely dependent on X if there exists a function f such that the set $\{(x, f(x))|\ x \in \mathrm{Dom}(X)\}$ is measurable and has probability 1. That is, $Y = f(X)$ with probability 1. If X is also completely dependent on Y, then we call X and Y mutually completely dependent.

This definition is taken from [30], where further insights into this matter can be found. The copulas on the boundary of the τ-ρ_S-region represent mutually completely dependent random variables. Examples of such copulas can be seen in Figure 2.3.

The limiting cases of this τ-ρ_S-region are mutually completely dependent and as such more of theoretical interest since they are not likely to occur in a setting with real data. Furthermore, the task of reaching an arbitrary point in the interior is still difficult. In practice, the frequently used copula

families only attain points in a very small subset of the τ-ρ_S-region, as noted by [12].

We have seen how to analytically calculate τ and ρ_S using the integral formulas (2.11) and (2.14). In cases where this is not feasible because we have no closed form of C, it is possible to resort to Monte Carlo simulations in order to gain approximate values. With some copula families, however, there is no need for either calculation because the values of τ and ρ_S can be expressed as functions of the correlation of the components.

Theorem 2.25. *For d-dimensional normal distributed $X \sim N_d(\mu, \Sigma)$, looking at the components X_i and X_j for $1 \leq i \neq j \leq d$, we see that*

$$\begin{aligned} \tau(X_i, X_j) =& \frac{2}{\pi} \arcsin(R_{ij}), \\ \rho_S(X_i, X_j) =& \frac{6}{\pi} \arcsin\left(\frac{R_{ij}}{2}\right), \qquad \text{with } R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}, \end{aligned} \tag{2.17}$$

where R_{ij} is the linear correlation between X_i and X_j.

Proof. Because we are only looking at two components of the random vector X, we can assume X is two-dimensional. τ and ρ_S are only dependent on the copula. Therefore, w.l.o.g. we also assume that $X \sim N_2(0, P)$ with $P = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. For Spearman's rho, we look at the probability $\mathbb{P}((X_1 - Y_1) > 0, (X_2 - Z_2) > 0)$, where X, Y and Z are independent copies of X. It is then clear that $W = (X_1 - Y_1, X_2 - Z_2)^\top \sim N_2(0, P + I_2)$. Because of Proposition 1.8, we have that

$$\begin{aligned} \rho_S(X_1, X_2) &= 12\, \mathbb{P}(W_1 > 0, W_2 > 0) - 3 = 12\left(\frac{1}{4} + \frac{\arcsin \rho(W)}{2\pi}\right) - 3 \\ &= \frac{6}{\pi} \arcsin \frac{R_{12}}{2}. \end{aligned}$$

□

Using the empirical estimates for ρ and (2.17), estimation of the correlation matrix of a Gaussian copula is possible. Empirically, these values are close to those obtained by maximum likelihood estimation, according to [3].

Theorem 2.26. *For d-dimensional X from an elliptical distribution, looking at the components X_i and X_j for $1 \leq i \neq j \leq d$, we see that*

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(R_{ij}), \tag{2.18}$$

where R_{ij} is the correlation between X_i and X_j.

Proof. W.l.o.g. assume that X is two-dimensional. We look at the probability $\mathbb{P}((X_1 - Y_1) > 0, (X_2 - Y_2) > 0)$, where X and Y are independent copies of X. Define $W = X - Y$. Because of Proposition 1.7, we know that $W \sim E_2(0, \Sigma, \psi^2)$. From Proposition 1.8, it follows that

$$\begin{aligned}\tau(X_1, X_2) &= 4\,\mathbb{P}(W_1 > 0, W_2 > 0) - 1 = 4 \cdot \left(\frac{1}{4} + \frac{\arcsin \rho(W)}{2\pi}\right) - 1 \\ &= \frac{2}{\pi} \arcsin R_{12}.\end{aligned}$$

□

From the proof, it is easily seen why the formula for Kendall's tau of elliptical distributions from Theorem 2.26 has no analogous version for Spearman's rho. In order to apply Proposition 1.7, the two elliptical random variables need to have the same dispersion parameter Σ. This is not the case for the calculation of Spearman's rho.

The restriction to *continuous* margins of X is not to be taken lightly. It is crucial to almost all statements and conclusions in this section. In fact, even fundamental properties are no longer valid in the case of discrete marginal distributions, as can be seen in [20]. In the ambits of τ and ρ_S, this begins with the difficulty of the definition that assumes the probability of coinciding values is zero. The consequence is that the equivalence of the probabilistic definition (2.10) and the analytical integral representation (2.11) of Kendall's tau (and respectively (2.13) and (2.14) for Spearman's rho) is no longer valid without the assumption of continuous margins. Even worse than that, the concordance measures may become margin-dependent in the case of discrete margins.

Example 4. We will show an example of margin-dependence of τ and ρ_S in the case of discrete margins of X. Let X_1 and X_2 be Bernoulli distributed with $\mathbb{P}(X_1 = 0) = p$ and $\mathbb{P}(X_2 = 0) = q$. The copula of X is C, therefore we have $\mathbb{P}(X_1 = 0, X_2 = 0) = C(p, q) =: r$. Then $\tau(X) = \rho_S(X) = C(p, q) - pq$. The calculation for Kendall's tau is straightforward:

$$\begin{aligned}\tau(X) &= \mathbb{P}((X_1 - Y_1)(X_2 - Y_2) = 1) - \mathbb{P}((X_1 - Y_1)(X_2 - Y_2) = -1) \\ &= \mathbb{P}((X = (1,1) \wedge Y = (0,0)) \uplus (X = (0,0) \wedge Y = (1,1))) \\ &\quad - \mathbb{P}((X = (1,0) \wedge Y = (0,1)) \uplus (X = (0,1) \wedge Y = (1,0))) \\ &= 2\,(\mathbb{P}(X = (1,1))\mathbb{P}(Y = (0,0)) - \mathbb{P}(X = (1,0))\mathbb{P}(Y = (0,1))) \\ &= 2\,((1 - p - q + r)r - (q - r)(p - r)) = r - pq.\end{aligned}$$

The result for ρ_S can be obtained analogously. Because τ and ρ depend on the actual choice of p and q, they are not independent from the margins. Kendall's tau and Spearman's rho are in general *no copula properties* in the case of discrete margins.

2.4 Tail Dependence

While Kendall's tau and Spearman's rho look at the dependence between a pair of random variables X_1 and X_2 along their whole domain, it is also of interest to see how extreme values play together. That is, how likely is it for X_2 to reach a value in its tail given that X_1 realises a tail value. The limit $u \to 0$ of this very conditional probability $\mathbb{P}(X_2 \leq F_{X_2}^{-1}(u) | X_1 \leq F_{X_1}^{-1}(u))$ is the tail dependence. The tail dependence is a measure of asymptotic dependence. Thus, a tail dependence equal to 0 means that we have asymptotic independence of X_1 and X_2. That is, extreme events in the tails occur asymptotically independently from each other. Because the property of tail dependence is invariant under monotonous transformation, it is a copula property. We define it using the copula. The equivalence to the mentioned conditional probability can easily be verified.

Definition 2.27. Let C be a two-dimensional copula, then the upper and lower tail dependence are defined as the limit of the quantile dependence function of C. That is

$$\lambda_U(C) = \lim_{u \to 1^-} \frac{\bar{C}(u,u)}{1-u}, \tag{2.19}$$

$$\lambda_L(C) = \lim_{u \to 0^+} \frac{C(u,u)}{u}, \tag{2.20}$$

if the limits exist. It is evident that $\lambda_U(C) = \lambda_L(\tilde{C})$. Formula (2.19) can easily be evaluated by noting that $\bar{C}(u,u) = 1 - 2u + C(u,u)$.

For many copula families, we have a closed form for the tail dependence. In particular, the tail dependence of the t-copula and the Gaussian copula are given in the following theorem.

Theorem 2.28. *For the multivariate t-distributed vector $X \sim St_d(\mu, \Sigma, \nu)$, $1 \leq i \neq j \leq d$ and linear correlation $R_{ij} > -1$, the tail dependence is:*

$$\lambda_U(X_i, X_j) = \lambda_L(X_i, X_j) = 2\, \bar{t}_{\nu+1}\left(\sqrt{\nu+1}\frac{\sqrt{1-R_{ij}}}{\sqrt{1+R_{ij}}}\right) \tag{2.21}$$

The tail dependence of normally distributed random vectors is zero.

Proof. The idea for the proof can be found in [13]. □

2.5 Copula Families

There are two important branches of copula families. The elliptical copulas and the Archimedean copulas. We will introduce them both and highlight their advantages and disadvantages.

2.5.1 Elliptical Copulas

Elliptical copulas are obtained by Sklar's theorem from elliptical distributions. Most notably, we have the Gauss-copula that is the copula of a multidimensional normal distribution $X \sim N_d(0, \Sigma)$. It has the positive definite correlation matrix $\Sigma \in \mathbb{R}^{d \times d}$ as an input, i.e. $\binom{d}{2}$ parameters. The Gauss copula has no tail dependence. For $\Sigma = I_d$, the Gaussian copula reaches $C^{\perp}$. For Σ being a matrix of ones, the limiting case C^{+} is reached.

The t-copula is the copula of a multivariate t-distributed random vector $X \sim St_d(0, \Sigma, \nu)$. In comparison with the Gauss copula, it has one additional parameter, i.e. the degrees of freedom ν. Its tail dependence is in general non-zero and given by (2.21). Because the t-distribution converges to the normal distribution for $\nu \to \infty$, the Gauss copula is a limiting case of the t-copula.

For both the t-copula and the Gauss copula, the parameters can be easily estimated by applying Theorems 2.25 or 2.26. All elliptical copulas are radially symmetric with $\tilde{C} = C$. Therefore, they have the same upper and lower tail dependence. Their advantages are the abundant number of parameters and the easy estimation thereof.

Elliptical distributions and, in consequence, elliptical copulas are typically easy to be simulated from. Their rank correlation and tail dependence is easily calculated, see [14]. The drawback is that there is no closed form for C. They are restricted to radial symmetry and are not able to exhaustively reproduce the behaviour of real data, as has been noted by [17]. Nonetheless, they remain very popular in practical applications.

2.5.2 Archimedean Copulas

An Archimedean copula is not derived from an underlying distribution by means of Sklar's theorem, but it is itself a probability distribution derived from a generator function.

Definition 2.29. Let $\varphi : [0, 1] \to [0, \infty]$ be a continuous, strictly decreasing function with $\varphi(1) = 0$. We define the pseudo-inverse of φ by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases}$$

From the definition, the pseudo-inverse $\varphi^{[-1]}$ is continuous and decreasing. $\varphi^{[-1]}(\varphi(t)) = t$ holds in any case.

$$\varphi(\varphi^{[-1]}(t)) = \begin{cases} t, & 0 \leq t \leq \varphi(0), \\ \varphi(0), & \varphi(0) \leq t \leq \infty. \end{cases}$$

If $\varphi(0) = \infty$, then the pseudo-inverse and the inverse coincide.

Theorem 2.30. *Let $\varphi : [0,1] \to [0,\infty]$ be a continuous, strictly decreasing function with $\varphi(1) = 0$. Define $C : [0,1]^2 \to [0,1]$ by*

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)).$$

Then C is a copula if and only if φ is convex.

Proof. First, we show that for $u_1 \leq u_2$,

$$C(u_2, v) - C(u_1, v) \leq u_2 - u_1 \tag{2.22}$$

holds if and only if φ is convex. If we set $a = \varphi(u_1)$, $b = \varphi(u_2)$ and $c = \varphi(v)$, then this equation (2.22) is equivalent to

$$\varphi^{[-1]}(a) + \varphi^{[-1]}(b + c) \leq \varphi^{[-1]}(b) + \varphi^{[-1]}(a + c),$$

which, for $a \neq b$, is equivalent to

$$\frac{\varphi^{[-1]}(a) - \varphi^{[-1]}(b)}{a - b} \leq \frac{\varphi^{[-1]}(a + c) - \varphi^{[-1]}(b + c)}{a - b}.$$

This holds if φ (and therefore, $\varphi^{[-1]}$) is convex. In the other direction: if (2.22) holds, then for $0 \leq s < t$ and $a = (s + t)/2, b = s$ and $c = (t - s)/2$,

$$\varphi^{[-1]}\left(\frac{s + t}{2}\right) \leq \frac{\varphi^{[-1]}(s) + \varphi^{[-1]}(t)}{2}.$$

Because $\varphi^{[-1]}$ is continuous, it follows that $\varphi^{[-1]}$ is convex.

What is left to show is that (2.22) is equivalent to C being a copula. If C is a copula, inequality (2.22) is contained in the Lipschitz continuity of C, i.e. $|C(x) - C(y)| \leq \|x - y\|_1$. Showing the other direction is more involved.

It is easy to show the boundary conditions since

$$C(0, u_2) = \varphi^{[-1]}(\varphi(0) + \varphi(u_2)) = 0$$

holds because of monotonicity and

$$C(1, u_2) = \varphi^{[-1]}(\varphi(1) + \varphi(u_2)) = \varphi^{[-1]}(\varphi(u_2)) = u_2.$$

We still have to show the 2-monotonicity of C.

For $u_1 < u_2$ and $v_1 < v_2$, we know that because of continuity, there is a t such that $C(t, v_2) = v_1$, i.e. $\varphi(t) + \varphi(v_2) = \varphi(v_1)$.

$$\begin{aligned}
C(u_2, v_1) - C(u_1, v_1) &= \varphi^{[-1]}(\varphi(u_2) + \varphi(v_1)) - \varphi^{[-1]}(\varphi(u_1) + \varphi(v_1)) \\
&= \varphi^{[-1]}(\varphi(u_2) + \varphi(v_2) + \varphi(t)) \\
&\quad - \varphi^{[-1]}(\varphi(u_1) + \varphi(v_2) + \varphi(t)) \\
&= C(C(u_2, v_2), t) - C(C(u_1, v_2), t) \\
&\overset{(2.22)}{\leq} C(u_2, v_2) - C(u_1, v_2).
\end{aligned}$$

□

φ is called the Archimedean copula generator of C. If $\varphi(0) = \infty$, then φ is a strict Archimedean copula generator.

Proposition 2.31. *Let C be an Archimedean copula with generator φ, then*

i) C is symmetric: $C(u, v) = C(v, u)$.

ii) C is associative: $C(C(u, v), w) = C(u, C(v, w))$.

iii) $C(u, u) < u$ for all $0 < u < 1$.

Proof. i) follows from the definition. As for ii),

$$\begin{aligned} C(C(u,v),w) &= \varphi^{[-1]}(\varphi(\varphi^{[-1]}(\varphi(u)+\varphi(v)))+\varphi(w)) \\ &= \varphi^{[-1]}(\varphi(u)+\varphi(v)+\varphi(w)) = \cdots = C(u, C(v,w)). \end{aligned}$$

For iii), we easily see that

$$C(u,u) = \varphi^{[-1]}(2\varphi(u)) < \varphi^{[-1]}(\varphi(u)) = u.$$

□

Theorem 2.32. *If C is an associative copula with $C(u, u) < u$ for all $0 < u < 1$, then C is Archimedean*

Proof. Can be found in [34]. □

Proposition 2.33. *The Kendall distribution function of a copula C is the distribution function of $C(U, V)$ where $(U, V)^\top \sim C$. If C is an Archimedean copula with generator φ, then its Kendall distribution function is calculated by*

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}, \qquad 0 \le t \le 1.$$

K_C can also be used to test distributional hypotheses on a given dataset in a goodness-of-fit setting, as will be briefly discussed in Section 3.5.

Theorem 2.34. *Let C be an Archimedean copula with generator φ, then Kendall's tau can be calculated by*

$$\tau(C) = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt. \tag{2.23}$$

Proof. For $(U, V)^\top \sim C$, let us denote by K_C the distribution function of $C(U, V)$. Using the integral formula for τ, we have

$$\begin{aligned} \tau(C) &= 4\mathbb{E}[C(U,V)] - 1 = 4\int_0^1 t dK_C(t) - 1 = 3 - 4\int_0^1 K_C(t)dt \\ &= 3 - 4\int_0^1 \left(t - \frac{\varphi(t)}{\varphi'(t^+)}\right)dt \\ &= 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt. \end{aligned}$$

□

The generalisation of Theorem 2.30 to obtain Archimedean copulas in higher dimensions is not straightforward, as we need stricter conditions on the generator φ.

Theorem 2.35. *If $\varphi : [0,1] \to [0,\infty]$ is a strict Archimedean copula generator and C is defined by*

$$C(u_1, \ldots, u_d) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)),$$

then C is a d-dimensional copula if and only if the inverse φ^{-1} is completely monotonic, i.e. $\forall k \in \mathbb{N}$ and $t > 0$

$$(-1)^k \frac{d^k}{dt^k} \varphi^{-1}(t) \geq 0.$$

Proof. Can be found in [27]. □

Proposition 2.36. *Let φ be a strict Archimedean copula generator that satisfies the requirements of Theorem 2.35 and let $C_{(l)}$ denote an l-dimensional Archimedean copula with generator φ. Then we have that for $k \leq d$ and $I_1 \uplus \cdots \uplus I_k = \{1, \ldots, d\}$,*

$$C_{(d)}(u_1, \ldots, u_d) = C_{(k)}(C_{(|I_1|)}(u_{I_1}), \ldots, C_{(|I_k|)}(u_{I_k})).$$

Proof. This follows easily from using the fact that $\varphi(\varphi^{-1}(t)) = t$ in the equation

$$C_{(d)}(u_1, \ldots, u_d) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_d)).$$

□

Archimedean copulas that are constructed via Theorem 2.35 are all exchangeable (see (2.8)) by construction. If we want to change this, we need to generalise them into nested Archimedean copulas. Nested Archimedean copulas arise when we replace arguments of Archimedean copulas with other (nested) Archimedean copulas.

Definition 2.37 (Nested Archimedean copulas)**.** Denote by $\mathcal{A}_N$ the set of nested Archimedean copulas.

1. If C is an Archimedean copula, then $C \in \mathcal{A}_N$.

2. Let C be a 2-dimensional Archimedean copula and $C_1, C_2 \in \mathcal{A}_N$. We define the function
$$C^* = C \circ \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}.$$
If C^* is a copula, then $C^* \in \mathcal{A}_N$.

Example 5. Let φ_1 and φ_2 be two strict Archimedean generators and define

$$C(u_1, u_2, u_3) = \varphi_2^{-1}(\varphi_2 \circ \varphi_1^{-1}(\varphi_1(u_1) + \varphi_1(u_2)) + \varphi_2(u_3)).$$

This is equivalent to a construction by taking C_i to be an Archimedean copula with generator φ_i and setting

$$C(u_1, u_2, u_3) = C_2(C_1(u_1, u_2), u_3).$$

This is an application of 2. in Definition 2.37 with $C_2(u) = u$ as a "one-dimensional copula", which is included in the definition. If φ_1^{-1} and φ_2^{-1} are completely monotonic decreasing functions and $\varphi_2 \circ \varphi_1^{-1}$ is a completely monotonic increasing function, then C is indeed a copula. It is not exchangeable, although u_1 and u_2 are an exchangeable group for C. The bivariate margins are again Archimedean copulas. C_{12} has generator φ_1, C_{13} and C_{23} have generator φ_2. This can be generalised for higher dimensions with according additional conditions on the generators, see [13]. For four dimensions, this construction with three generators $\varphi_1, \varphi_2, \varphi_3$ is possible:

$$C(u_1, u_2, u_3) = C_3(C_1(u_1, u_2), C_2(u_3, u_4)).$$

This gives two exchangeable groups. This construction can be found in [25]. Another possibility for arbitrary higher dimensions is the recursive equation

$$C_d(u_1, \ldots, u_d; \varphi_1, \ldots, \varphi_d) = \varphi_1^{-1}(\varphi_1(u_1) + \varphi_1(C_{d-1}(u_2, \ldots, u_d; \varphi_2, \ldots, \varphi_d))),$$

which only leaves u_{d-1} and u_d as exchangeable. The conditions for C being a copula are simple generalisations of the three-dimensional case, i.e. complete monotonicity of the composition of generators and inverses.

The most popular d-dimensional Archimedean copulas with one parameter are Clayton, Gumbel and Frank. They are named after the people who first described them. i.e. David Clayton (1978), Emil Gumbel (1960) and David Frank (1979) respectively. The Gumbel family is sometimes also referred to by the name Gumbel-Hougaard.

Definition 2.38. Clayton, Gumbel and Frank copula family in d dimensions

Copula	Generator $\varphi(t)$	Parameter	$C(u_1, \ldots, u_d)$
Clayton	$t^{-\alpha} - 1$	$\alpha > 0$	$\left(\sum_{i=1}^{d} u_i^{-\alpha} - d + 1\right)^{-\frac{1}{\alpha}}$
Gumbel	$(-\log(t))^{\alpha}$	$\alpha > 1$	$\exp\left(-\left(\sum_{i=1}^{d}(-\log(u_i)^{\alpha}\right)^{\frac{1}{\alpha}}\right)$
Frank	$\log\left(\frac{\exp(-\alpha u)-1}{\exp(-\alpha)-1}\right)$	$\alpha > 0$	$-\frac{1}{\alpha}\log\left(1 + \frac{\prod_{i=1}^{d}\left(e^{-\alpha u_i}-1\right)}{\left(e^{-\alpha}-1\right)^{d-1}}\right)$

The range of the parameter is valid for arbitrary dimension. In two dimensions larger ranges are possible: the Clayton copula can then also take $\alpha \in (-1, 0)$, while the Frank copula also accepts $\alpha < 0$.

Those copulas have the limiting case $C^{\perp}$ for α approaching the respective lower bounds and the limiting case C^{+} for $\alpha \to \infty$. Furthermore, it is evident that the k-dimensional margins of these three copulas are themselves copulas of the same family. We can therefore limit the investigation of pairwise dependences to the two-dimensional copulas.

Proposition 2.39. *In the two-dimensional case, the Archimedean copulas from Definition 2.38 have the following properties:*

Copula	*Kendall's tau* $\tau(\alpha)$	$\lambda_L(\alpha)$	$\lambda_U(\alpha)$
Clayton	$\frac{\alpha}{\alpha+2}$	$2^{-\frac{1}{\alpha}}$	0
Gumbel	$1 - \frac{1}{\alpha}$	0	$2 - 2^{\frac{1}{\alpha}}$
Frank	$1 - \frac{4}{\alpha} + \frac{4}{\alpha^2}\int_0^{\alpha} \frac{t}{e^t - 1} dt$	0	0

Furthermore, the Frank copula is radially symmetric. That is, $C = \tilde{C}$. *This is only true in the case of* $d = 2$.

Proof. Kendall's tau is calculated by formula (2.23). Now, we will only look at the lower tail dependence. The analogous calculation for the upper tail dependence is left to the reader. For the Clayton case, we have

$$\lambda_L = \lim_{u \to 0^+} \frac{\left(2u^{-\alpha} - 1\right)^{-\frac{1}{\alpha}}}{u} = \lim_{u \to 0^+} \left(2 - u^{\alpha}\right)^{-\frac{1}{\alpha}} = 2^{-\frac{1}{\alpha}}.$$

In the case of the Gumbel copula, we have

$$\lambda_L = \lim_{u \to 0^+} \frac{\exp\left(-\left(2(-\log u)^{\alpha}\right)^{\frac{1}{\alpha}}\right)}{u} = \lim_{u \to 0^+} u^{2^{\frac{1}{\alpha}} - 1} = 0.$$

The calculations for τ can be found in [37].

For the Frank copula, $\tilde{C}$ is given by

$$\begin{aligned}
\tilde{C}(u, v) &= -1 + u + v - \frac{1}{\alpha} \log\left(1 + \frac{\left(e^{-\alpha(1-u)} - 1\right)\left(e^{-\alpha(1-v)} - 1\right)}{e^{-\alpha} - 1}\right) \\
&= -\frac{1}{\alpha} \log\left(e^{-\alpha(u+v)} \frac{1 + e^{-\alpha(1-u-v)} - e^{\alpha u} - e^{\alpha v}}{e^{-\alpha} - 1}\right) \\
&= -\frac{1}{\alpha} \log\left(1 + \frac{e^{-\alpha(u+v)} + 1 - e^{-\alpha v} - e^{-\alpha u}}{e^{-\alpha} - 1}\right) = C(u, v).
\end{aligned}$$

Therefore, the Frank copula is radially symmetric in the case of $d = 2$. For higher dimensions, a counterexample can be found easily by directly evaluating C and $\tilde{C}$. □

We can see that the Frank copula has radially symmetric two-dimensional margins. This means that if we look at the components of a random vector X only using pairwise measures (τ, ρ_S, λ_L and λ_U), we cannot distinguish between a Frank copula and its survival copula. However, they are apparently distinct in all dimensions $d > 2$. Because the difference doesn't manifest itself in many measures and diagnostics, it is difficult to assess their properties in comparison. But as it is in the case of the self-evidently non-symmetric Archimedean copulas (e.g. Clayton and Gumbel), it is worth noting that by taking the survival copula, we virtually gain new parametric copula families without much additional effort.

The radial symmetry of the two-dimensional Frank copula is remarkable from another point of view: it is the only Archimedean copula to satisfy this property $C = \tilde{C}$, as noted by [37] and proven by Frank in [18].

2.5.3 Construction of New Copulas

The class of Archimedean copulas is very handy because it is rather easy to compute, but the limitations on the number of parameters is the downside of this. It is an important field of research to construct new families of copulas with desirable properties in terms of tractability of both computations and simulations in conjunction with an adequate number of parameters in the model. Considering actual data, the need to fit asymmetric tail behaviour and a wide variety of dependence structures arises. The different kinds of dependence structure manifest themselves in the dependence measures τ and ρ_S. We know of the attainable τ-ρ_S-region, as we have already seen. But there still persist big gaps in the set that can be reached by using the most popular copula families. For the Gaussian copula, the values of τ and ρ_S are fixed to one another as seen in Theorem 2.25. Therefore, they form a curve in the centre of the region. The t-copula expands this line into a thin stripe. But even the other frequently used copula families seem to fail reaching the most points in the τ-ρ_S-region, as can be seen in Figure 5 of [12]. One way to fix this problem is to apply transformations on copulas we already know. There are several methods to achieve this. In this section, we want to take a short look into the following two methods.

1) Copula Transformation

One can use a continuous, concave function $\gamma : [0,1] \to [0,1]$ and construct a transformed copula C_γ defined by

$$C_\gamma(x_1, \ldots, x_d) = \gamma^{-1}\big(C(\gamma(x_1), \ldots, \gamma(x_d))\big). \tag{2.24}$$

In arbitrary dimensions d, γ needs to have an inverse γ^{-1} that is absolutely monotonic of order d, i.e. $\forall k \leq d$ and t, $\frac{d^k}{dt^k}\gamma^{-1}(t) \geq 0$. This is sufficient for C_γ to be a copula, see [36].

Proposition 2.40. *If C is a two-dimensional copula and $\gamma : [0,1] \to [0,1]$ is a concave function and twice differentiable on $(0,1)$ that satisfies $\gamma(0) = 0$, $\gamma(1) = 1$, then C_γ defined in equation (2.24) is again a copula.*

Proof. Because $\gamma(0) = 0$, $\gamma(1) = 1$, the margins of C_γ satisfy the copula properties, e.g.

$$C_\gamma(u_1, 1) = \gamma^{-1}(C(\gamma(u_1), \gamma(1))) = \gamma^{-1}(C(\gamma(u_1), 1)) = \gamma^{-1}(\gamma(u_1)) = u_1.$$

In every point where C has a density, the density of C_γ is given by

$$\frac{\partial^2 C_\gamma}{\partial u_1 \partial u_2}(u_1, u_2) = \frac{\gamma'(u_1)\gamma'(u_2)}{\gamma'(C_\gamma(u_1, u_2))} \cdot \left(c(v_1, v_2) - \frac{\gamma''(C_\gamma(u_1, u_2))}{(\gamma'(C_\gamma(u_1, u_2)))^2} \frac{\partial C}{\partial u_1}(v_1, v_2) \frac{\partial C}{\partial u_2}(v_1, v_2) \right),$$

where $v_k = \gamma(u_k)$.

Because of concavity, we have that $\gamma'' \leq 0$. Therefore, the expression in the brackets is non-negative. Thus, $\frac{\partial^2 C_\gamma}{\partial u_1 \partial u_2}(u_1, u_2) \geq 0$. The rectangle inequality follows then by integration. □

From the proof, it becomes obvious that γ being concave is only sufficient and not a necessary condition for C_γ being a copula. This makes the set of possible transforming functions even larger.

This transformation broadens the set of τ-ρ_S-combinations that we can reach. It does this by shifting the attainable set of the copula family being transformed. In fact, bounds for Kendall's tau of the transformed copula can be given (see [12]), while the tail dependence remains unchanged by this transformation. Because of the vast class of possible transforming functions, this approach is very promising. Further theoretical elaborations on this transformation can be found in [11] and [46]. Some examples using different copula families and transforming functions are presented in [12].

2) Convex Combination

As summarized in [5], it is also possible to mix two given copulas C_1 and C_2 by convex combination

$$C(\cdot) = \mu C_1(\cdot) + (1 - \mu) C_2(\cdot), \qquad \mu \in [0, 1].$$

This is equivalent to taking C as distribution function of a random vector U with the conditional distribution $(U|D = i) \sim C_i$, where $D - 1$ is Bernoulli distributed with $\mathbb{P}(D = 1) = \mu$. The tail dependence of the convex combination C is easily calculated by the convex combination of the tail dependences

$$\lambda_L(C) = \mu \lambda_L(C_1) + (1 - \mu) \lambda_L(C_2).$$

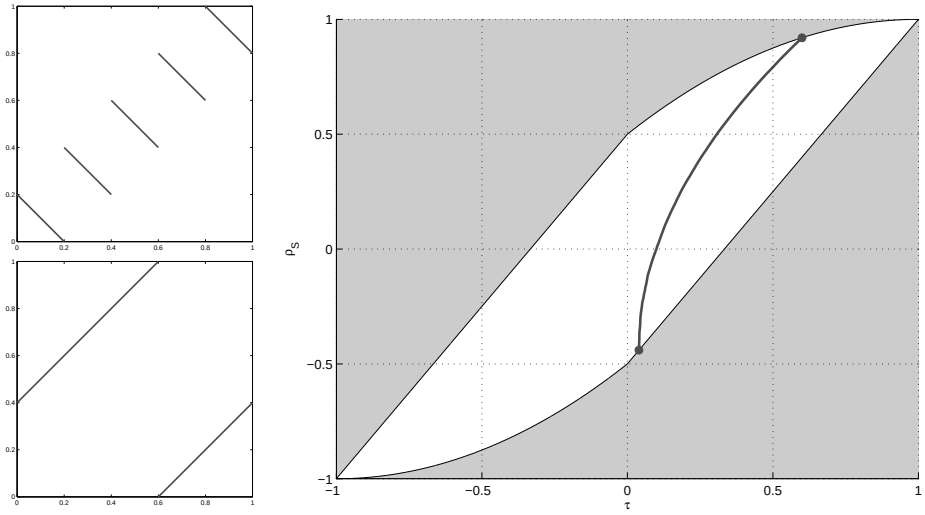

Figure 2.3: A path (right) of (τ, ρ_S)-pairs attainable by convex combination of two copulas (denoted by the two dots) which lie on the boundary of the τ-ρ_S-region. Left: the two copulas representing mutually completely dependent random variables, plotted as functions $u_2(u_1)$ according to Definition 2.24. The convex combination for $0 < \mu < 1$ does not represent mutually complete dependence anymore, because there are two possible values for $U_2|U_1 = u_1$ and vice versa.

For Spearman's rho, we see from (2.14) that

$$\rho_S(C) = \mu\rho_S(C_1) + (1-\mu)\rho_S(C_2).$$

For Kendall's tau, the same calculations and integration by parts yield

$$\begin{aligned}\tau(C) = \mu^2\tau(C_1) + (1-\mu)^2\tau(C_2) \\ + 2\mu(1-\mu)\left(4\int_{[0,1]^2} C_1(u,v)dC_2(u,v) - 1\right).\end{aligned}$$

This analytically very feasible construction can overcome some of the drawbacks of Archimedean copulas. For example, the upper and lower tail dependence can both be adjusted at the same time by mixing the Clayton and Gumbel copulas. It can of course be readily generalised into a combination of an arbitrary number of copulas C_i. As with the copula transformation we have seen in 1), it is difficult to comprehend how the attainable τ-ρ_S-region is changed by the transformation. Only by evaluating the integral do we know the path of points $(\tau(\mu), \rho_S(\mu))$ we can reach for $0 \leq \mu \leq 1$. An example for this can be seen in Figure 2.3.

The approach is simple with respect to both computation and simulation and has a very simplistic interpretation. Nonetheless, the wish for a model

with a larger number of parameters conflicts severely with the desirable simplicity and ease of interpretation which convex combinations of a bigger number of copulas cannot provide. Therefore, it is better to stick with convex combinations of only a small number of copulas. Thus, we remain relatively restricted in our number of parameters. But the possibility of combining properties of different families still gives us additional flexibility.

One other way of achieving the construction of new copulas is to numerically extract copulas from multivariate distributions that are already known, as we will see in a later chapter. This way one can also mix copulas in a more sophisticated and selective way instead of indiscriminately mixing by convex combination.

2.6 The Empirical Copula

In practical applications, the starting point of any task is analysing a set of data. Analogously to the one-dimensional case where we can estimate a distribution function F by its empirical counterpart $\hat{F}$, we can also do this with copulas. Empirical copulas were introduced by Deheuvels in 1979 under the denomination of *empirical dependence functions* D_n. He used them for distribution-free tests of independence.

Definition 2.41. Let $x_n = (x_{1n}, \ldots, x_{dn})^\top \in \mathbb{R}^d,\ 1 \leq n \leq N$ be an i.i.d. sample with continuous marginal distributions F_k.

The empirical copula (also called: Deheuvels empirical copula)[7] is defined as

$$\hat{C}\left(\frac{n_1}{N}, \ldots, \frac{n_d}{N}\right) = \frac{1}{N} \sum_{n=1}^{N} \prod_{k=1}^{d} \mathbb{1}\left(x_{kn} \leq x_k^{(n_k)}\right), \tag{2.25}$$

where $(x_1^{(i)}, \ldots, x_d^{(i)})$ is the ascending order statistic obtained from the sample, i.e. $x_k^{(i)}$ is the i-th smallest entry of component k in our sample.

This empirical copula $\hat{C}$ is only defined on the lattice $(n_1/N, \ldots, n_d/N)$ with $0 \leq n_k \leq N\ \forall k$. As such it is not a copula in the narrow sense of the Definition 2.5, but only a subcopula. However, since it is derived from the empirical distribution which is naturally a discrete distribution, Sklar's Theorem 2.11 implicates that only the values on this exact lattice are fixed. It further secures the existence of a copula $\hat{C}_*$ that has the same values on the lattice $\{0, 1/N, 2/N, \ldots, (N-1)/N, 1\}^d$.

More generally, (2.25) can very conveniently be written using the joint empirical distribution function $\hat{F}$ and the marginal empirical quantile functions $\hat{F}_i^{-1}$. This yields

$$\hat{C}\left(u_1, \ldots, u_d\right) = \hat{F}\left(\hat{F}_1^{-1}(u_1), \ldots, \hat{F}_d^{-1}(u_d)\right). \tag{2.26}$$

Furthermore, $\hat{C}$ is not to be confused with the notation for the survival copula found in other literature. Therefore, the survival copula is denoted by $\tilde{C}$ in this thesis.

The empirical copula frequency is defined by

$$\hat{c}\left(\frac{n_1}{N}, \ldots, \frac{n_d}{N}\right) = \frac{1}{N}, \tag{2.27}$$

if there is an index m and a data point $x_m = (x_{1m}, \ldots, x_{dm})^\top$ such that $x_m = (x_1^{(n_1)}, \ldots, x_d^{(n_d)})^\top$. $\hat{c}$ is zero otherwise. We thus exclude cases of doubled entries where there exist n and m such that $x_{kn} = x_{km}$ for all $1 \leq k \leq d$. With continuous marginal distributions, these cases will almost surely not arise. The two following equations hold:

$$\hat{C}\left(\frac{n_1}{N}, \ldots, \frac{n_d}{N}\right) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} \hat{c}\left(\frac{i_1}{N}, \ldots, \frac{i_d}{N}\right),$$

$$\hat{c}\left(\frac{n_1}{N}, \ldots, \frac{n_d}{N}\right) = \sum_{u \in \bigotimes_{i=1}^{d} \left\{\frac{n_i-1}{N}, \frac{n_i}{N}\right\}} (-1)^{s(u)} \hat{C}(u), \quad \text{with } s(u) = \#\{i : u_i = \tfrac{n_i-1}{K}\},$$

where the first equation represents the usual relationship between a density function and a distribution function. The sum in the second equation is defined analogously to the sum in the rectangle inequality (2.4).

The empirical copula can be used to estimate different dependence measures in a non-parametric way, as done by Deheuvels in [9].

Example 6. One could compute Spearman's rho by using Deheuvels empirical copula and formula (2.14), which yields

$$\hat{\rho}_S = \frac{12}{(N-1)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\hat{C}\left(\frac{i}{N}, \frac{j}{N}\right) - \frac{i}{N} \cdot \frac{j}{N} \right). \tag{2.28}$$

Assuming a Gaussian copula, the correlation matrix Σ can then be calculated by using formula (2.17) which connects ρ_S with Σ.

Proposition 2.42. *Equation* (2.28) *is equivalent to the usual sample version of Spearman's rho seen in formula* (2.15).

Proof. Looking at the sample version $\hat{\rho}_S$,

$$\hat{\rho}_S = \frac{12}{N(N^2-1)} \left(\sum_{n=1}^{N} R_n \cdot S_n - \frac{N(N+1)^2}{4} \right),$$

it is obvious that all we have to show is

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \hat{C}\left(\frac{i}{N}, \frac{j}{N}\right) = \frac{1}{N} \sum_{n=1}^{N} R_n \cdot S_n.$$

This follows from

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{C}\left(\frac{i}{N},\frac{j}{N}\right) &= \sum_{i=1}^{N}\sum_{j=1}^{N}\hat{C}\left(\frac{R_i}{N},\frac{S_j}{N}\right)\\ &= \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=1}^{N}\mathbb{1}\left(x_{1n}\le x_1^{(i)}\right)\sum_{j=1}^{N}\mathbb{1}\left(x_{2n}\le x_2^{(j)}\right)\right)\\ &= \frac{1}{N}\sum_{n=1}^{N}(K-R_n+1)(K-S_n+1) = \frac{1}{N}\sum_{n=1}^{N}R_n\cdot S_n.\end{aligned}$$

□

A similar result holds for Kendall's tau; the proof can be found in [37].

Theorem 2.43. *Let $X_n, n\in\mathbb{N}$ be an i.i.d. sequence of random variables on $\mathbb{R}$ with the distribution function F. The empirical distribution function $\hat{F}$ be defined by*

$$\hat{F}_N(x) = \frac{1}{N}\sum_{n=1}^{N}\mathbb{1}(X_n\le x). \tag{2.29}$$

Then,

i) (Theorem of Glivenko-Cantelli)

$$\sup_{x\in\mathbb{R}}\left|\hat{F}_N(x)-F(x)\right|\to 0 \quad a.s. \tag{2.30}$$

ii) For $x\in\mathbb{R}$, we have pointwise

$$\sqrt{N}\left(\hat{F}_N(x)-F(x)\right)\xrightarrow{d} N\big(0,F(x)(1-F(x))\big). \tag{2.31}$$

Proof. For i), let $\epsilon>0$. There exist $-\infty=t_0<t_1<\cdots<t_k=\infty$ such that for all $0\le i\le k-1$,

$$F(t_{i+1}^-)-F(t_i)\le\frac{\epsilon}{2}. \tag{2.32}$$

Take $x\in\mathbb{R}$, then $x\in[t_i,t_{i+1})$ for some i. Because of the monotonicity of F and $\hat{F}_N$, we have

$$\hat{F}_N(t_i)-F(t_{i+1}^-)\le\hat{F}_N(x)-F(x)\le\hat{F}_N(t_{i+1}^-)-F(t_i).$$

By using (2.32), we get

$$\hat{F}_N(t_i)-F(t_i)-\frac{\epsilon}{2}\le\hat{F}_N(x)-F(x)\le\hat{F}_N(t_{i+1}^-)-F(t_{i+1}^-)+\frac{\epsilon}{2}.$$

For every fixed $y\in\mathbb{R}$, $\hat{F}_N(y)-F(y)\to 0$ a.s. because of the strong law of large numbers. Therefore, the result (2.30) follows.

ii) follows from the central limit theorem. □

The theorem of Glivenko-Cantelli also holds in higher dimensions. (2.31) even converges weakly to a Gaussian process. As described in [16], the empirical copula by Deheuvels also shows weak convergence.

Theorem 2.44. *Let C be a copula and F_i continuous marginal distribution functions. F is defined by*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

Let X_n be an i.i.d. sequence of random vectors with distribution function F. $\hat{C}_N$ is the empirical copula of $\{X_n\}_{1 \le n \le N}$ defined in (2.26). If the copula C has continuous partial derivatives, then the empirical copula process

$$\sqrt{N}\left(\hat{C}_N - C\right)(u_1, \dots, u_d) \tag{2.33}$$

converges weakly to the limiting Gaussian process $\mathbb{G}_C$ in $l^\infty([0,1]^d)$.

The proof of this can be seen in [16]. It uses the functional delta method from [47]. For $d = 2$, the limiting Gaussian process has the form

$$\mathbb{G}_C(u_1, u_2) = \mathbb{B}_C(u_1, u_2) - \frac{\partial C(u_1, u_2)}{\partial u_1}\mathbb{B}_C(u_1, 1) - \frac{\partial C(u_1, u_2)}{\partial u_2}\mathbb{B}_C(1, u_2),$$

where $\mathbb{B}_C$ is a Brownian bridge on $[0,1]^2$ with covariance

$$\mathbb{E}\Big[\mathbb{B}_C(u_1, u_2) \cdot \mathbb{B}_C(v_1, v_2)\Big] = C(u_1 \wedge v_1, u_2 \wedge v_2) - C(u_1, u_2)C(v_1, v_2),$$

for all $u_1, u_2, v_1, v_2 \in [0,1]$. We can see that the limiting distribution is decidedly more complex than in the one-dimensional case of the empirical distribution function.

2.7 The Rosenblatt Transform

The Rosenblatt transform is a transformation of a random vector X into an independent uniformly distributed vector. As such, it is of high value in verifying a distribution assumption and testing hypotheses.

Definition 2.45. Let X be a d-dimensional random vector and F be its distribution function with the margins $X_k \sim F_k$, $\forall\ 1 \le k \le d$. The Rosenblatt transform $z = \mathcal{R}(x)$ of the d-dimensional vector $x = (x_1, \dots, x_d)^\top$ is given by

$$\begin{aligned}
z_1 &= \mathbb{P}(X_1 \le x_1) = F_1(x_1), \\
z_2 &= \mathbb{P}(X_2 \le x_2 | X_1 = x_1) = F_{2|1}(x_2|x_1), \\
&\vdots \\
z_d &= \mathbb{P}(X_k \le x_d | X_{d-1} = x_{d-1}, \dots, X_1 = x_1) = F_{d|d-1,\dots,1}(x_d|x_{d-1}, \dots, x_1).
\end{aligned}$$

The Rosenblatt transform of X is denoted by $Z = \mathcal{R}(X)$.

As can be seen from the name, the Rosenblatt transform was first proposed by Murray Rosenblatt in 1952, see [38].

Theorem 2.46. *The Rosenblatt transform of a d-dimensional random vector X with continuous distribution function F and margins $X_k \sim F_k$, $\forall\, 1 \leq k \leq d$, is independent uniformly distributed on $[0,1]^d$.*

Proof. Let $Z = \mathcal{R}(X)$ be the transform. The margins Z_k are standard uniformly distributed because $F_{k|k-1,\dots,1}$ is continuous for all k. For $k \geq 1$, the independence follows from induction. Let $z \in [0,1]^d$, then

$$\begin{aligned} F_{Z_{1\dots k+1}}(z_1,\dots,z_{k+1}) &= \mathbb{P}(Z_1 \leq z_1,\dots,Z_{k+1} \leq z_{k+1}) \\ &= \mathbb{E}[\mathbb{P}(Z_1 \leq z_1,\dots,Z_{k+1} \leq z_{k+1}|Z_1,\dots,Z_k)] \\ &= \mathbb{E}[\mathbb{P}(Z_{k+1} \leq z_{k+1}|Z_1,\dots,Z_k)\cdot \mathbb{1}(Z_1 \leq z_1,\dots,Z_k \leq z_k)] \\ &= z_{k+1} \;\cdot\; F_{Z_{1\dots k}}(z_1,\dots,z_k) \\ &= z_{k+1} \;\cdot\; \prod_{i=1}^{k} z_i = \prod_{i=1}^{k+1} z_i. \end{aligned}$$

For $k = d$, we then have that $Z = \mathcal{R}(X)$ is indeed uniformly distributed on $[0,1]^d$. □

Corollary 2.47. *Let X be a d-dimensional random vector X with continuous distribution function F and density f, then*

1. *the Rosenblatt transform $Z = \mathcal{R}(X)$ can be calculated by*

$$\begin{aligned} z_k &= F_{k|k-1,\dots,1}(x_k|x_{k-1},\dots,x_1) \\ &= \frac{\partial^{k-1} F_{1,\dots,k}(x_1,\dots,x_k)}{\partial x_1 \dots \partial x_{k-1}} \Bigg/ \frac{\partial^{k-1} F_{1,\dots,k-1}(x_1,\dots,x_{k-1})}{\partial x_1 \dots \partial x_{k-1}} \\ &= \frac{\partial^{k-1} F_{1,\dots,k}(x_1,\dots,x_k)}{\partial x_1 \dots \partial x_{k-1}} \Bigg/ f_{1,\dots,k-1}(x_1,\dots,x_{k-1}). \end{aligned} \tag{2.34}$$

2. *Let C be the copula of X. If the margins of X have a strictly increasing distribution function, the Rosenblatt transform of an arbitrary vector $u \in [0,1]^d$ with respect to the copula C is given by $z = \mathcal{R}(u)$ and can be easily calculated by* (2.34) *where we set $x_i = F_i^{-1}(u_i)$.*

Proof. For every d-dimensional distribution function G with density g and a partition $I \uplus J = \{1,\dots,d\}$, we have

$$\frac{\partial G(x_I, x_J)}{\partial x_I} = g_I(x_I) \;\cdot\; G_{J|I}(x_J|x_I).$$

This follows by induction because w.l.o.g. for $h > 0$,

$$\begin{aligned}\frac{\partial G(x)}{\partial x_1} &= \lim_{h\to 0} \frac{G(x_1+h, x_2, \ldots, x_d) - G(x_1, x_2, \ldots, x_d)}{h} \\ &= \lim_{h\to 0} \frac{\mathbb{P}(x_1 < X_1 \leq x_1 + h)}{h} \cdot \mathbb{P}(X_2 \leq x_2, \ldots, X_d \leq x_d | x_1 < X_1 \leq x_1 + h) \\ &= g_1(x_1) \cdot G_{2,\ldots,d|1}(x_2, \ldots, x_d | x_1).\end{aligned}$$

Therefore, (2.34) holds. □

The Rosenblatt transform can be used to test whether a sample $\mathbf{u}$ is distributed according to a given copula C. It is especially suited for this as testing $\mathbf{v} = \mathcal{R}(\mathbf{u})$ for independence and uniformity can be easier than testing $\mathbf{u}$ directly for the more complicated hypothesis. We will see this in Section 3.5.

Chapter 3

Copula Estimation and Fitting

The task of delivering a fitted copula model to a certain given dataset is not an easy one. The procedure consists in three steps. The first part is the analysis of the data in order to decide roughly on the parametric copula family to be used. In the second step, we estimate the parameters. This can be done in a variety of ways. Lastly, we have to review the resulting fitted model and verify that it does indeed fit the data. In the contrary case, it can be advisable to go back to step one and try a different parametric copula family.

3.1 Choosing the Right Copula

When being confronted with multivariate data in a copula setting, the first and essential step is analysing the data for distinctive properties that suggest a certain copula family. The most important properties to assess are symmetry ($\tilde{C} \stackrel{?}{=} C$), exchangeability ($C(u,v) \stackrel{?}{=} C(v,u)$) and tail dependence. Only if an adequate copula family is chosen, the parameter estimation can yield a good result. Evidently asymmetric data forbids using an elliptical copula.

While real data can only be adequately modelled by using a sufficiently high number of parameters, in practice, it is desirable to find a simplistic model with parameters that are accessible for interpretation of the model. This also helps to prevent the problem of overfitting the model. For doing this, there is also tools like the AIC and the BIC that are designed to control the numbers of parameters by imposing a penalty on their number. What they don't do is make sure that the model has a sound way of being interpreted. If a model is complex and the reasoning behind the construction is unclear, then it will be difficult to justify its use. Therefore, we should always strive for simplicity while preserving as much complexity in the model

as is necessary. This means that we should always compare the complex model with a simpler one. If the advantage is negligible, the simple model should be chosen.

3.2 ML Estimation of the Parameters

Statistical inference on copulas has to deal with two problems: making assumptions for identifying the marginal distributions and choosing the copula function itself. Let X be a d-dimensional random vector. Assume our parametric distribution model for X consists of two parameter vectors θ_C and θ_M. θ_C are the parameters of the copula while θ_M are the parameters for the margins. Therefore, the density of X is $f(\cdot|\theta_C, \theta_M)$. The marginal distribution functions are given by $F_k(\cdot|\theta_M)$; their density is $f_k(\cdot|\theta_M)$. Let $(x_{kn})_{1\leq k\leq d, 1\leq n\leq N}$ denote our sample of X, consisting of N data points.

The log-likelihood function of $(x_{kn})_{1\leq k\leq d, 1\leq n\leq N}$ is then given by:

$$\begin{aligned} l(\theta_C, \theta_M) &= \log\left(\prod_{n=1}^{N} f(x_{1n}, \ldots, x_{dn}|\theta_C, \theta_M)\right) \\ &= \sum_{n=1}^{N} \log c(F_1(x_{1n}|\theta_M), \ldots, F_d(x_{dn}|\theta_M)|\theta_C) + \sum_{n=1}^{N}\sum_{k=1}^{d} \log f_k(x_{kn}|\theta_M). \end{aligned} \tag{3.1}$$

Given the marginal distribution functions and an appropriate parametric copula, maximum likelihood can be applied on the parameters of the marginal distributions and the copula at the same time.

$$\hat{\theta}_{\mathrm{MLE}} = \underset{(\theta_C, \theta_M)\in\Theta_C\times\Theta_M}{\arg\max}\; l(\theta_C, \theta_M) \tag{3.2}$$

Because the function has to be numerically optimised with respect to all of the parameters in (θ_C, θ_M) simultaneously, this method of finding the so-called *exact maximum likelihood estimator* can be computationally expensive.

It is known that if we observe $X_1, \ldots, X_n$, where $X_i \overset{\text{iid}}{\sim} f(x|\theta)$, ML estimators are consistent *under suitable regularity conditions*, as can be read in [6]. These conditions are:

(M1) The parameter is identifiable, i.e. $\theta \neq \theta' \Rightarrow f(x|\theta) \neq f(x|\theta')$.

(M2) The densities $f(x|\theta)$ have common support and $f(x|\theta)$ is differentiable in θ.

(M3) The parameter space contains an open set of which the true parameter value θ_0 is an interior point.

Under further assumptions we can guarantee asymptotic normality of the maximum likelihood estimator $\hat{\theta}_{\mathrm{MLE}}$. The conditions are:

(M4) $\int f(x|\theta)dx$ can be differentiated three times under the integral sign.

(M5) For all x, $f(x|\theta)$ is three times differentiable with respect to θ, with a continuous third derivative. For any $\theta_0 \in \Theta$, there exist $c > 0$ and $M(x)$ (both possibly depending on θ_0) such that

$$\left|\frac{\partial^3 \log f(x|\theta)}{\partial \theta^3}\right| \leq M(x), \qquad \forall x,\ \theta \in (\theta_0 - c, \theta_0 + c),\ \ \mathbb{E}_{\theta_0}[M(X)] < \infty.$$

Under the conditions (M1)-(M5), the ML estimator is asymptotically normal and efficient, i.e. it asymptotically reaches the Cramér-Rao bound of variance:

$$\sqrt{N}\left(\hat{\theta}_{\text{MLE}} - \theta_0\right) \xrightarrow{d} N(0, I^{-1}(\theta_0)),$$

where θ_0 is the true parameter value and $I(\theta_0)$ is Fisher's information matrix defined by

$$I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial \log f(X|\theta)}{\partial \theta}\right)^2\right].$$

If we take a look at the form of the likelihood function in (3.1), it becomes clear that the latter part is a term only involving the parameter θ_M of the marginals. Therefore, it is possible to approximate the exact MLE $\hat{\theta}_{\text{MLE}}$ by first maximising the second term in the parameters of the marginal distributions θ_M and then the first term in the copula parameter θ_C using $\hat{\theta}_M$. This method of inference for the margins is called IFM estimation and is computationally less involved.[8]

$$\begin{aligned}
\hat{\theta}_M &= \underset{\theta_M \in \Theta_M}{\arg\max} \left(\sum_{n=1}^{N}\sum_{k=1}^{d} \log f_k(x_{kn}|\theta_M)\right), \\
\hat{\theta}_C &= \underset{\theta_C \in \Theta_C}{\arg\max} \left(\sum_{n=1}^{N} \log c(F_1(x_{1n}|\hat{\theta}_M), \ldots, F_d(x_{dn}|\hat{\theta}_M)|\theta_C)\right), \\
\hat{\theta}_{\text{IFM}} &= \left(\hat{\theta}_C, \hat{\theta}_M\right).
\end{aligned} \tag{3.3}$$

Under the usual regularity conditions for asymptotic maximum likelihood theory on both the joint distribution and the margins, the IFM estimator is asymptotically normal, see [25]. These conditions are stated below in (C1)-(C6).

Various comparisons suggest that, although the methods of assessment can prove difficult, the IFM estimation is highly efficient compared with the exact ML estimation. Additionally, the IFM estimator (3.3) can not only be used as it is but it can also be taken as a starting point for obtaining the MLE (3.2) via numerical optimisation.

Both methods (3.2) and (3.3) have the flaw that choosing the parametric family for the marginals can be a difficult task. There are many choices for

marginal distributions. But at its core, letting the choice of the marginals influence the copula model is conflicting with what copula theory is aiming to do, i.e. keeping marginal distributions and dependence separate.

It is also possible to estimate the copula parameters without specifying the marginals. Because the copula is an invariant under strictly increasing transformation of the margins of X, it is only reasonable to resort to the maximally invariant property, i.e. the ranks of the observation. This way, the estimation of θ_C is truly margin-free. We thus avoid a parametric model of the marginal distribution by instead using the empirical distribution function $\hat{F}_k(x_{kn})$ to estimate the marginal distribution $F_k(x_{kn})$ in a non-parametric way by

$$\hat{F}_k(x_{kn}) = \frac{1}{N+1} \sum_{l=1}^{N} \mathbb{1}(x_{kl} \leq x_{kn}), \tag{3.4}$$

which then yields a pseudo-likelihood function to be maximised. This semiparametric method of estimating θ_C is called the Canonical Maximum Likelihood (CML), as can be seen in [7]. It is semiparametric because the margins are estimated non-parametrically and the copula itself is taken to be in a parametric family.

$$\hat{\theta}_{\text{CML}} = \underset{\theta_C \in \Theta_C}{\arg\max} \left(\sum_{n=1}^{N} \log c(\hat{F}_1(x_{1n}), \ldots, \hat{F}_d(x_{dn}) | \theta_C) \right) \tag{3.5}$$

The three different approaches to maximum likelihood parameter estimation are compared in [3].

Theorem 3.1. *Under the conditions (C1)-(C6), the semiparametric estimator $\hat{\theta}_{CML}$ is consistent and $\sqrt{N}\left(\hat{\theta}_{CML} - \theta\right)$ is asymptotically normal.*

This result was proven in [19]. In any case, the asymptotic variance of $\hat{\theta}_{\text{CML}}$ is obviously larger than in the case where the marginal distributions are known. The regularity conditions that can be found in [32] or [44] are as follows:

(C1) The parameter space Θ is an open interval.

(C2) The distributions $f(x|\theta)$ have a common support that is independent of θ.

(C3) For all x, $f(x|\theta)$ is three times differentiable with respect to θ, with a continuous third derivative. For each $\theta_0 \in \Theta$, there exists a bounding function (that can depend on the true parameter value θ_0) such that for θ in a neighbourhood of θ_0 the following inequalities hold:

$$\left| \frac{\partial^3 \log f(x|\theta)}{\partial \theta^3} \right| \leq H(x), \qquad \mathbb{E}_\theta[H(X)] < \infty.$$

(C4) The integral $\int f(x|\theta)dx$ can be differentiated three times under the integral sign.

(C5) For all $\theta \in \Theta$, the Fisher information satisfies $0 < I(\theta) < \infty$

(C6) For any $\theta_0 \in \Theta$ there exist $c > 0$ and $M(x)$ (both possibly depending on θ_0) such that

$$\left|\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2}\right| \leq M(x), \qquad \forall x,\ \theta \in (\theta_0 - c, \theta_0 + c),\ \ \mathbb{E}_{\theta_0}[M(X)] < \infty.$$

The differentiability under the integral sign can be guaranteed by the existence of bounding functions for each $\theta_0 \in \Theta$, such that for θ in a neighbourhood of θ_0 the following inequalities hold:

$$\left|\frac{\partial f(x|\theta)}{\partial \theta}\right| \leq g(x), \qquad \left|\frac{\partial^2 f(x|\theta)}{\partial \theta^2}\right| \leq h(x),$$
$$\int g(x)dx < \infty, \qquad \int h(x)dx < \infty.$$

When comparing different models, it is important to choose the model as simple as possible while fitting the copula to the data as well as possible. In order to do this, one can look at the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which impose a penalty on every parameter in the model:

$$\text{AIC} = -2 \log L(\hat{\theta}|\mathbf{X}) + 2k_\theta, \tag{3.6}$$
$$\text{BIC} = -2 \log L(\hat{\theta}|\mathbf{X}) + \log(N)k_\theta, \tag{3.7}$$

where $\log L(\cdot|\mathbf{X})$ is the log-likelihood function, N is the sample size and k_θ is the number of parameters in the model. Small values of AIC and BIC are preferred. For practical purposes, the BIC penalises the number of parameters more heavily than the AIC. While they are of course closely related, it is sometimes suggested that the AIC is better than the BIC from a theoretical point of view, as argued in [4]. In practice, both of them are frequently used.

3.3 Using Measures of Dependence

Besides the MLE and the estimation via the generalised method of moments (GMM), one can also estimate the parameters of a copula by trying to match most closely certain empirical properties with their theoretical counterparts of the parametric copula. There are two copula properties that are often used to fit a parametric copula to empirical data. Firstly, there is measures of dependence (see [7]) and secondly, there is the measure of tail dependence

(see [5]). All of these have in common that they only look at the two-dimensional case and thus can only tell us something about the pairwise dependence the components of a multidimensional random vector X exhibit. However, there do exist multidimensional generalisations for some of them, see [42].

Using known relations between dependence measures and parameters of copulas can provide us with a convenient way of parameter estimation. Using Theorem 2.26, Kendall's tau can be used to obtain estimates for elliptical copulas. Spearman's rho has a similar relation to the Gaussian copula, as seen in Theorem 2.25. In the case of the t-copula, $\hat{\Sigma}$ is usually obtained via Kendall's tau. $\hat{\nu}$ is then chosen with maximum likelihood estimation. One problem that can arise in this context is that the estimator $\hat{\Sigma}$ is not a positive definite matrix. In that case, one can use Algorithm 5.55 from [13] to obtain a positive definite correlation matrix that is close to the original estimate. The algorithm for adjusting a symmetric matrix R^* into a positive definite correlation matrix R consists in four steps:

1. Compute the spectral decomposition $TDT^\top = R^*$.

2. Replace all non-positive eigenvalues in D with $\delta > 0$. This yields D'.

3. $S = TD'T^\top$. The diagonal elements of S are not necessarily one.

4. R follows by norming the matrix, i.e. $R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} \cdot S_{jj}}}$.

With this algorithm, a valid estimation $\hat{\Sigma}$ of the correlation matrix is possible. Most of the time, no correction is necessary in smaller dimensions, but cases do occur. Especially in higher dimensions this problem is quite diffused. The approach of the eigenvalue method is of course rather brutal in that we don't really know what we did to the underlying matrix and how similar the result actually is to the original matrix. There are other, more sophisticated approaches that try to find the nearest correlation matrix as measured by the Frobenius norm, with the possibility of preventing certain entries more strongly from being adjusted than others by using a weighted norm, see [24].

For Archimedean copulas, it is convenient to use Theorem 2.34 and the relations that follow easily from (2.23) for many Archimedean models. However, as can be seen in Proposition 2.39, numerical methods can become necessary to find the parameter in some cases. The known relations for the tail dependence of Archimedean copulas (as seen in Proposition 2.39) can be an easy way to fit Archimedean copulas to empirical data. In the multidimensional case, this method suffers from the mismatch between an abundance of empirical tail dependences and a lack of free parameters.

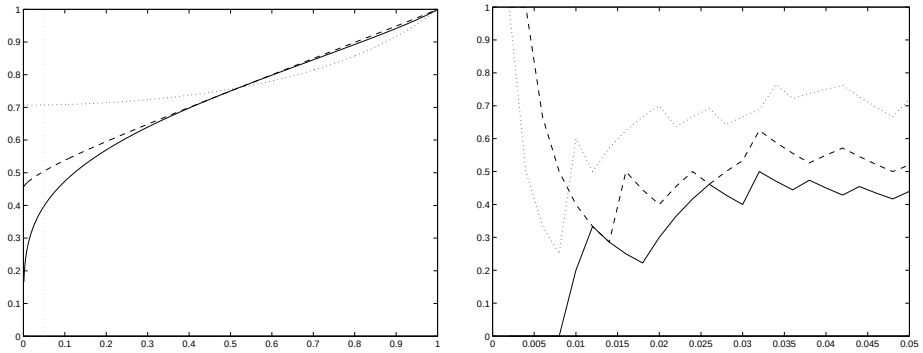

Figure 3.1: Quantile dependence function $\lambda_L(q)$ for three copulas (each $\tau = 0.5$). Gauss-copula, $t(3)$-copula (dashed), Clayton (dotted). Left: theoretical values. Right: sample version for a sample of $N = 500$ and quantiles $q \leq 0.05$.

3.4 Estimation of Tail Dependence

The easiest way to estimate tail dependence is to substitute the copula C in the formulas (2.19) and (2.20) with its empirical estimator $\hat{C}$. This way we get a non-parametric, "naive" estimator for the tail dependence. As we have only finitely many data points and the empirical copula is (uniquely) defined only on a lattice, it is impossible to take the limits of the fractions. Therefore, we can only take a look at their trajectories for $1 \leq i \leq N$.

$$\hat{\lambda}_U\left(\frac{i}{N}\right) = \frac{1 - 2\frac{i}{N} + \hat{C}\left(\frac{i}{N}, \frac{i}{N}\right)}{1 - \frac{i}{N}},$$
$$\hat{\lambda}_L\left(\frac{i}{N}\right) = \frac{\hat{C}\left(\frac{i}{N}, \frac{i}{N}\right)}{\frac{i}{N}},$$

where $\hat{C}$ is the empirical copula defined in (2.25). These are also called sample versions of the *quantile dependence functions*.

The difficulty lies now in choosing an adequate index i_0 to evaluate the expressions above. One can observe the jumps in the trajectory becoming bigger as the index i approaches the edge, with the denominator approaching zero. Now one could just look at the plot of those functions and select a part where there's a relative stability in the trajectory. As a more rigid approach, we try a bootstrapping method that computes S different bootstrapped trajectories and takes their average $\lambda_U^*(\cdot)$ and $\lambda_L^*(\cdot)$. Taking the average smooths out the bumps in the original functions. We are also given bootstrap confidence intervals for further investigation. This can be seen in Figure 3.2.

$\lambda_U(u)$ and $\lambda_L(u)$ are respectively decreasing and increasing functions on $[0, 1]$ (see Figure 3.1). As their empirical estimates $\hat{\lambda}_U(\cdot)$ and $\hat{\lambda}_L(\cdot)$ should have the same property, we choose the last i such that the functions λ_L^* and

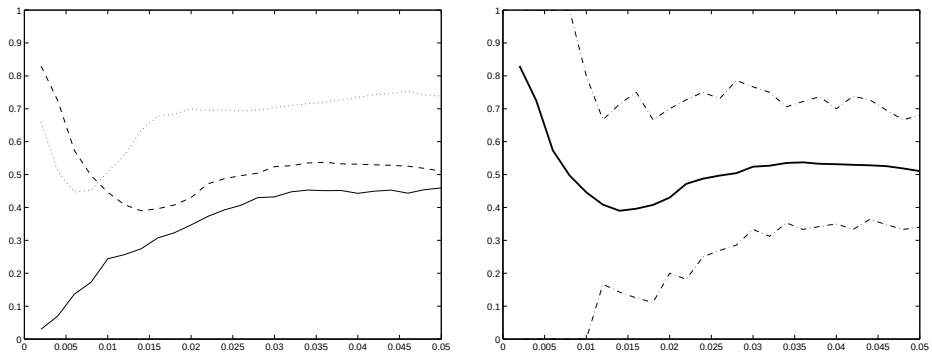

Figure 3.2: Left: bootstrapped ($S = 100$) trajectories $\lambda_L^*(q)$ of the sample quantile dependence function for the copulas from Figure 3.1. Right: bootstrapped trajectory for the t-copula with 90% confidence interval.

λ_U^* show this property. Taking $\lambda_L^*(\cdot)$, what we want to find is a period of relative stability between the unstable low quantiles near the edge and the monotonically increasing stretch of the higher quantiles. The key lies in the interpretation of this goal. To ease the criterion of monotonicity, we say that they should satisfy it up to a tolerance ϵ_1. If $\lambda_L^*(\cdot)$ is approximately constant over a certain interval, which again is up to a tolerance ϵ_2, we take the median of $\hat{\lambda}_L$ over this interval. To illustrate this convoluted approach, we provide Figure 3.3.

Using this approach, we are provided with a clearly defined algorithm to obtain estimates for the tail dependence. However, as shown in [5] with an empirical study, the efficiency of the estimators depends on the true value of λ_L and λ_U. With higher tail dependence, standard error and bias decrease. It is especially difficult to reach converging estimates in cases of copulas with a tail dependence lower than 0.2. With increasing sample size N, the estimators converge empirically. In the cases listed in [5], the estimates with a sample size of $N = 5000$ show quite good convergence to the real value.

The main limitation of this approach is the fact that the estimation is entirely dependent on the choice of the index i_0. However well we specify the rules for the choice, it remains somewhat arbitrary. Its strength lies in the fact that it is non-parametric, i.e. we do not have to choose a copula family a priori. Because we don't assume properties of the underlying copula by choosing a parametric family before estimating the tails, it is the data and the estimates derived from it that propose us a certain choice of copula family. If the estimates seem to support a model with symmetric and non-zero tail dependences, a t-copula might already be the right and easy choice. If the tails are evidently asymmetric, further investigation can be necessary.

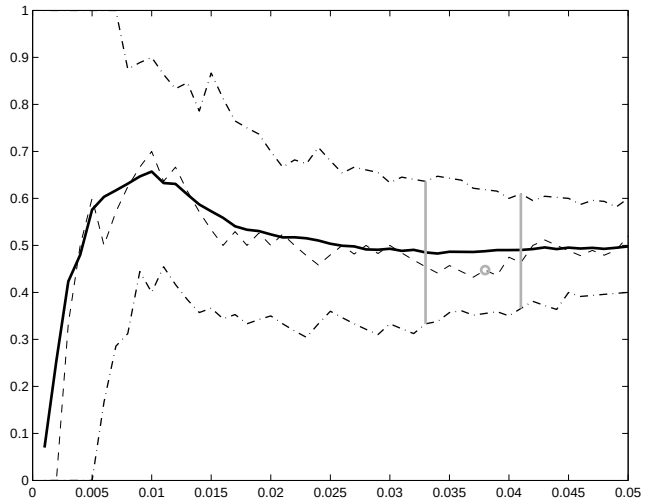

Figure 3.3: Estimation procedure on a sample ($N = 1000$) of the $t(3)$-copula from Figure 3.2. We set $\epsilon_1 = \epsilon_2 = 0.005$. Estimated value $\hat{\lambda}_L = 0.4474$. The theoretical value is $\lambda_L = 0.454$. The grey bars indicate the "stable" part of the trajectory of $\lambda_L^*(\cdot)$ (thick), i.e. indices $33, \ldots, 41$. The point where $\hat{\lambda}_L(\frac{38}{1000})$ (dashed) is evaluated is marked. Note that the bootstrapped 90% pointwise confidence interval explodes to $[0, 1]$ for small values q.

3.5 Goodness of Fit

In general, one will fit a copula to a given dataset by trying to match rank correlation measures like Kendall's tau or by employing maximum likelihood estimation. After the parameters are estimated, it is only natural to ask how well the estimated copula reproduces the behaviour observed in the original dataset. This can be done by comparing empirical and theoretical values of ρ_S, τ, λ_L and λ_U.

One other useful method to examine how good a copula fits the data is to take a look at the quantiles, i.e. C itself. Because of Theorem 2.44, we have the theoretical basis on which to do goodness-of-fit tests. As $\hat{C}$ should approach C with large sample sizes, all that is left to do is measuring the distance. To obtain this, we can take different approaches. [5] are using a sort of L^2 distance, which measures how far the empirical copula and the estimated copula $C_{\hat{\theta}}$ differ from each other. They approximate the following integral:

$$\int_{[0,1]^d} \Big(C_\theta(u_1, \ldots, u_d) - C(u_1, \ldots, u_d) \Big)^2 du_1 \cdots du_d, \tag{3.8}$$

where we take the distance between a parametric copula and the actual (or in practice: the observed) copula. By changing the integration measure, we

get the Cramér-von Mises statistic

$$\int_{[0,1]^d} \Big(C_\theta(u_1,\dots,u_d) - C(u_1,\dots,u_d)\Big)^2 dC(u_1,\dots,u_d), \tag{3.9}$$

It is also possible to look at the Kolmogorov-Smirnov statistics, i.e.

$$\sup_{u\in[0,1]^d} \left|C_\theta(u_1,\dots,u_d) - C(u_1,\dots,u_d)\right|. \tag{3.10}$$

We can take the sample version of those by plugging in the empirical copula. As the empirical copula is not uniquely defined in between the lattice points, the only way to take a sample version of these is by taking the sum or maximum over the sample points. To account for sample size, the statistics are normed using $\sqrt{N}\cdot(C_\theta(\cdot) - \hat{C}^*(\cdot))$.

Because of the different definitions of the empirical distribution function of the margins with either N or $N+1$ as the denominator in equation (3.4), we once again clearly define the way the grades $\mathbf{u}$ are calculated from the sample $\mathbf{x} = (x_{kn})_{1\le k\le d, 1\le n\le N}$ and the empirical copula function used. We compute the grades $\mathbf{u}$ via

$$u_{kn} = \hat{F}_k(x_{kn}) = \frac{1}{N+1}\sum_{l=1}^{N} \mathbb{1}(x_{kl} \le x_{kn}),$$

as done by the CMA separation Algorithm 1. Therefore, the empirical copula used is $\hat{C}^*$ defined by

$$\hat{C}^*(v_1,\dots,v_d) = \frac{1}{N}\sum_{n=1}^{N}\prod_{k=1}^{d} \mathbb{1}\left(u_{kn} \le v_k\right), \tag{3.11}$$

which suits the choice of the empirical distribution function. For such $\mathbf{u}$, the relation to Deheuvels empirical copula (2.25) is given by

$$\hat{C}(u) = \hat{C}^*\Big(\tfrac{N}{N+1}\cdot u\Big).$$

Due to its nature as a step function, $\hat{C}^*$ cannot be a copula, but it is asymptotically the same as Deheuvels empirical copula. In the following definition, the usage of $\hat{C}^*$ (along with the proposed choice of $\hat{F}_k$) is preferred because it prevents the copula frequency $\hat{c}^*$ from being non-zero at the edges of $[0,1]^d$, where both copulas are already fixed to their values by construction, as has been mentioned by [23].

Definition 3.2. Given a copula C_θ and the empirical copula $\hat{C}^*$ of a set of points with grades $\mathbf{u} = (u_{kn})_{1\le k\le d, 1\le n\le N}$

i) the L^2-distance is defined as

$$D_2(C_\theta, \mathbf{u}) = \frac{1}{N^{d-1}} \sum_{n_1=1}^{N} \cdots \sum_{n_d=1}^{N} \left(C_\theta(u_{1n_1}, \ldots, u_{dn_d}) - \hat{C}^*(u_{1n_1}, \ldots, u_{dn_d}) \right)^2, \tag{3.12}$$

ii) the Cramér-von Mises statistic is defined by

$$M(C_\theta, \mathbf{u}) = \sum_{n=1}^{N} \left(C_\theta(u_{1n}, \ldots, u_{dn}) - \hat{C}^*(u_{1n}, \ldots, u_{dn}) \right)^2, \tag{3.13}$$

iii) the Kolmogorov-Smirnov statistic is defined by

$$K(C_\theta, \mathbf{u}) = \sqrt{N} \cdot \max_{1 \le n \le N} \left| C_\theta(u_{1n}, \ldots, u_{dn}) - \hat{C}^*(u_{1n}, \ldots, u_{dn}) \right|. \tag{3.14}$$

The formula for the L^2-distance (3.12) indicates that the number of terms in the sum is N^d, which is exponential in the dimension of the sample points. Therefore, direct evaluation is only viable for small dimensions, namely $d = 2$.

In all three cases, small values indicate a better fit of the model. Our null hypothesis is

$$H_0 : C \in \{C_\theta : \theta \in \Theta\},$$

Large values of M or K support the rejection of the model C_θ of the copula. In one dimension, the limiting distribution of the Cramér-von Mises statistic of estimated distribution functions is known and can be tabulated. The same holds for the Kolmogorov-Smirnov statistic where we have an explicit formula to compute p-values in the one dimensional case. In the multi-dimensional case we have in the context of copulas, things get more complicated. It is theoretically possible to obtain approximate p-values of both the Cramér-von Mises statistic (3.13) and the Kolmogorov-Smirnov statistic (3.14). However, the limiting distribution depends on both the copula family that is assumed and the unknown parameters. Approximative tests using assumed limiting distributions are not advisable. In practice, one can compute approximate p-values using a parametric bootstrap procedure. Algorithm 3 has exactly this purpose. Because the distributions are approximated by parametric bootstrapping, the tests based on any statistic mentioned in this section are expected to hold their nominal level, which they empirically do in most cases. This has been analysed in [23], [21] and [22]. Of course it is also possible to simply compare the values of the statistics for different models and choose the model with the smallest value, as can be seen in [5] for the distance measure D_2. This approach does, however, not guarantee that the chosen model will be the most *likely* one (from a strictly statistical point of view).

The same statistics can be applied to the Rosenblatt transformed $\mathcal{R}_\theta(\mathbf{u})$. According to Theorem 2.46, the Rosenblatt transformed sample should be approximately independent with uniform margins under the hypothesis that $\mathbf{u}$ has copula C_θ. If we use the real copula C for the transformation, $\mathcal{R}_\theta(\mathbf{u})$ is exactly uniformly distributed. But because we can only use the estimated copula $C_{\hat{\theta}}$, the uniform distribution on the hypercube is just a heuristic and can be compared to the idea of a simple t-statistic being approximately normally distributed although the variance has been estimated. So it is not exactly true, but will suffice for large enough samples.

If we apply the statistics already seen to the transformed sample $\mathcal{R}_\theta(\mathbf{u})$, the assumed copula function C_θ in (3.12), (3.13) and (3.14) is then the independence copula $C^\perp$. Setting $\mathbf{v} = \mathcal{R}_\theta(\mathbf{u})$, we get the following statistics:

$$D_2^{\mathcal{R}}(C_\theta, \mathbf{u}) = \frac{1}{N^{d-1}} \sum_{n_1=1}^{N} \cdots \sum_{n_d=1}^{N} \left(\hat{C}^*_{\mathcal{R}_\theta}(v_{1n_1}, \ldots, v_{dn_d}) - v_{1n_1} \cdots v_{dn_d} \right)^2,$$

$$M^{\mathcal{R}}(C_\theta, \mathbf{u}) = \sum_{n=1}^{N} \left(\hat{C}^*_{\mathcal{R}_\theta}(v_{1n}, \ldots, v_{dn}) - v_{1n} \cdots v_{dn} \right)^2, \tag{3.15}$$

$$K^{\mathcal{R}}(C_\theta, \mathbf{u}) = \sqrt{N} \cdot \max_{1 \le n \le N} \left| \hat{C}^*_{\mathcal{R}_\theta}(v_{1n}, \ldots, v_{dn}) - v_{1n} \cdots v_{dn} \right|, \tag{3.16}$$

where $\hat{C}^*_{\mathcal{R}_\theta}$ is the empirical copula of $\mathbf{v}$ defined as in (3.11).

As before, small values of $D_2^{\mathcal{R}}$, $M^{\mathcal{R}}$ and $K^{\mathcal{R}}$ lead to the non-rejection of the hypothesis $\mathbf{U} \sim C_\theta$. In that case, we are led to have little reason for doubting our model. Approximate p-values can only be calculated by parametric bootstrapping since they, again, depend on the model and the unknown parameter θ. Tests based on the Rosenblatt transform of bivariate data have also been investigated in [10]. They empirically showed that a parametric bootstrap procedure is necessary to obtain useful tests in a semiparametric setting.

Another way of testing the goodness of fit is looking at the distribution of Kendall's distribution function K_C (as seen in Proposition 2.33). This can be especially useful in cases of Archimedean copulas where there is a closed form for K_C. But one needs to be careful that in general, there can be two different copulas with the same Kendall distribution function K_C. Therefore, tests using K_C are about a different null hypothesis $H_0' \supseteq H_0$ with

$$H_0' : K_C \in \{K_\theta : \theta \in \Theta\}.$$

Consequently, an inconspicuous test statistic on K_C with acceptance of H_0' does not imply acceptance of the null hypothesis H_0. Thus, tests based on K_C are not generally consistent. Nonetheless, a Cramér-von Mises statistic of K_C can be advisable in the Archimedean case because of its sheer convenience. Empirically, there are no advantages to a corresponding Kolmogorov-Smirnov statistic of K_C, as can be seen in [23].

Chapter 4

Algorithms

In this chapter, we will first gain a short overview on how copula computations can be implemented using the traditional approach. Afterwards, we introduce the CM-algorithm that improves the runtime significantly by confining the calculations to approximations and rank statistics. Lastly, we will see a simple algorithm on how to simulate test statistics that we have no known satisfactory analytical approximation for. In this lack of analytical feasibility, we resort to Monte Carlo simulation. It can be used to calculate the p-value of a given statistic. The first part of this chapter is largely taken from [35], while the second part follows the reasoning in [23].

4.1 Traditional Copula Computations

Consider a random vector $X = (X_1, \ldots, X_d)^\top$ with a joint distribution function F_X and marginal distribution functions $F_{X_1}, \ldots, F_{X_d}$. The vector

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_d \end{pmatrix}, \qquad U_k = F_{X_k}(X_k),\ 1 \leq k \leq d \tag{4.1}$$

has then the joint distribution C, i.e. the copula of X. Reversely, by taking different marginal distribution functions $F_{Y_1}, \ldots, F_{Y_d}$, one obtains

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix}, \qquad Y_k = F^{-1}_{Y_k}(U_k),\ 1 \leq k \leq d. \tag{4.2}$$

Then, Y has the marginal distributions Y_k and the same copula C as X, as follows from Sklar's Theorem 2.11.

Because most of the time, it is not possible to perform those steps analytically and obtain the distribution function F_Y in closed form, one has to use Monte Carlo scenarios. We start from a parametric N-dimensional

distribution F_X^θ with analytically feasible marginals $F_{X_k}^\theta$. We draw J Monte Carlo scenarios $\{x_{1j}, \ldots, x_{dj}\}$, $1 \leq j \leq J$. Using the marginal distribution functions $F_{X_k}^\theta$, we map the scenarios into copula scenarios.

To do the reverse, we again need parametric marginal distributions whose inverse distribution functions can be computed analytically or numerically. The step that restricts the variety of possible copulas the most is the separation step which requires the analytical form of the marginals. If we substitute those for the empirical distribution function of the margins, the computations become easy. In fact, the CMA here uses the same idea as the Canonical Maximum Likelihood (CML) estimation.

4.2 CM-Algorithm

The Copula Marginal Algorithm proposed in 2011 by Meucci [35] consists in two separate procedures, i.e. the separation step and the combination step. With the CM-algorithm, it is possible to use weighted scenarios to extract the copula from a joint distribution.

4.2.1 The Separation Step

The separation step in Algorithm 1 takes (probability-weighted) scenarios of a joint distribution as an input and returns the grade scenarios along with the empirical margins of the scenarios. It avoids the costly inversion of the marginal distribution functions using linear inter-/extrapolation.

Algorithm 1: CMA-Separation Step

Input: Scenario vectors $(x_{1n}, \ldots, x_{dn}) \in \mathbb{R}^d$ with probabilities p_n. $(1 \leq n \leq N)$

Output: Scenario grades $(u_{1n}, \ldots, u_{dn}) \in [0,1]^d$, $1 \leq n \leq N$, interpolation points of marginal distribution functions $\{(\bar{x}_{kj}, \bar{u}_{kj})\}_{1 \leq j \leq N}$, $1 \leq k \leq d$.

for $k = 1, \ldots, d$ **do**
  $(\bar{x}_{kj})_{1 \leq j \leq N} = (x_{ka_j})_j = \text{Sort}(\{x_{kn}\}_{1 \leq n \leq N})$;
  Compute b_n such that $\bar{x}_{kb_n} = x_{kn}$;
  Take the cumulative sum $\bar{u}_{kj} = \sum_{t=1}^{j} p_{a_t}$;
  Set $u_{kn} = \bar{u}_{kb_n}$;
end
return *grades* $\{u_{kn}\}$ *and margins* $\{(\bar{x}_{kj}, \bar{u}_{kj})\}_{1 \leq j \leq N}$.

To avoid the separated grades becoming 1 at the upper end, a scaling parameter $\alpha < 1$ is applied. Both $\{u_{kn}\}$ and $\{\bar{u}_{kj}\}$ are multiplied with α, which is chosen as $\frac{N}{N+1}$. In the case of an unweighted sample with $p_n = 1/N \; \forall n$,

this choice yields values of the grades that coincide with the empirical distribution function (3.4).

4.2.2 The Combination Step

The combination step in Algorithm 2 takes grade scenarios and marginal distribution (in the form of interpolation points) as an input and returns scenarios with the given marginals.

Algorithm 2: CMA-Combination Step

Input: Scenario grades $(u_{1n}, \dots, u_{dn}) \in [0,1]^d$ and interpolation points of marginal distribution functions $\{(\bar{y}_{kj}, \bar{u}_{kj})\}_{1 \le j \le N},\ 1 \le k \le d$.

Output: Scenario vectors $(y_{1n}, \dots, y_{dn}) \in \mathbb{R}^d$.

1 **for** $k = 1, \dots, d$ **do**
2 $\quad y_{kn} = \text{Interpolation}\big(u_{kn} \big| \{(\bar{u}_{kj}, \bar{y}_{kj})\}_{1 \le j \le N}\big)$;
3 **end**
4 **return** $(y_{1n}, \dots, y_{dn})$.

4.2.3 Explanatory Example

In order to understand the algorithms more clearly, let us illustrate a small example. We have the two-dimensional data $\mathbf{x}$ with 4 data points.

$$\mathbf{x} = \begin{pmatrix} 19 & 28 & 25 & 24 \\ 26 & 23 & 29 & 20 \end{pmatrix}, \qquad p = \begin{pmatrix} 0.1 & 0.2 & 0.4 & 0.3 \end{pmatrix}.$$

We want to separate the copula from the margins and then combine it with Pareto margins that satisfy

$$\mathbb{P}(X_i > x) = \frac{1}{x}, \quad \forall x \ge 1.$$

For the separation step, we sort each column and keep track of the a_{kn} and b_{kn} such that $\bar{x}_{kj} = x_{ka_{kj}}$ and $\bar{x}_{kb_{kn}} = x_{kn}$.

$$\bar{x} = \begin{pmatrix} 19 & 24 & 25 & 28 \\ 20 & 23 & 26 & 29 \end{pmatrix}, \quad a = \begin{pmatrix} 1 & 4 & 3 & 2 \\ 4 & 2 & 1 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 4 & 3 & 2 \\ 3 & 2 & 4 & 1 \end{pmatrix}.$$

We sort p according to a, put it into two rows and take the cumulative sum to get $\bar{u}$. We sort $\bar{u}$ according to b to get u.

$$\bar{p} = \begin{pmatrix} 0.1 & 0.3 & 0.4 & 0.2 \\ 0.3 & 0.2 & 0.1 & 0.4 \end{pmatrix}, \qquad \bar{u} = \begin{pmatrix} 0.1 & 0.4 & 0.8 & 1 \\ 0.3 & 0.5 & 0.6 & 1 \end{pmatrix},$$
$$u = \begin{pmatrix} 0.1 & 1 & 0.8 & 0.4 \\ 0.6 & 0.5 & 1 & 0.3 \end{pmatrix}.$$

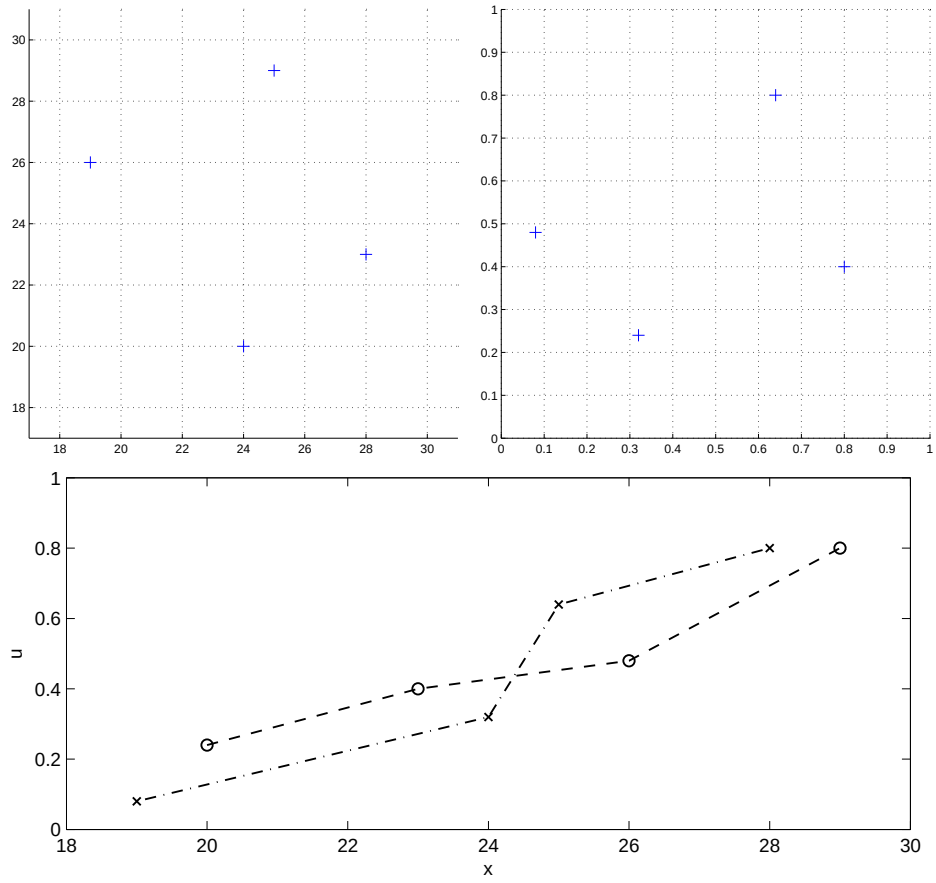

Figure 4.1: Explanatory separation example of the CM-algorithm in two dimensions with $N = 4$. Top left: original data. Top right: separated copula. Bottom: empirical marginal distribution functions obtained by interpolation.

As has mentioned before, a scaling factor of $\frac{N}{N+1} = \frac{4}{5}$ is applied to avoid reaching 1 at the upper end of the distribution function. This yields the output

$$\bar{u} = \begin{pmatrix} 0.08 & 0.32 & 0.64 & 0.8 \\ 0.24 & 0.40 & 0.48 & 0.8 \end{pmatrix}, \qquad u = \begin{pmatrix} 0.08 & 0.8 & 0.64 & 0.32 \\ 0.48 & 0.4 & 0.8 & 0.24 \end{pmatrix}.$$

This separation step is illustrated in Figure 4.1

Now for the combination step, we need interpolation points $\{\bar{u}_{kj}, \bar{y}_{kj}\}$. We choose

$$\bar{y}_1 = \bar{y}_2 = \begin{pmatrix} 1 & 2 & 3 & 5 & 10 \end{pmatrix}.$$

Then we can easily compute $\bar{u}$ by $1 - \bar{u}_{kj} = \frac{1}{\bar{y}_{kj}}$.

$$\bar{u}_1 = \bar{u}_2 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{2}{3} & \frac{4}{5} & \frac{9}{10} \end{pmatrix}.$$

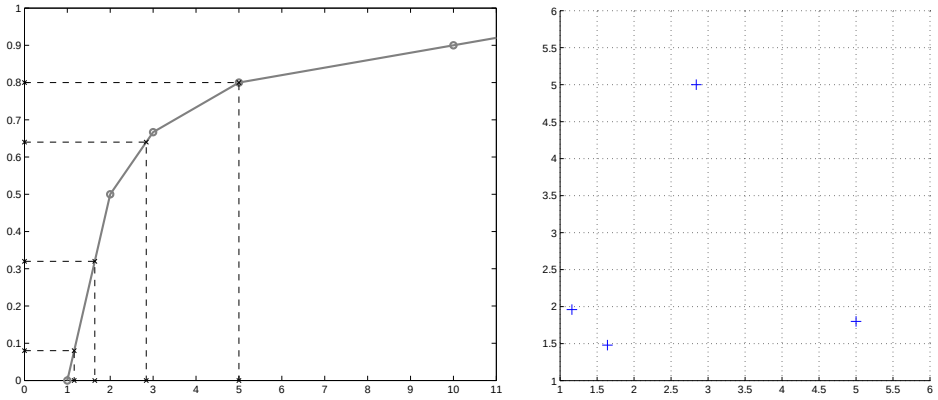

Figure 4.2: Explanatory combination example of the CM-algorithm in two dimensions with the data from Figure 4.1. Left: the interpolation process for dimension $k = 1$. Right: the recombined data.

We obtain the recombinated scenarios y_{kn} by interpolation. This yields

$$y = \begin{pmatrix} 1.16 & 5 & 2.84 & 1.64 \\ 1.96 & 1.8 & 5 & 1.48 \end{pmatrix}.$$

This combination step is illustrated in Figure 4.2.

4.2.4 Application

There are several possibilities to apply this algorithm in order to gain interesting new results. The strength of the algorithm lies in the fact that it is parameter-free. It can extract the copula from every multidimensional distribution and can glue together every margin, parametric or empiric, with every copula. Therefore, what is lost in accuracy is gained in the broadest possible range of copulas that can be achieved, within the range of numerical scenario computation. One application is to generate panic distributions to gain a so called panic copula, which we will see in the next chapter.

Another application is arbitrary copula transformations that are possible using this algorithm. By definition, a copula $C = F_U$ represents a random vector U that lives on the unit cube $[0, 1]^d$, with the margins U_k being uniformly distributed. Applying a transformation T on this random variable U will alter the dependence between the components of the vector. In general, however, it will also change both the marginal distribution and the domain of the random vector. To regain the copula of $T(U)$, another separation step is necessary, with possibly unpleasant marginal distribution functions to invert. As this does not pose a problem for the CMA, this procedure is easily done. In [35], Meucci proposes the following transformation: by the means of the combination algorithm, the vector U is first brought into the form

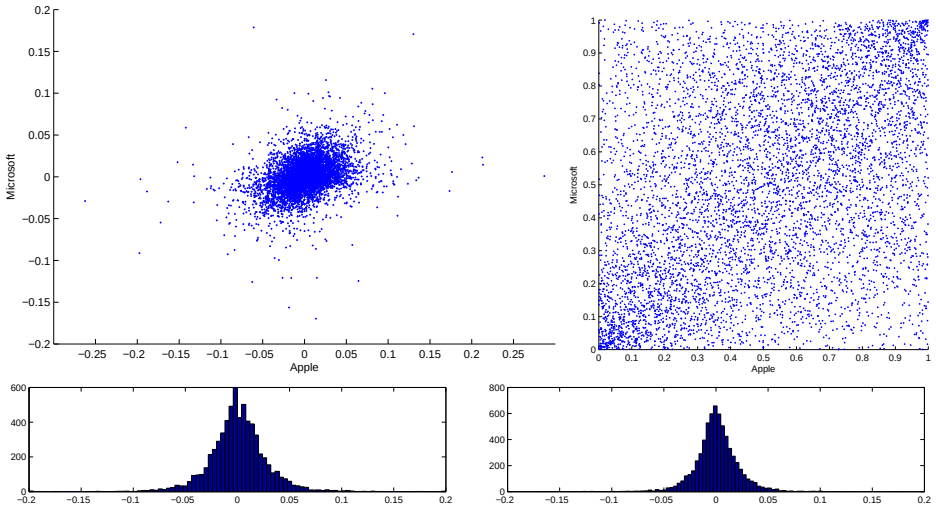

Figure 4.3: Top Left: $N = 6578$ daily log-returns of the stocks of *Apple* and *Microsoft*. Top Right: the separated copula. Bottom: histograms of the marginal distributions.

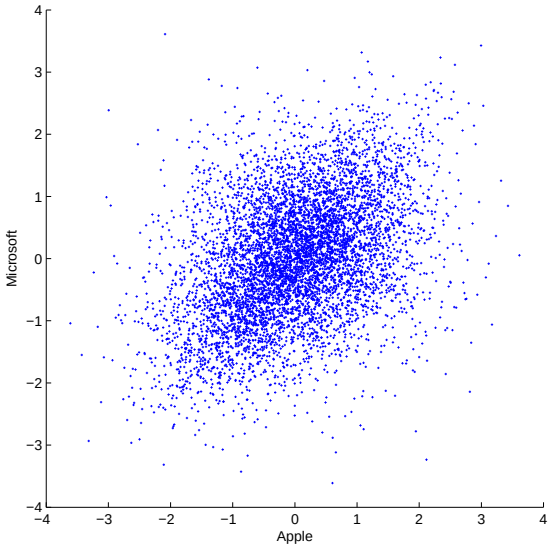

Figure 4.4: The empirical copula of Figure 4.3 recombined with standard normal margins.

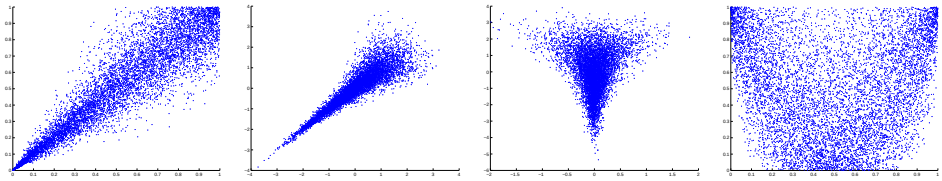

Figure 4.5: The two-dimensional Clayton(7) copula is transformed by a rotation of $\varphi = \frac{\pi}{4}$. From left to right: copula scenarios, scenarios after combination with standard normal margins, rotated scenarios, transformed copula after the separation step.

of standard normally distributed margins. Then a linear transformation is applied. The copula of the transformed vector is harvested by applying the separation algorithm.

Empirically, the rotation in two dimensions put forth by Meucci in [35] leads to a convergence to the independence copula $C^{\perp}$ if applied repeatedly with an angle $\varphi \neq \frac{k\pi}{2},\ k \in \mathbb{Z}$. For the Gaussian copula, this can easily be proven analytically.

While these kinds of computations are highly feasible even in higher dimensions, the question that remains to be answered is how to interpret said transformations. In fact, one might use the scheme mentioned above and fix the margins to arbitrary marginal distributions, apply any $\mathbb{R}^d$-valued function and extract the copula. But the explanation of what is happening during this reshuffling is unclear. Nonetheless, a rotation in the two-dimensional case is illustrated in Figure 4.5 in order to clarify the approach.

The strength of the algorithm lies also in the possibility of weighted scenarios. Therefore, we can especially target the tails with an increased number of Monte Carlo scenarios.

4.3 Parametric Bootstrapping

When looking at statistics to evaluate the goodness of fit (see Section 3.5), we are confronted with the task of approximating p-values for them. If the distribution of the test statistic is unknown to us or unpractical to calculate, we can resort to a *parametric bootstrapping* procedure. As with the CM-algorithm before, we gain universal applicability. Theoretical inspection of this algorithm can be found in [22].

We are interested in verifying the fit of a parametric copula to the pseudo-sample $\mathbf{u} = (u_1, \ldots, u_N)$ that it was obtained via the empirical distribution function of the margins in Algorithm 1. We assume that u_n are independent realisations of the multidimensional random variable $U \sim C$, where C is a copula. This is approximately true because of the asymptotic results regarding the empirical copula.

Assume a parametric copula C_θ is given. We estimate the parameter using a given estimator $\theta_0 = \hat{\theta} = T(\mathbf{u})$. The hypothesis is now that for the true copula C of the sample, the following holds:

$$H_0 : C \in \{C_\theta : \theta \in \Theta\}, \tag{4.3}$$

which we test one-tailed using the statistic $S(\cdot)$. Small values of S indicate a good fit. In Algorithm 3, the procedure for obtaining an approximate p-value

$$p = \mathbb{P}\left(S(\mathbf{U}, T(\mathbf{U})) > S(\mathbf{u}, \theta_0)\right) \tag{4.4}$$

for this hypothesis is explained, where $\mathbf{U} = (U_1, \ldots, U_N)^\top$ is a random matrix with $U_i \overset{\text{iid}}{\sim} C_{\theta_0}$. This p-value will in general not only depend on the parametric copula family C_θ but also on the true value of the parameter θ, see [23].

Algorithm 3: Parametric Bootstrapping

Input: Sample $\mathbf{u}$ of size N. Parametric copula C_θ, estimator $T(\cdot)$, test statistic function $S(\cdot)$ and a large integer K.

Output: Approximate p-value (4.4) for the hypothesis H_0 from (4.3).

$\theta_0 = T(\mathbf{u})$;
$S_0 = S(\mathbf{u}, \theta_0)$;
for $k = 1, \ldots, K$ **do**
    Simulate N data points $\mathbf{v}_k = (v_{kl})_{1 \le l \le N}$ from C_{θ_0};
    Estimate $\hat{\theta}_k = T(\mathbf{v}_k)$;
    Calculate $S_k = S(\mathbf{v}_k, \hat{\theta}_k)$;
end
$p = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}(S_k > S_0)$;
return p.

This algorithm lets us approximate the p-value in an unproblematic way. That is, if we have copulas that we can easily simulate from. Using the CM-algorithm, this includes all copulas of multivariate distributions that are easily simulated. What the writing of the algorithm conceals is that the computation of S and T might be difficult. The test statistic S in particular can prove to be very time-consuming. In the case of test statistics that involve calculations of computationally difficult $C_\theta(\cdot)$, a Monte Carlo approach can alleviate those difficulties. The copula function can again be approximated by simulation of a sample and taking the empirical copula thereof.

Chapter 5

Panic Copula

5.1 Motivation

With the CMA, we have a tool to generate copulas out of every joint distribution without large computational effort. For stress-testing, we want to create an asymmetric panic copula that especially targets the down-side. In this model, panic is triggered endogenously by the quantiles of a panic distribution. The panic copula is going to be extracted from a jointly distributed d-dimensional random variable X that is constructed using the approach mentioned by Meucci in [35]. The market is assumed to be in one of the two following states. First, there's the calm market $X^{(C)}$ which is normally distributed with mean 0 and some correlation matrix ρ_C. Then there is the panic stricken market $X^{(P)}$ which is again normally distributed with mean 0 but higher correlation ρ_P. The two states of the market are realised as two sets of independent random variables. If $X^{(P)}$ reaches under a certain threshold quantile $q \in (0,1)$, panic is triggered. So X is defined as follows:

$$
\begin{aligned}
X &\overset{d}{=} (\mathbb{1}_d - B) \circ X^{(C)} + B \circ X^{(P)}, \\
B_k &= \begin{cases} 1 & \text{if } \Phi(X_k^{(P)}) < q, \\ 0 & \text{otherwise,} \end{cases} \qquad \forall 1 \leq k \leq d, \\
X^{(C)} &\sim N_d\left(0, \rho_C\right), \qquad X^{(P)} \sim N_d\left(0, \rho_P\right).
\end{aligned}
$$

From this construction, it can be seen that there are three parameters to be chosen in this setting. The two correlation matrices and the panic threshold q. For dimensions higher than two, in order to keep the number of parameters one has to estimate low, it might be a good option to choose a panic correlation matrix ρ_P with high positive and homogeneous correlation r:

$$
\rho_P = \begin{pmatrix} 1 & r & \dots & r \\ r & 1 & \dots & r \\ \vdots & \vdots & \ddots & \vdots \\ r & r & \dots & 1 \end{pmatrix} \tag{5.1}
$$

5.2 Generalised Construction of a Panic Copula

For further examination, we define the panic copula in the most general way possible.

Definition 5.1 (Generalised panic copula)**.** Let F_Y and F_Z be two continuous distributions on $\mathbb{R}^d$ and $q \in (0,1)^d$. We define $b_k = F_{Z_k}^{-1}(q_k)$ for all $1 \leq k \leq d$. Define X by

$$X \overset{d}{=} (\mathbb{1}_d - B) \circ Y + B \circ Z, \tag{5.2}$$

$$B_k = \begin{cases} 1 & \text{if } F_{Z_k}(Z_k) < q_k, \\ 0 & \text{otherwise,} \end{cases} \qquad \forall 1 \leq k \leq d, \tag{5.3}$$

$$Y \sim F_Y, \qquad Z \sim F_Z,$$

where Y and Z are independent from each other.

The copula of X is called a panic copula with panic threshold q. If Y and Z are Gaussian distributions, then we call the copula of X a Gaussian panic copula. If they are Student-t distributions, we call it a t-panic copula.

It can be seen easily that the mean of X will be smaller than the weighted means of Y and Z, since Z is selectively taken only on its downside. That is, $\mathbb{E}[X_k] < q_k\mathbb{E}[Z_k] + (1-q_k)\mathbb{E}[Y_k]$. In the case of the Gaussian panic copula, the distribution X_k has a strictly negative mean of $-(2\pi)^{1/2}\exp(-b_k^2/2)$. Obviously, the mean has no direct influence on the copula of X.

As two special cases of panic copulas in Definition 5.1, we have the Gaussian panic copula (mentioned in the first section of this chapter) and the t-panic copula, where Y and Z are Student-t distributed. Compared to the Gaussian panic copula, the t-panic copula has two new parameters for the respective degrees of freedom ν_C and ν_P. Thus, the t-panic copula we obtain has a positive tail dependence, as shown in Corollary 5.6.

Proposition 5.2. *If X is a random vector defined by* (5.2) *and C is its copula, then all l-dimensional margins $C_{k_1,\ldots,k_l}$ of C are panic copulas. Furthermore, if C is a t-panic copula, then $C_{k_1,\ldots,k_l}$ is again a t-panic copula.*

Proof. Let $1 \leq l < d$ and $\{k_1,\ldots,k_l\} \subseteq \{1,\ldots,d\}$. If C is the copula of $(X_1,\ldots,X_d)^\top$, then $C_{k_1,\ldots,k_l}$ is the copula of $(X_{k_1},\ldots,X_{k_l})^\top$. The random vector B in equation (5.2) is defined component-wise, as can be seen in (5.3). Therefore, plugging $(Y_{k_1},\ldots,Y_{k_l})^\top$, $(Z_{k_1},\ldots,Z_{k_l})^\top$ and $(q_{k_1},\ldots,q_{k_l})^\top$ into (5.2) yields exactly $(X_{k_1},\ldots,X_{k_l})$. Thus, $C_{k_1,\ldots,k_l}$ is a panic copula constructed by the components of Y and Z. This proves both statements. □

Theorem 5.3. *If Z has continuous margins and we define X as in* (5.2)*, the distribution functions of X and its marginals X_k are given by:*

$$\begin{aligned} F_X(x) &= \sum_{I \uplus J = (1,\dots,d)} F_{Y_I}(x_I)\, \mathbb{P}((Z_J \le \min(x_J, b_J)) \wedge (Z_I \ge b_I)), \\ &= \sum_{I \uplus J = (1,\dots,d)} F_{Y_I}(x_I) \sum_{\emptyset \subseteq K \subseteq I} (-1)^{|K|}\, F_{(Z_J, Z_K)}(\min(x_J, b_J), b_K), \quad (5.4) \\ F_{X_k}(x_k) &= F_{Z_k}(\min(x_k, b_k)) + (1 - q_k) F_{Y_k}(x_k). \quad (5.5) \end{aligned}$$

Proof. This is easily seen by the law of total probability. We take the sum over all partitions of $(1, \dots, d)$ and condition on the events that exactly the components Z_J fall under the panic threshold.

$$\begin{aligned} F_X(x) &= \sum_{I \uplus J = (1,\dots,d)} \mathbb{P}(X \le x | (Z_I \ge b_I) \wedge (Z_J < b_J))\, \mathbb{P}((Z_I \ge b_I) \wedge (Z_J < b_J)) \\ &= \sum_{I \uplus J = (1,\dots,d)} \mathbb{P}(Y_I \le x_I)\, \mathbb{P}(Z_J \le x_J | (Z_I \ge b_I) \wedge (Z_J < b_J))\, \mathbb{P}((Z_I \ge b_I) \wedge (Z_J < b_J)) \\ &= \sum_{I \uplus J = (1,\dots,d)} \mathbb{P}(Y_I \le x_I)\, \mathbb{P}((Z_J \le \min(x_J, b_J)) \wedge (Z_I \ge b_I)) \end{aligned}$$

It is easily seen by applying the inclusion-exclusion principle that

$$\mathbb{P}(Z_J \le \min(x_J, b_J) \wedge Z_I > b_I) = \sum_{\emptyset \subseteq K \subseteq I} (-1)^{|K|} F_{(Z_J, Z_K)}(\min(x_J, b_J), b_K)$$

The marginal distribution is the one-dimensional special case of this formula. □

The marginal distribution function seen above is not to be easily inverted, with the notable exception of the normally distributed case or other cases with $F_{Y_k} = F_{Z_k}$ that are tabulated or can be analytically inverted. But with the CMA, exact inversion is not necessary in any case.

Corollary 5.4. *In the special case of $d = 2$ we have the distribution function:*

$$\begin{aligned} F_X(x_1, x_2) = {} & F_Z(\min(x_1, b_1), \min(x_2, b_2)) \\ & + F_{Y_1}(x_1)\, (F_{Z_2}(\min(x_2, b_2)) - F_Z(b_1, \min(x_2, b_2))) \\ & + F_{Y_2}(x_2)\, (F_{Z_1}(\min(x_1, b_1)) - F_Z(\min(x_1, b_1), b_2)) \\ & + F_Y(x_1, x_2) \bar{F}_Z(b_1, b_2). \end{aligned}$$

The density function is then:

$$\begin{aligned} f_X(x_1, x_2) = {} & f_Z(x_1, x_2) \mathbb{1}(x < b) \\ & + f_{Y_1}(x_1) \mathbb{1}(x_2 < b_2) f_{Z_2}(x_2) \mathbb{P}(Z_1 \ge b_1 | Z_2 = x_2) \\ & + f_{Y_2}(x_2) \mathbb{1}(x_1 < b_1) f_{Z_1}(x_1) \mathbb{P}(Z_2 \ge b_2 | Z_1 = x_1) \\ & + f_Y(x_1, x_2) \bar{F}_Z(b_1, b_2). \end{aligned}$$

In many cases, formula (5.4) can be written more conveniently using survival functions. Let's take the case of $d = 3$ as an example. With $x \in \mathbb{R}^3$ and $b \in \mathbb{R}^3$, we have the distribution function

$$\begin{aligned} F_X(x) = {} & F_Z(\min(x, b)) \\ & + F_{Y_1}(x_1) \left(F_{Z_{2,3}}(\min(x_{2,3}, b_{2,3})) - F_Z(b_1, \min(x_{2,3}, b_{2,3})) \right) \\ & + \cdots + \cdots \\ & + F_{Y_{1,2}}(x_{1,2}) \left(\bar{F}_{Z_{1,2}}(b_{1,2}) - \bar{F}_Z(b_{1,2}, \min(x_3, b_3)) \right) \\ & + \cdots + \cdots \\ & + F_Y(x) \bar{F}_Z(b), \end{aligned}$$

where the minimum is taken componentwise and the subindices $v_{i,j}$ denote the corresponding entries of the vector, i.e. (v_i, v_j).

Looking at equation (5.4), it is quite clear that evaluation of the distribution function in high dimensions becomes challenging, as we have an exponential number of summands. The measures of dependence of a panic copula can be calculated by the integral formulas (2.11) and (2.14) for τ and ρ_S. However, if there is no closed form for the distribution function of Y and especially Z, or if the marginals aren't easily inverted, it will be more convenient to approximate τ and ρ_S by Monte Carlo simulation and formulas (2.12) and (2.15). In dimensions higher than 2, it is also computationally faster to compute the value of C using the empirical copula of a Monte Carlo simulation, if a lot of evaluations need to be done, as is the case when calculating M and K in (3.13) and (3.14).

It is noteworthy that the parameter q_i of the panic threshold represents a quantile of the distribution Z_i but not of X_i. Therefore, it is no quantile of the panic copula. Although it is related to the quantiles of X_i, the connection is more intricate, as can be observed in equation (5.5). If Y_i and Z_i have the same distribution, the panic threshold $b_i = F_{Z_i}^{-1}(q_i)$ corresponds to the quantile $2q_i - q_i^2$ of X_i. If the distributions of Y_i and Z_i are different, this quantile may vary. This means that in fact, the influence of the panic distribution Z reaches over a significantly larger share of the copula than the parameter q might lead us to think. Therefore, we need to pay attention not to confuse q with an actual quantile of the dataset during the fitting of a panic copula to real data.

5.3 Tail Dependence of the Panic Copula

Let us take a look at the two-dimensional, general case of a panic copula. There, the calm market distribution Y has tail dependences $\lambda_L^{(C)}$ and $\lambda_U^{(C)}$, and the panic market distribution Z respectively $\lambda_L^{(P)}$ and $\lambda_U^{(P)}$. What can we say about the tail dependence of the panic copula λ_L and λ_U?

Theorem 5.5. *In the two-dimensional case, let $q_1 = q_2 = q$ and Y and Z such that their marginals are the same in both dimensions. Denote $b = F_{Z_1}(q)$. If the domain of the margins for both Y and Z are the same (let's say $[a, c]$, $a < c$), then the tail dependences of the general panic copula are calculated by:*

$$\lambda_U = \frac{\bar{F}_Z(b,b)}{1-q}\lambda_U^{(C)}, \tag{5.6}$$

$$\lambda_L = \lambda_L^{(P)} \cdot \lim_{x\to a^+} \frac{F_{Z_1}(x)}{F_{X_1}(x)} + \bar{F}_Z(b,b)\ \lambda_L^{(C)} \cdot \lim_{x\to a^+} \frac{F_{Y_1}(x)}{F_{X_1}(x)}, \tag{5.7}$$

given that the limits do exist.

Proof. Because $q \in (0,1)$ and the domains of Y_i and Z_i are the same, we have that $a < b < c$. For the upper tail dependence, we then have

$$\begin{aligned}
\lambda_U = \lambda_U(X_1, X_2) &= \lim_{u\to 1^-} \mathbb{P}(X_2 > F_{X_2}^{-1}(u) | X_1 > F_{X_1}^{-1}(u)) \\
&= \lim_{u\to 1^-} \mathbb{P}(X_2 > F_{X_1}^{-1}(u) | X_1 > F_{X_1}^{-1}(u)) = \lim_{x\to c^-} \mathbb{P}(X_2 > x | X_1 > x) \\
&= \lim_{x\to c^-} \frac{\bar{F}_X(x,x)}{\bar{F}_{X_1}(x)} = \lim_{x\to c^-} \frac{\bar{F}_Z(b,b)\bar{F}_Y(x,x))}{1-(F_{Z_1}(b)+(1-q)F_{Y_1}(x))} \\
&= \lim_{x\to c^-} \frac{\bar{F}_Z(b,b)\bar{F}_Y(x,x))}{(1-q)\bar{F}_{Y_1}(x)} = \frac{\bar{F}_Z(b,b)}{1-q}\lambda_U^{(C)}.
\end{aligned}$$

For the lower tail dependence, we can see that

$$\begin{aligned}
\lambda_L &= \lim_{x\to a^+} \frac{F_Z(x,x) + 2F_{Y_1}(x)(F_{Z_1}(x) - F_Z(b,x)) + \bar{F}_Z(b,b)F_Y(x,x)}{F_{X_1}(x)} \\
&= \lim_{x\to a^+} \frac{F_Z(x,x)}{F_{X_1}(x)} + 2\lim_{x\to a^+} \frac{F_{Y_1}(x)(F_{Z_1}(x)-F_Z(b,x))}{F_{X_1}(x)} + \bar{F}_Z(b,b)\lim_{x\to a^+} \frac{F_Y(x,x)}{F_{X_1}(x)} \\
&= \lim_{x\to a^+} \frac{F_Z(x,x)}{F_{Z_1}(x)}\frac{F_{Z_1}(x)}{F_{X_1}(x)} + 0 + \bar{F}_Z(b,b)\lim_{x\to a^+} \frac{F_Y(x,x)}{F_{Y_1}(x)}\frac{F_{Y_1}(x)}{F_{X_1}(x)} \\
&= \lambda_L^{(P)} \cdot \lim_{x\to a^+} \frac{F_{Z_1}(x)}{F_{X_1}(x)} + \bar{F}_Z(b,b)\lambda_L^{(C)} \cdot \lim_{x\to a^+} \frac{F_{Y_1}(x)}{F_{X_1}(x)}.
\end{aligned}$$

□

The condition that the margins of Y and Z have the same domain is crucial, as without it we would get $\lambda_L = \lambda_L^{(P)}$ if Z_i reaches lower than Y_i. Furthermore, degenerate cases where b lies outside the domain of Y_i only yield impractical results. This is not the aim of this construction and therefore we exclude such cases.

For the two cases of panic copulas we have seen, we can now state the following corollary.

Corollary 5.6. *In the two-dimensional case,*

1. *the Gaussian panic copula introduced at the beginning of this chapter has no tail dependence if* $\rho_C, \rho_P < 1$.

2. *the t-panic copula with* $Y \sim St_2(0, \rho_C, \nu_C)$ *and* $Z \sim St_2(0, \rho_P, \nu_P)$ *has a lower and upper tail dependence* $\lambda_L, \lambda_U > 0$ *if* $\rho_C, \rho_P \neq -1$.

And these results are obviously applicable in the multidimensional case when looking at pairwise tail dependence.

Proof. If Y and Z have the same marginal distributions, then the limits in (5.7) are trivially $0 < \frac{1}{2-q} < \infty$. Because the normal distribution has no tail dependence, the first part follows from (5.6) and (5.7).

For the second part, we have to check that one of the two limits in (5.7) is not zero. Suppose $\nu_P < \nu_C$, then

$$\begin{aligned} \lim_{x\to-\infty} \frac{F_{Y_1}(x)}{F_{Z_1}(x)} &\overset{\text{de l'H.}}{=} \lim_{x\to-\infty} \frac{f_{Y_1}(x)}{f_{Z_1}(x)} = 0, \\ \Rightarrow \quad \frac{F_{Z_1}(x)}{F_{X_1}(x)} &= \frac{1}{1 + (1-q)\frac{F_{Y_1}(x)}{F_{Z_1}(x)}} \overset{x\to-\infty}{\longrightarrow} 1. \end{aligned} \tag{5.8}$$

The case $\nu_P > \nu_C$ yields the same result for the other limit. Because $\lambda_L^{(C)}$ and $\lambda_L^{(P)}$ are not zero, we then see that $\lambda_L > 0$. □

Furthermore, we have seen that the limits in (5.7) only depend on the expression $L = \lim_{x\to-\infty} \frac{f_{Y_1}(x)}{f_{Z_1}(x)}$, which can be computed easily for many parametric distributions as long as the marginal distribution is known. Formula (5.7) then takes the form

$$\lambda_L = \lambda_L^{(P)} \cdot \frac{1}{1 + (1-q)L} + \lambda_L^{(C)} \cdot \frac{\bar{F}_Z(b,b)}{\frac{1}{L} + (1-q)}.$$

The limit cases with $L \in \{0, \infty\}$ are easily deduced from this formula by noting that $L \geq 0$.

5.4 The Parameters and their Estimation

The parameters to be chosen for the general panic copula defined in (5.2) consist of the parameters of Y and Z respectively, and the panic threshold q. Since all of the distribution functions are explicitly known, it is in theory possible to use the method of maximum-likelihood estimation for the parameters. Given the grades $u_{kn} = F_{X_k}(x_{kn})$ that are computed non-parametrically as in the CML estimation, we have the pseudo-log-likelihood

function

$$l(q, \theta^{(C)}, \theta^{(P)}|\mathbf{u}) = \sum_{n=1}^{N} \log f_X(F_{X_1}^{-1}(u_{1n}), \ldots, F_{X_d}^{-1}(u_{dn})) - \sum_{n=1}^{N}\sum_{k=1}^{d} \log f_{X_k}(F_{X_k}^{-1}(u_{kn})), \tag{5.9}$$

depending on the panic threshold q as well as the parameters $\theta^{(C)}$ of the calm market distribution Y and $\theta^{(P)}$ of the panic market distribution Z.

However, the fact that the likelihood function is in general not continuous along the planes of $x_k = b_k$ poses a problem in maximising it. Its plot looks very jagged (see Figure 6.5) because of the many small jumps that occur in every point q_k where a point u of the pseudo-sample corresponds to a point x that lies on said plane. Moreover, Theorem 3.1 is not applicable in this case because the log-likelihood function is not continuous in all parameters. Therefore, we have no guaranteed consistency or asymptotical normality for the parameter estimators of the panic copula. What weighs even more are the problems that the discontinuities cause for the maximisation. In practice, we would consider a hill-climbing procedure that separates the optimisation in q from the other parameters that don't cause discontinuities. For the optimisation in q, a pattern search method is applied. Lastly, we will restrict the panic copula model to equal panic thresholds $q_k = q$ for all components k. This will spare us the messy task of multidimensional optimisation of a non-continuous function. However, it has to be said that this is a very restrictive assumption and we only make it because it is necessary to render the estimation of q even practicable.

The function l is smooth with respect to all parameters but the panic threshold q, which makes the optimisation steps in finding the maximum for fixed q unproblematic. The disadvantage of the hill-climbing optimisation is that we will only find a local optimum. This can be problematic, given the jagged nature with the large number of discontinuities of the function to be maximised. Furthermore, different starting values can lead to different outcomes of the hill-climbing process. If we start with a value of q that is too small or a correlation $\Sigma^{(P)}$ of the panic distribution that is excessively high, the optimisation usually converges to a model with $\hat{q} \approx 0$ and unstable estimates of the panic distribution parameters $\hat{\theta}^{(P)}$ stemming from the fact that the data used for the estimation of the panic correlation includes very few points. If not enough data points fall under the panic threshold q, the estimation of $\theta^{(P)}$ becomes practically infeasible. Therefore, the case of $\hat{q} \approx 0$ can be taken as a convergence to the simpler model which is contained in the panic setting as a nested model. If this happens, new attempts using other starting points with lower correlation in the panic distribution and/or higher panic threshold should be made.

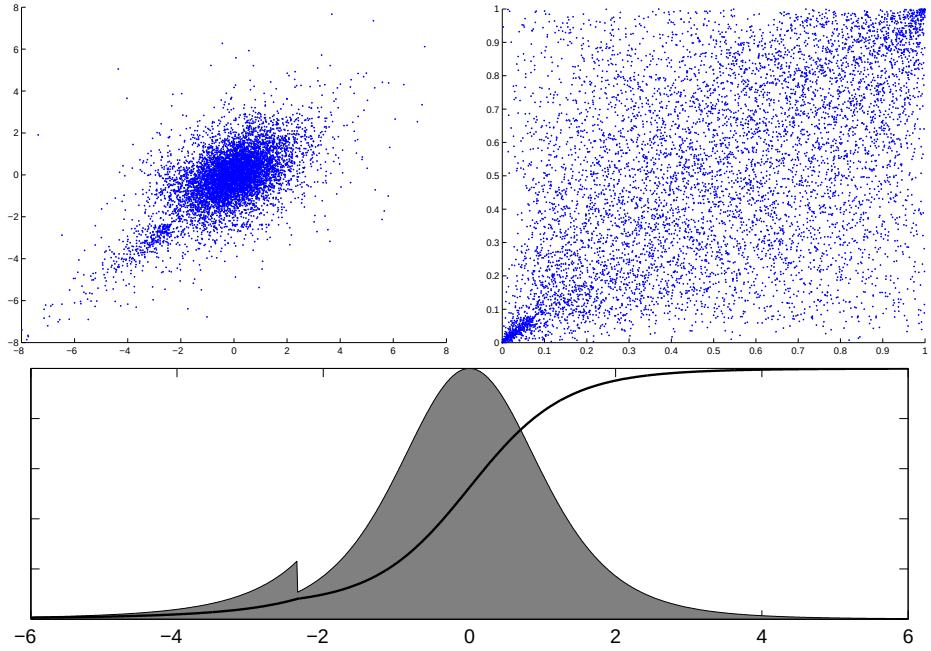

Figure 5.1: A t-panic copula is constructed. $\rho_C = 0.5$, $\rho_P = 0.98$, $\nu_C = 5$, $\nu_P = 3$, $q = 0.05$. Left: scenarios of X. Right: separated copula. Bottom: note the discontinuity in the density function of the margins of X. Because we have no point masses, the distribution function (bold) is continuous.

5.5 The t-Panic Copula

We are taking a look at the case where Y and Z are both from the multi-dimensional Student-t distribution. We want to apply maximum likelihood estimation for the parameters $q, \Sigma^{(C)}, \Sigma^{(P)}, \nu_C, \nu_P$. In this case, F_X and f_X depend on all of the parameters while F_{X_k} and f_{X_k} don't depend on the correlation matrices. In Figure 5.1, you can see a two-dimensional example of such a t-panic copula.

Because the Gaussian copula, the t-copula and the t-panic copula are nested models, we can expect to have more flexibility in our model, letting us achieve better results than the other models. From a practical point of view, it makes sense to restrict the model to positive correlation only. To examine the pairwise properties of the model, let us look at the two-dimensional case. Now clearly, the t-panic copula can handle asymmetric cases and asymmetric tail dependence. We know that the attainable τ-ρ_S-points of the t-panic copula have to be a superset of the attainable points of the t-copula. In lack of closed formulas to calculate the correlation measures exactly, we resort to numerical simulation of copulas with a number of different parameters. Somewhat surprisingly, the size of the attainable set

is significantly larger than before, as can be seen in Figure 5.2. This is of course a very desirable property. As mentioned in [12], the points that can be obtained by established copula models form a rather thin stripe along the line that is reachable using the Gaussian copula. In comparison, the attainable set of the t-panic copula accomplishes an enormous improvement. The limiting cases on the upper edge of the attainable set are the Gaussian copula, which fits into the more general model by setting $\nu_C = \infty$ and $q = 0$. Looking at the cases near the lower edge of the set (as shown in Figure 5.2), we can conclude that they are of little use in a practical setting. They have correlations ρ_C and ρ_P that are either 1 or 0 respectively. Intuitively, this was to be expected as it conforms with the already mentioned fact that the boundary of the τ-ρ_S-region (the dashed line in the plot) comprises only mutually completely dependent random variables. The closer we get to the edges, the more the copula starts to resemble a deterministic relation between the components, so to speak.

5.5.1 Estimation of the Parameters

If X is d-dimensional, we have $d(d-1)+3$ parameters that need estimating. While the likelihood function is smooth in the other parameters, the estimation of q introduces discontinuities and thus poses some difficulty for the optimisation process. Therefore, we propose a two-step procedure: assuming the panic threshold is somewhat low, we first estimate the parameters of the upper, calm market distribution using only the upper half of the dataset, which we assume to stem back largely to realisations of the regular market Y and not the panic driven market Z. We take the so found estimators $\hat{\Sigma}^{(C)}$ and $\hat{\nu}_C$ as starting points for the optimisation procedure.

In order to maximise the likelihood function (5.9), we have to invert the marginal distribution functions, which is difficult if Y and Z have different marginal distributions. This is avoided by applying the CMA combination step with appropriate margins to the grades $(u_{kn})_{k=1,2,\, 1\leq n\leq N}$ and then evaluating (5.9). Perhaps the biggest difficulty lies in the estimation of q. Because of the numerous discontinuities of the likelihood function, we can only try to reach a maximum with no telling if it is global or merely local.

Because of the difficult task of optimising a function that is not continuous in the variable q, we apply hill-climbing optimisation. That is, we optimise with respect to q either by using a pattern search algorithm or by evaluating the function on a mesh of different points and searching the optimum there. The important thing is to use a method that does not use the gradient and can cope with discontinuities. Then, we optimise with respect to the other parameters (with no discontinuities) and repeat the process. We have fairly good starting values for $\hat{\Sigma}^{(C)}$ and $\hat{\nu}_C$. The starting values for $\Sigma^{(P)}$ and ν_P need to be chosen somewhat arbitrarily because there is no way of estimating them without specifying the panic threshold q. For q, we

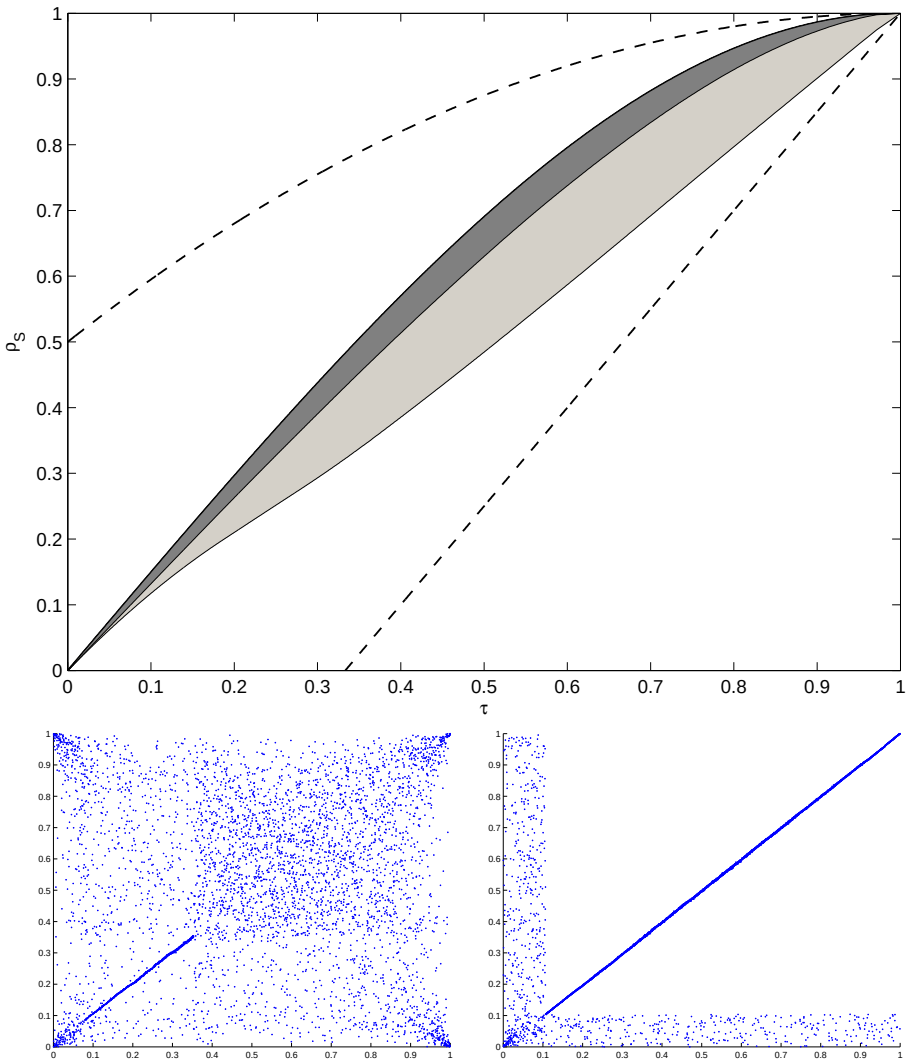

Figure 5.2: The attainable set within the τ-ρ_S-region (dashed bounds) for the t-copula (dark grey) and the t-panic copula with positive correlations (dark and light grey area). Below, simulated limit cases of t-panic copulas near and on the lower edge of the set are shown.

do not need a starting value per se if we evaluate the likelihood function on a mesh of points. We avoid making this delicate choice arbitrarily by starting the optimisation procedure with the maximisation in q on this mesh. Therefore, we only need an interval to search the maximum in.

Let us illustrate this on the example of a two-dimensional case. Here, the correlation matrices reduce to two single parameters ρ_C and ρ_P. We have a dataset of N random vectors given, i.e. $(x_{kn})_{k=1,2,\ 1\leq n\leq N}$.

1. We apply the separation step of the CM-algorithm, which then returns the grades $(u_{kn})_{k=1,2,\ 1\leq n\leq N}$ of the empirical copula.

2. To estimate the parameters of the calm market ρ_C and ν_C, we only look at indexes n_i, $1 \leq i \leq I$ for which $u_{1n_i} + u_{2n_i} > 1$. This means they lie in the upper half of the distribution. We mirror them along the axis $u_1 + u_2 = 1$ and get a new dataset

$$\begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix} = \begin{cases} (u_{1n_i}, u_{2n_i})^\top & \text{if } 1 \leq i \leq I, \\ (1 - u_{1n_{i-I}}, 1 - u_{2n_{i-I}})^\top & \text{if } I+1 \leq i \leq 2I. \end{cases}$$

   $\hat{\rho}_C$ and $\hat{\nu}_C$ are then easily obtained from $(v_{1i}, v_{2i})_{1\leq i\leq I}^\top$ by the means of the usual parameter estimation of the bivariate t-distribution. That is, compute Kendall's tau to obtain $\hat{\rho}_C$ and estimate ν_C with its MLE.

3. Choose reasonable starting values for $\hat{\rho}_P$ and $\hat{\nu}_P$, e.g. $\hat{\rho}_P = \sqrt[3]{\hat{\rho}_C}$ and $\hat{\nu}_P = \hat{\nu}_C$. In any case, $\hat{\rho}_P > \hat{\rho}_C$ should hold.

4. The pseudo-log-likelihood function $l(q, \rho_C, \rho_P, \nu_C, \nu_P|\mathbf{u})$ is then maximised using a hill-climbing approach. This means we maximise $l(\cdot|\mathbf{u})$ in alternation with respect to q and $(\rho_C, \rho_P, \nu_C, \nu_P)$.

For higher dimensions, the second step is performed on every pair (i, j) of dimension indices. The obtained matrix needs to be checked on positive definiteness. In case it is not positive definite, we adjust it using the eigenvalue method, see [40]. Because we are confident in the accuracy of our starting value for ρ_C, we can delay maximising $l(\cdot|\mathbf{u})$ with respect to ρ_C for the first steps, saving some runtime. Convergence usually occurs within the first few hill-climbing steps.

Alternatively to this pseudo-ML approach, it would also be possible to try to match the estimated tail dependence best. This means shifting the focus of fitting the copula to its tails. While this approach is well-known from two-dimensional Archimedean copulas, it is more involved here. We

have the following equations:

$$\hat{\lambda}_U(X_i, X_j) = \frac{\bar{F}_{(Z_i,Z_j)}(b,b)}{1-q} 2\, \bar{t}_{\nu_C+1}(s_C)\,, \tag{5.10}$$

$$\hat{\lambda}_L(X_i, X_j) = \begin{cases} 2\, \bar{t}_{\nu_P+1}(s_P) & \text{if } \nu_P < \nu_C, \\ \bar{F}_{(Z_i,Z_j)}(b,b)\; 2\, \bar{t}_{\nu_C+1}(s_C) & \text{if } \nu_P > \nu_C, \\ \frac{2}{2-q}\bar{t}_{\nu_P+1}(s_P) + \frac{2}{2-q}\bar{F}_{(Z_i,Z_j)}(b,b)\; \bar{t}_{\nu_C+1}(s_C) & \text{if } \nu_P = \nu_C, \end{cases} \tag{5.11}$$

where $s_C = \sqrt{\nu_C + 1}\frac{\sqrt{1-\Sigma_{ij}^{(C)}}}{\sqrt{1+\Sigma_{ij}^{(C)}}}$ and $s_P = \sqrt{\nu_P + 1}\frac{\sqrt{1-\Sigma_{ij}^{(P)}}}{\sqrt{1+\Sigma_{ij}^{(P)}}}$.

Now while this is a system of $2 \cdot \binom{d}{2}$ equations in $d(d-1)+3$ variables, we have to remark that there is no justification for altering the estimate $\hat{\Sigma}^{(C)}$ greatly just to fit the tails of the empirical copula. We can take different approaches of introducing the equations (5.10) and (5.11) into the estimation procedure. During the hill-climbing process above, we can choose to estimate $\Sigma^{(P)}$ either by maximising the likelihood function or by matching the tail dependences (5.11). Depending on the dimension d, some of these methods can be worth trying.

When handling said equations, one needs to be very careful. In the case that $\nu_P > \nu_C$, the parameter ν_P appears only in $\bar{F}_{(Z_i,Z_j)}(b,b)$ and can therefore be difficult to get a reasonable estimate of. In general, taking the tail dependence into consideration during the estimation procedure can be messy in its execution and the results will be mostly inferior to MLE. This is why we did not further pursue this scheme.

It is also an important thing to notice that in the case of the t-panic copula, all k-dimensional margins of the copula are again t-panic copulas. This offers us a way to avoid a problem that arises in higher dimensions. In practice, the likelihood function is difficult to evaluate if we have high-dimensional t-distributions in our model. Large numbers of $(d-1)$-dimensional distribution functions of the multivariate t-distibution need to be calculated with every evaluation of (5.9). This makes fitting a t-panic copula in dimensions $d > 3$ runtime-intensive and difficult, which is unsatisfactory. To circumvent this, it is possible to only look at the pairwise marginal copulas. They are, as we already know, also t-panic copulas and can be estimated rather quickly. What this yields is the matrices $\hat{\Sigma}^{(C)}$, $\hat{\Sigma}^{(P)}$ as well as matrices with the estimations of $\hat{\nu}_C$, $\hat{\nu}_P$ and $\hat{q}$ for each pair $X_i, X_j,\ 1 \le i, j \le d$. The estimated values of the panic threshold q can be alarmingly scattered with values reaching from 0 up to 0.5. Because we need to find one single value, different methods can be tried to approximate a good result. We take the widely scattered values of $\hat{q}$ as a sign that blindly taking the mean is not a good option. Instead, we obtain $\hat{q}$, $\hat{\nu}_C$ and $\hat{\nu}_P$ by

maximising

$$l^*(q, \Sigma^{(C)}, \Sigma^{(P)}, \nu_C, \nu_P | \mathbf{u}) = \sum_{i=1}^{d} \sum_{j=1}^{i-1} l_{ij}\big(q, \Sigma_{ij}^{(C)}, \Sigma_{ij}^{(P)}, \nu_C, \nu_P \big| u_i, u_j\big), \quad (5.12)$$

where $l_{ij}(\cdot)$ is the pseudo-log-likelihood function of the pair (u_i, u_j). If we want to restrict $\Sigma^{(P)}$ to homogeneous correlation matrices, we can use (5.12) to estimate $\Sigma^{(P)}$ as well.

5.6 Mixing of Two Copulas

If we have two copulas $C^{(C)}$ and $C^{(P)}$ that are easy to get simulated data from, we can use our usual method from formula (5.2) for $Y \sim C^{(C)}$ and $Z \sim C^{(P)}$ to obtain a mixed panic copula. Because the margins are uniform, the formulas from Theorem 5.5 then simplify to

$$\lambda_U = \frac{\bar{C}^{(P)}(q, q)}{1 - q} \lambda_U^{(C)}, \quad (5.13)$$

$$\lambda_L = \frac{1}{2 - q} \left(\lambda_L^{(P)} + \bar{C}^{(P)}(q, q)\ \lambda_L^{(C)} \right), \quad (5.14)$$

where q is the panic threshold from (5.3). In two dimensions, an example for such a panic copula might be using a Clayton copula as $C^{(C)}$ and a Gumbel copula as $C^{(P)}$. The parameter θ_C of the Gumbel copula can be estimated from Kendall's tau of the upper half of the distribution (as described above). q is estimated by maximum likelihood, while $\hat{\theta}_P$ for the Clayton copula can be set such that the tail dependence in equation (5.14) is matched best. If we want to take the equations (5.13) and (5.14) for both tail dependences into consideration, an iterative procedure will be necessary. CML estimation like in the t-panic copula case is also possible.

We might call the resulting construction an *Archimedean panic copula*, although this term is slightly misleading as the resulting copula is in fact not Archimedean any more. This approach is easily scaled up to higher dimensions. Neither the calculation of the likelihood function nor the maximisation poses a big problem in this setting. The computation of the estimates is highly feasible. The only drawback is the small number of merely 3 parameters. Moreover, the mixing of exchangeable copulas again yields an exchangeable copula. In contrast to the t-panic copula, higher dimensions don't pose an insurmountable problem to the computations and our estimation procedure. The distribution function of Y, Z and all its lower-dimensional margins are known and given in closed form. Therefore, no computational problems will occur. Still, the small number of only 3 parameters forbids using this model in high dimensions when fitting models to real data. In lower dimensions, they are an interesting model because of the easy implementation and straight-forward fitting procedure.

Chapter 6

Empirical Explorations and Application

In order to gain an understanding of how effectively the panic copula models are able to describe real data, we will take a look at stock data of German companies that we have harvested using Yahoo Finance. We will look at daily returns of sets of 3 stocks. In three dimensions, the t-panic copula is still relatively easy to compute. Therefore, we will concentrate on the three-dimensional case first. We will compare different models and different ways of estimation. We will also check the t-panic copula model against the simpler t-copula model to make sure the more complicated model does bring something new to the table and fits real data sufficiently better than the known models. Furthermore, we will try to extend our approach to arbitrary higher dimensions by using only the two-dimensional margins and the corresponding pairwise likelihood functions in the estimation process. In practice, the t-panic copula model will suffer from runtime constraints in high dimensions and the mentioned approach can mitigate this problem.

In a first step, we will empirically approach the problem of finding sufficiently good quantiles for testing the goodness of fit of the model. As we know, we can obtain approximate p-values using the parametric bootstrapping Algorithm 3. However, we might want to examine how large the dependence between the test statistic's distribution and both the parameters of the model and the size of the sample N is. If possible, our goal is to find a one-size-fits-all table of approximate quantiles for all sample sizes and any form of the three-dimensional t-panic copula.

Subsequently, we will ask the question of how different the t-panic copula is from the usual t-copula. We will use simulated data from both models and investigate how accurately the goodness-of-fit measures can distinguish between the two models. That is, we want to know whether the t-panic copula model is sufficiently different from the established model in order to justify the added inconvenience its implementation brings.

Lastly, we will not only concentrate on the t-panic copula but also take a look at the Archimedean panic copula models. With the Archimedean copulas we introduced, we can already use them for a panic copula construction in 36 different combinations. We will compare the results with the other models.

6.1 Finding Quantiles for Goodness-of-Fit Tests

In the one-dimensional case, the limiting distributions of the Cramér-von Mises statistic and the Kolmogorov-Smirnov statistic can be tabulated. This is a very handy feature. But as we have already come to know, the property does not hold for the multidimensional copula case. We have seen in Section 3.5 that the distributions of the Cramér-von Mises statistic $M(C_{\hat{\theta}}, \mathbf{u})$ and the Kolmogorov-Smirnov statistic $K(C_{\hat{\theta}}, \mathbf{u})$ are dependent on the chosen model, the estimator $\hat{\theta}$ and the specifically estimated parameters. Furthermore, they also depend on the sample size N. Because we don't have an analytic form of the distribution to compute the p-value of the statistic, we have to do simulations. This means that we have to find approximations by parametric bootstrapping, as described in Algorithm 3, i.e. we repeatedly simulate data from the copula $C_{\hat{\theta}}$, conduct an estimation of the parameters on the simulated data and compute the statistics M and K thereof. And for every sample size and every set of estimated parameters, we have to start a new bootstrapping procedure. This can prove computationally expensive. However, because of Theorem 2.44, we can expect the distribution of M and K to converge for big sample sizes N. Working with smaller sample sizes would of course speed up the parametric bootstrapping procedure.

If our goal is not to conduct a precise statistical test on the sample, it is also possible to just compare the values of the test statistics directly. As the values stem from different distributions for different models or parameters, they are not comparable per se. A smaller value for one model can be more improbable than a larger value for another model. Therefore, we are not comparing the likelihood of the sample being from one model or another. By comparing the values of the statistics directly, we just assume that smaller values will be better in any case. And especially when we are looking at the same model with different parameters, the distributions of the test statistics can be expected to be somewhat similar. While this cannot tell us on an absolute scale whether or not a model is a good fit for the sample, it can suggest which models fit better among a set of given models.

In this section, we want to try to evaluate if it is possible to find approximate quantiles for the statistics M and K which would be very useful in assessing the goodness of fit of a copula model. Because of the limitations explained above, we can expect our attempt to fail. We don't use any real market data in this section, we just simulate scenarios from the t-panic

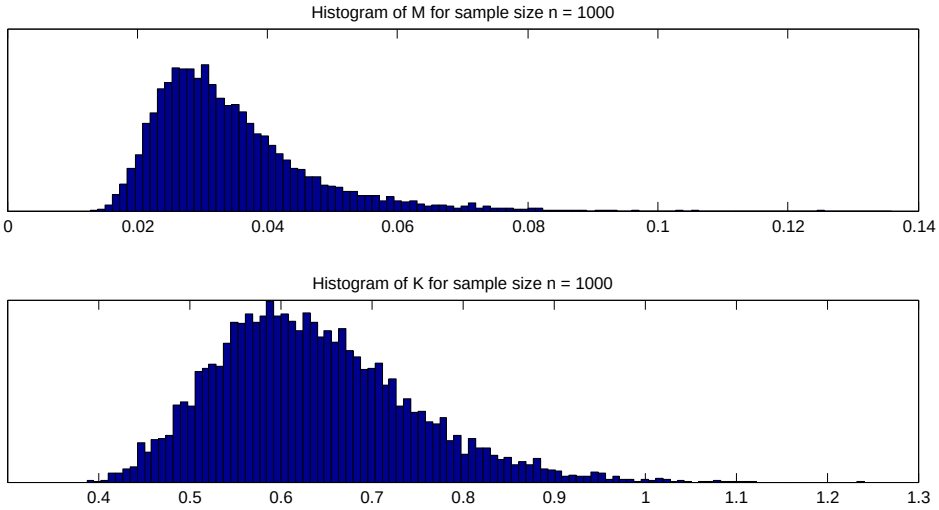

Figure 6.1: The bootstrapped distribution of the Cramér-von Mises statistic M and the Kolmogorov-Smirnov statistic K, based on 7500 samples from the t-panic copula with sample size $N = 1000$ (and varying parameters). This is for illustration purposes only, as the distributions vary depending on the sample size N and the parameters chosen.

copula.

In the case of the three-dimensional t-panic copula model, we conducted the following study: setting the sample size $N = 1000$ and drawing 750 samples from the t-panic copula for ten different parameter sets (amounting to a total of 7500 samples), we can see that the 95% bootstrap quantiles range from 0.053 to 0.065 for the Cramér-von Mises statistic M and from 0.81 to 0.87 for the Kolmogorov-Smirnov statistic K. The bounds of the confidence interval for the 95% quantile $q_{0.95}$ are then approximately given by the estimated quantiles of the levels $0.95 \pm 1.96\sqrt{K \cdot 0.95 \cdot 0.05}$. For $K = 1000$, this amounts to confidence intervals of typically ± 0.004 for M and ± 0.02 for K. This means that the quantiles are clearly not contained in the confidence interval of one single distribution. Therefore, the distributions of M and K will differ significantly for changing parameters. The distribution of the statistics are shown in Figures 6.1 and 6.2. The distribution of M somewhat resembles a scaled χ^2-distribution. However, it has a significantly higher skewness.

We have observed that the Cramér-von Mises statistic M shows a rather quick convergence, i.e. the quantiles don't change greatly with the sample size N. The Kolmogorov-Smirnov statistic K on the other hand seems to converge slowly with the quantiles rising quite strongly with growing N. For practical application, this means that regardless of the sample size and the specific parameters, a value M that is far beyond 0.065 should make us

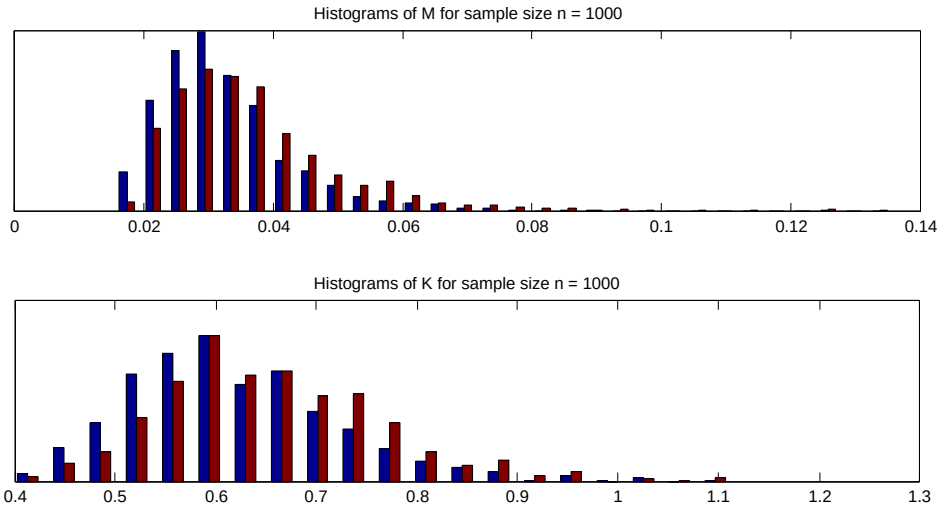

Figure 6.2: The bootstrapped distribution of M and K for two different sets of parameters, based on 750 samples from the t-panic copula with sample size $N = 1000$. In blue: a model with higher correlations $\Sigma_{ij}^{(C)}$ in the range of $0.5 - 0.7$ and $\nu_C = 5,\ \nu_P = 6$. In red: a model with lower correlations $\Sigma_{ij}^{(C)}$ in the range of $0.1 - 0.3$ and $\nu_C = 2,\ \nu_P = 3$.

suspicious.

Further investigation shows that the influence of the chosen estimator is not as large as expected. In fact, when the parameter $\Sigma^{(P)}$ is a homogeneous correlation matrix, the quantiles show only negligible variation depending on whether the homogeneous or the inhomogeneous estimator is used. Generalising this, the difference to using the more readily available estimation of the t-copula model is only marginally larger than the differences within the t-panic copula model. While they are of course nested models, it still surprises us to an extent that the restricted estimator in the t-copula model doesn't alter the quantiles more. One of the biggest influences on the distribution of M seems to be the degrees of freedom ν_C and ν_P. The lower those two parameters are, the more flattened the distribution of M becomes and therefore, the higher the 95% quantile of M is. The influence on K isn't as clearly evident. The quantiles for models with different parameters can be seen in Table 6.3.

On the possibility of finding easily available quantiles to compare the statistics M and K to, our conclusion has to be that it is not possible to tabulate values that can even remotely deliver reliable tests when applied to t-panic copula models of different sample sizes or parameters. This result was to be expected and is not a specific fault of the panic model. The t-copula yields similar results.

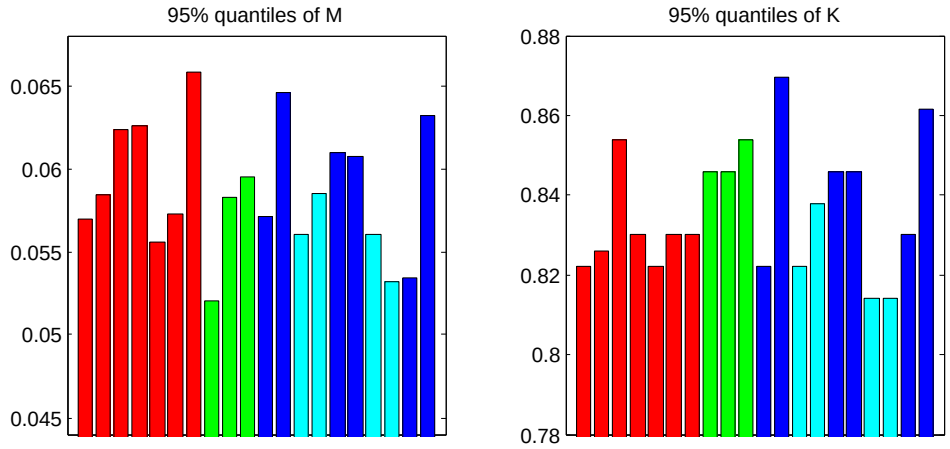

Figure 6.3: The 95% quantiles of M and K for twenty different sets of parameters, based on 750-1000 samples from the t-panic copula with sample size $N = 1000$. Models with the same parameters $\Sigma^{(C)}$ and $\Sigma^{(P)}$ are grouped together.

6.2 Differentiating between the t-Copula and the t-Panic Copula

Up to this point, all we did was to point out all the advantages a t-panic copula supposedly has in comparison to the decidedly simpler t-copula model. It is only reasonable to ask: can these promises be kept? How large exactly is the gap between the models? Can we distinguish between them even in the case of smaller sample sizes? In order to examine this important question, we simulate a range of different data from t-panic copulas and t-copulas and fit both models to the simulated data. Using the already discussed measures for the goodness of fit, we decide which models fits the data better. We then take a look at the hit rate, i.e. how often our decision was the right one.

We will test the three-dimensional case. The parameters for the simulated data will be randomly chosen. The decision on which model is to be considered better will be based on one of the measures we have already seen, i.e. AIC, BIC, M, K, $M^{\mathcal{R}}$ and $K^{\mathcal{R}}$. We decide on the model strictly by choosing the model with the smaller value of these statistics.

We fixed $K = 500$ sets of parameters, half of them with $q = 0$, corresponding to a plain t-copula model. The correlations in the matrix $\Sigma^{(C)}$ are uniformly distributed on $(0, 0.75)$, the correlations in the matrix $\Sigma^{(P)}$ are uniformly distributed on $(0.45, 1)$. In particular, it is noteworthy that $\Sigma^{(P)}$ is chosen as inhomogeneous. $\Sigma^{(P)}$ is inhomogeneous, as always. The matrices are then adjusted using the eigenvalue method, if necessary. The degrees of freedom are uniformly distributed integers satisfying $1 \leq \nu_C, \nu_P \leq 15$. q is uniformly distributed on $(0.005, 0.4)$. Our course of action was the following:

	AIC	BIC	M	K	$M^{\mathcal{R}}$	$K^{\mathcal{R}}$
$N = 100$ Panic	21.2	64.0	43.2	50.0	32.4	41.2
Non-Panic	30.4	8.0	43.6	38.8	48.0	50.0
$N = 200$ Panic	10.8	50.4	33.2	39.6	23.6	34.4
Non-Panic	27.6	3.2	36.4	34.0	40.0	40.0
$N = 300$ Panic	7.2	36.0	32.4	40.0	19.6	33.6
Non-Panic	16.0	2.0	23.2	27.6	35.2	33.2
$N = 400$ Panic	2.4	20.8	28.8	38.4	14.4	28.8
Non-Panic	12.8	1.2	24.0	24.0	28.8	27.2
$N = 500$ Panic	2.8	21.6	24.8	26.8	15.6	26.4
Non-Panic	10.8	0	25.2	24.4	25.2	24.4
$N = 600$ Panic	2.0	15.6	23.6	28.8	8.8	20.0
Non-Panic	10.0	0.4	24.0	23.6	27.2	25.2
$N = 800$ Panic	1.2	12.4	24.0	26.0	5.2	16.0
Non-Panic	7.6	0.4	18.0	19.6	17.6	16.8
$N = 1000$ Panic	1.2	8.4	18.4	20.4	4.8	7.6
Non-Panic	6.4	0	16.0	14.8	13.2	13.6
$N = 1500$ Panic	0.8	4.0	14.0	16.0	5.6	10.0
Non-Panic	2.0	0	13.2	11.6	9.2	11.2

Table 6.1: $K = 500$ different parameter sets have been used to simulate sample data of different sizes from the t-panic copula ("Panic") and the t-copula ("Non-Panic"). In the table, we see the rejection percentages of the actual copula model being rejected in favour of the other model. Ideally, this should be 0%. It can be seen that the numbers improve significantly for higher sample sizes.

1. We simulated samples of different sizes for every set of parameters.
2. We then performed both an ML estimation of a t-panic copula model and a standard estimation of a t-copula model.
3. Afterwards, we calculated the different measures mentioned above and decided on which model fits the data better by taking the smaller value. If the estimated $\hat{q}$ in the t-panic copula model was smaller than 0.005, we took it as a convergence to the non-panic model regardless of the values of the statistics.

The results can be seen in Table 6.1 where all the type I errors are shown. Each percentage in the table stems from 250 simulated samples.

The results state clearly that the accuracy of the decisions for a model is increasing with the growing sample size. In fact, it was to be expected that for small sample sizes, telling the models apart should be difficult. In a sample size of $N = 200$ and with a parameter $q = 0.15$, we can roughly expect a mere 15 sample points to fall under the panic threshold (in more than one dimension). This number of course depends on $\Sigma^{(P)}$, but we cannot expect more than $q \cdot N = 30$ points to stem back from the panic distribution in all dimensions. This makes the estimation of the 5 parameters intertwined with the panic distribution near impossible. Distinguishing between the models is therefore not practically feasible. The AIC seems to be biased towards the more complex model, whereas the BIC hardly ever chooses the more complex model when the underlying data is from the simpler model. M and K are of no use when we have smaller sample sizes. With small sample sizes, $M^{\mathcal{R}}$ and $K^{\mathcal{R}}$ consistently prefer the complex model more than their counterparts M and K do. They are more accurate in recognising the t-panic copula but also more likely to falsely prefer the complex model.

We can summarise:

- For samples smaller than $N = 300$, there is no reliable way of distinguishing between the models.
- For medium-sized samples, the AIC is very good at detecting the complex model and good at detecting the simple model.
- For medium-sized samples, $M^{\mathcal{R}}$ and $K^{\mathcal{R}}$ are more accurate than their counterparts M and K.
- In general, AIC and BIC perform best.

6.3 Fitting the Panic Copula to Real 3-Dimensional Data

In this section, we want to apply the new model to real stock data of three assets. We will fit several versions of the panic copula model along with the

regular t-copula model for comparison. They are compared using the already discussed measures. We will use the copula models to compute the Value-at-Risk for portfolios of three assets. Our model assumes all vectors of the daily stock returns to stem from the same distribution F. An assumption that is ubiquitous in the theoretical results on ML estimation is the independence of the random vectors, which in our case means that the logarithmic returns of the stocks are i.i.d. random vectors,

$$\begin{pmatrix} S_n^{(1)} \\ \vdots \\ S_n^{(d)} \end{pmatrix} = S_n \overset{iid}{\sim} F,$$

for a distribution F that we want to model using estimated marginal distributions and an estimated copula function. In this section, we set $d = 3$. The assumption about the distribution F is crucial from a theoretical point of view and can be seen as a minimum requirement on the historic data as the likelihood function we employ is only valid with a sample of independent random vectors. Note that this is not compatible with the somewhat popular GARCH models and other models that assume the log returns to be a heteroscedastic time series. For such approaches to the modelling of stock returns, see [1] and [15]. Of course, it is dubious whether this assumption actually holds, which has been noted by [5], who also suggests methods to avoid this strong assumption.

In order to calculate the Value-at-Risk of a portfolio, we need to model the joint distribution of the assets. After fitting their copula, we fit the margins using two different approaches. That is, we use a scaled t-distribution and a normal mixture model, as has been suggested in [2] and [28]. The conclusions on which approach is better vary. The scaled t-distribution includes fat tails but fails to address potential asymmetry. The mixture of normal distributions accounts for skewness, but has light tails. The marginal distribution hypotheses are evaluated using the Cramér-von Mises and Kolmogorov-Smirnov statistics, comparing them with the tabulated quantiles.

The Value-at-Risk (VaR) is a widely used risk measure in finance. In simple terms, it is the α-quantile of a loss distribution, given $\alpha \in (0, 1)$.

Definition 6.1. For $\alpha \in (0, 1)$, the *Value-at-Risk* at confidence level α of a portfolio with loss distribution L is given by

$$\begin{aligned} VaR_\alpha(L) &= \inf\{l \in \mathbb{R} : \mathbb{P}(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\} \\ &= F_L^{\leftarrow}(\alpha). \end{aligned}$$

If $\mathbb{E}[|L|] < \infty$, then the *expected shortfall* at confidence level α is defined as

$$ES_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 F_L^{\leftarrow}(u) du,$$

where $F_L^{\leftarrow}$ is the generalised inverse of F_L and thus the quantile function of L. The expected shortfall ES_α is also known as the Conditional Value-at-Risk $CVaR_\alpha$.

The Value-at-Risk does not give any information on how severe the incurred loss will be if it exceeds the quantile VaR_α. This flaw is mended by the expected shortfall. In the case of a continuous distribution L, the expected shortfall ES_α can be seen as the expected loss to occur in the case that the VaR_α is surpassed. That is, $ES_\alpha = \mathbb{E}[L|L \geq VaR_\alpha(L)]$. Typical values for the confidence level are $\alpha = 0.95$ or $\alpha = 0.99$. Depending on the field where the VaR is used, the time frame that the distribution L is looking at will encompass periods from 1 to 10 days in market risk management or one year in credit risk management. Because the VaR has found its way into the Basel II legislation, it is a very important tool of risk assessment. Further insights into it can be found in [13].

For an ordered sample $x_1 > \cdots > x_N$ of i.i.d. random variables from the distribution L, the sample version of the VaR is simply

$$\widehat{VaR}_\alpha = x_{\lfloor N(1-\alpha)\rfloor+1},$$

while the sample version of the expected shortfall is

$$\widehat{ES}_\alpha = \frac{1}{\lfloor N(1-\alpha)\rfloor + 1} \sum_{k=1}^{\lfloor N(1-\alpha)\rfloor+1} x_k.$$

Using these formulas, we can not only calculate the empirical estimates for given data but also use Monte Carlo simulation to compute approximate values for random variables L where the distribution function F_L is not easily available. It should be noted that the empirical estimator $\widehat{ES}_\alpha$ is rather unstable if N is not large enough because it relies on only $\lfloor N(1-\alpha)\rfloor+1$ data points, which can easily become a small number for large quantiles α. This is not a substantial problem in the ambit of Monte Carlo simulation since it is always possible to simulate more data. But the empirical estimation of the expected shortfall from a sample can be troublesome.

The first example we look at is the daily returns of the stocks of the German companies Daimler, Lufthansa and Merck. Using Yahoo Finance, we take the data from the period 1/2000 to 5/2016. We start from the daily adjusted closing prices $S_n^{(k)}$ of each stock and compute the logarithmic returns

$$x_{kn} = \log(S_{n+1}^{(k)}) - \log(S_n^{(k)}), \quad 1 \leq k \leq 3, \quad \forall n.$$

This makes for a sample of the size $N = 4105$. We examine the sample using the methods we already discussed. Because we use the CM-algorithm, we do not need to estimate the margins before estimating the copula. This semiparametric approach makes the copula estimator independent from the

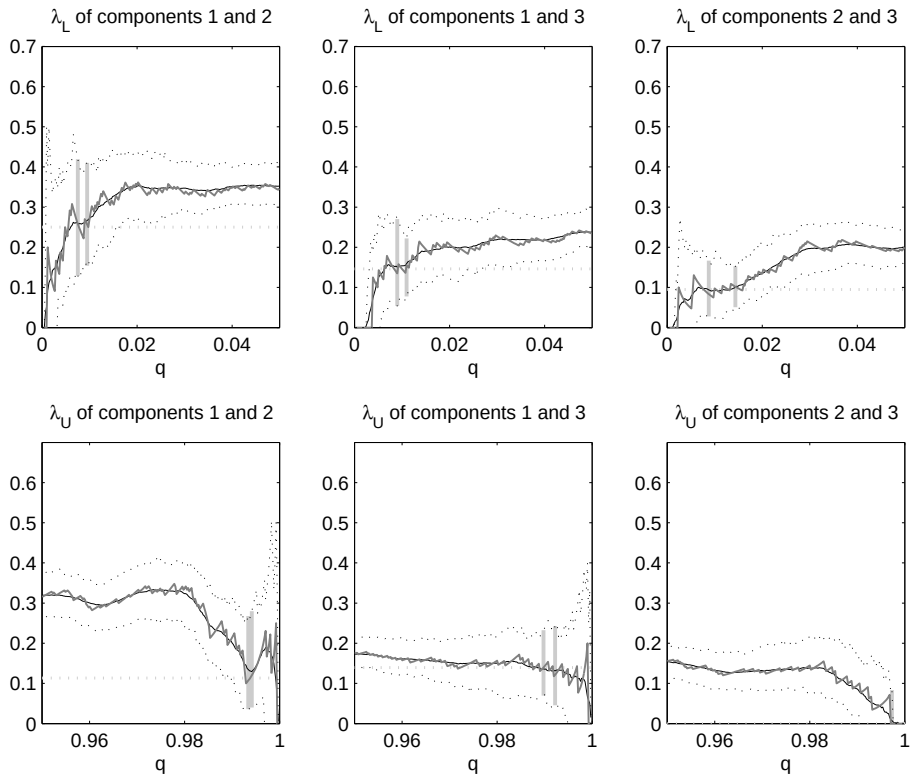

Figure 6.4: Empirical estimation of the tail dependences for daily stock returns from Daimler, Lufthansa and Merck with a sample size $N = 4105$. We see the trajectories $\hat{\lambda}$ with pointwise bootstrap confidence intervals. The vertical grey bars show the "stable" stretch of the trajectory, the horizontal dotted bar denotes the estimated values. Note that q represents the quantile used as an argument in the empirical quantile dependence function.

margin distribution model we choose. Because we will later want to compare our results using different kinds of margin distributions, this is a convenient and reasonable way of estimating the models. Otherwise we would have to estimate the copula separately for every set of marginal distributions we choose. The first data analysis gives us the following result:

$$\hat{\tau} = \begin{pmatrix} 1.0000 & 0.3573 & 0.2185 \\ 0.3573 & 1.0000 & 0.1809 \\ 0.2185 & 0.1809 & 1.0000 \end{pmatrix}, \quad \hat{\rho}_S = \begin{pmatrix} 1.0000 & 0.5024 & 0.3135 \\ 0.5024 & 1.0000 & 0.2613 \\ 0.3135 & 0.2613 & 1.0000 \end{pmatrix},$$

$$\hat{\lambda}_L = \begin{pmatrix} 1.0000 & 0.2500 & 0.1463 \\ 0.2500 & 1.0000 & 0.0948 \\ 0.1463 & 0.0948 & 1.0000 \end{pmatrix}, \quad \hat{\lambda}_U = \begin{pmatrix} 1.0000 & 0.1132 & 0.1395 \\ 0.1132 & 1.0000 & 0 \\ 0.1395 & 0 & 1.0000 \end{pmatrix}.$$

The lower tail dependence is higher that the upper tail dependence. Another sign for the stronger dependence in the lower half can be found in the

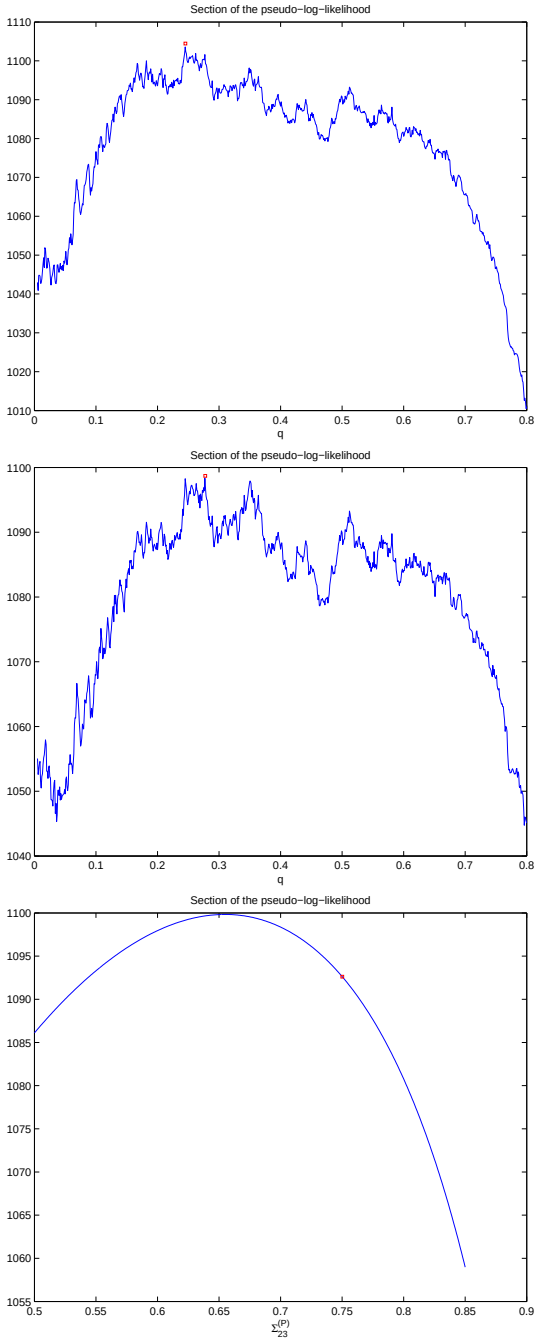

Figure 6.5: Estimation of t-panic copula models for daily stock returns from Daimler, Lufthansa and Merck with a sample size $N = 4105$. Top: a section of the pseudo-log-likelihood depending on q for the fitted model "ML-inhom.". Middle: the same section for "PW-inhom.". Please note that the q in these plots is the parameter of the panic copula and not a quantile per se, as explained in Section 5.2. Bottom: a section of the pseudo-log-likelihood depending on $\Sigma_{23}^{(P)}$ for the fitted model "ML-hom."

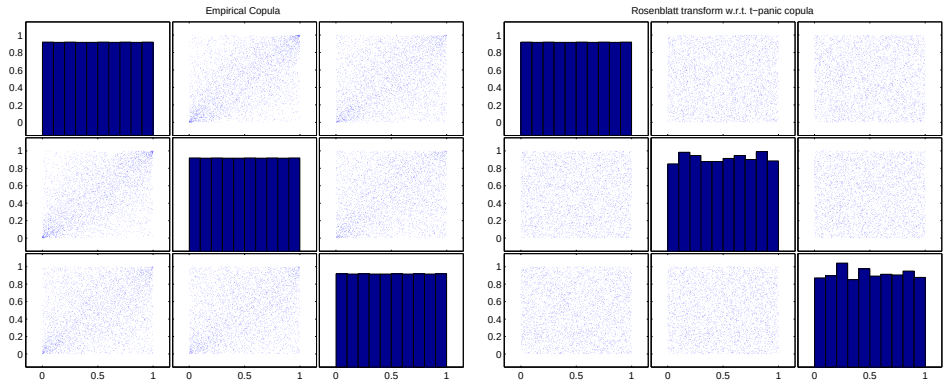

Figure 6.6: Left: Empirical copula of daily stock returns from Daimler, Lufthansa and Merck with a sample size $N = 4105$. Right: Rosenblatt transform of the of t-panic copula model "ML-inhom." fitted to the data. The Rosenblatt transform should be approximately uniformly distributed on $[0, 1]^3$.

correlation. If we only use the "upper half" of the distribution and mirror it to compute Kendall's tau as we did in the estimation procedure (see p.69), we obtain correlation coefficients that are higher (by about 0.04) than when using the "lower half". This discrepancy in the dependence on the upper and the lower half of the distribution justifies the use of a panic copula. Fitting a t-copula, we obtain the estimates $\hat{\nu} = 6$ and

$$\hat{\Sigma} = \begin{pmatrix} 1.0000 & 0.5322 & 0.3366 \\ 0.5322 & 1.0000 & 0.2804 \\ 0.3366 & 0.2804 & 1.0000 \end{pmatrix}.$$

We then fit two different t-panic copula models, i.e. with homogeneous and inhomogeneous $\Sigma^{(P)}$. We estimate them in two different ways. First, we do ML estimation, then we also try to maximise the sum of the pairwise likelihood functions l^* from equation (5.12), which leads to the models denoted by "PW-...". The results are shown in Table 6.2. Note that the p-values for the t-copula are zero, i.e. the hypothesis of a t-copula can be discarded at every confidence levels α. A small uncertainty remains because of the bootstrapping process needed to obtain the p-values, but on a confidence level of 95%, we can say that the distribution does not follow a t-copula. The ML estimates for the t-panic copula with inhomogeneous panic correlation ("ML-inhom.") are: $\hat{\nu}_C = 6,\ \hat{\nu}_P = 8,\ \hat{q} = 0.2454$,

$$\hat{\Sigma}^{(C)} = \begin{pmatrix} 1.0000 & 0.4860 & 0.2344 \\ 0.4860 & 1.0000 & 0.1688 \\ 0.2344 & 0.1688 & 1.0000 \end{pmatrix}, \quad \hat{\Sigma}^{(P)} = \begin{pmatrix} 1.0000 & 0.7840 & 0.6898 \\ 0.7840 & 1.0000 & 0.6343 \\ 0.6898 & 0.6343 & 1.0000 \end{pmatrix}.$$

We see that the estimate $\hat{\Sigma}^{(C)}$ in the t-panic copula is significantly lower than $\hat{\Sigma}$ is in the t-copula. This stems from the fact that the t-copula is symmetric,

	ML-inhom.	ML-hom.	PW-inhom.	PW-hom.	t-copula
$\hat{q}$	0.2454	0.1797	0.2776	0.1878	-
$\hat{\nu}_C$	6	6	5	5	6
$\hat{\nu}_P$	8	13	6	8	-
AIC	**-2190.9**	-2171.2	-2179.4	-2152.9	-2142.8
BIC	**-2134.0**	-2127.0	-2122.5	-2108.7	-2117.5
M	0.0491	0.0459	0.0578	**0.0330**	0.1930
$p(M)$	$0.091^{\cdot}$	0.135	0.030^{*}	0.472	0^{***}
K	0.8350	0.7336	1.0262	**0.6360**	1.2096
$p(K)$	0.117	0.324	0.010^{**}	0.695	0^{***}
$M^{\mathcal{R}}$	0.0508	0.0481	0.0588	**0.0476**	0.1654
$K^{\mathcal{R}}$	**0.8355**	0.8535	0.9706	0.8921	1.2566

Table 6.2: Fitted 3-dimensional t-panic copulas in comparison with the t-copula model, using the daily stock returns from Daimler, Lufthansa and Merck with a sample size $N = 4105$. "ML" means maximum likelihood estimation, "PW" means the maximisation of l^* in equation (5.12). The p-values are computed by parametric bootstrapping with $K = 1000$ iterations. Significance levels in the p-values are marked with stars. The best value in each line is printed bold.

while the t-panic copula can single out the higher correlated lower part of the distribution. In combination, we found τ and ρ_S of the t-panic copula and the t-copula to be virtually the same. They fit the empirical estimates equally well.

In Table 6.3, the same investigation is done using Archimedean panic copulas. We did not use all 36 possible combinations, instead we only used Clayton and Gumbel copulas in positions where their tail dependences seem reasonable in our setting. The Archimedean panic copulas do not provide a good fit for the data, as the statistics M and K are far beyond the quantiles we established to be approximately accurate in the first section of this chapter. Every measure of fit is worse than the values in Table 6.2. Therefore, we conclude that in general, Archimedean panic copulas do not fit 3-dimensional data well.

We compute the Value-at-Risk and the expected shortfall of a portfolio by using Monte Carlo simulation. That is, we simulate a high number of scenarios for the assets in our portfolio, calculate the value of the portfolio for each scenario and use the empirical estimates $\widehat{VaR}_\alpha$ and $\widehat{ES}_\alpha$ to compute approximate values for VaR_α and ES_α. In order to simulate scenarios for the assets, we need to fit the margins of their distribution. If we have the marginal distributions and the copula model, we can then use the CMA combination step to obtain scenarios for the assets. If we want to fit the marginal distributions, we need to decide on a model. The model of lognormal distributed returns has been proven to be inadequate in the past. Hence,

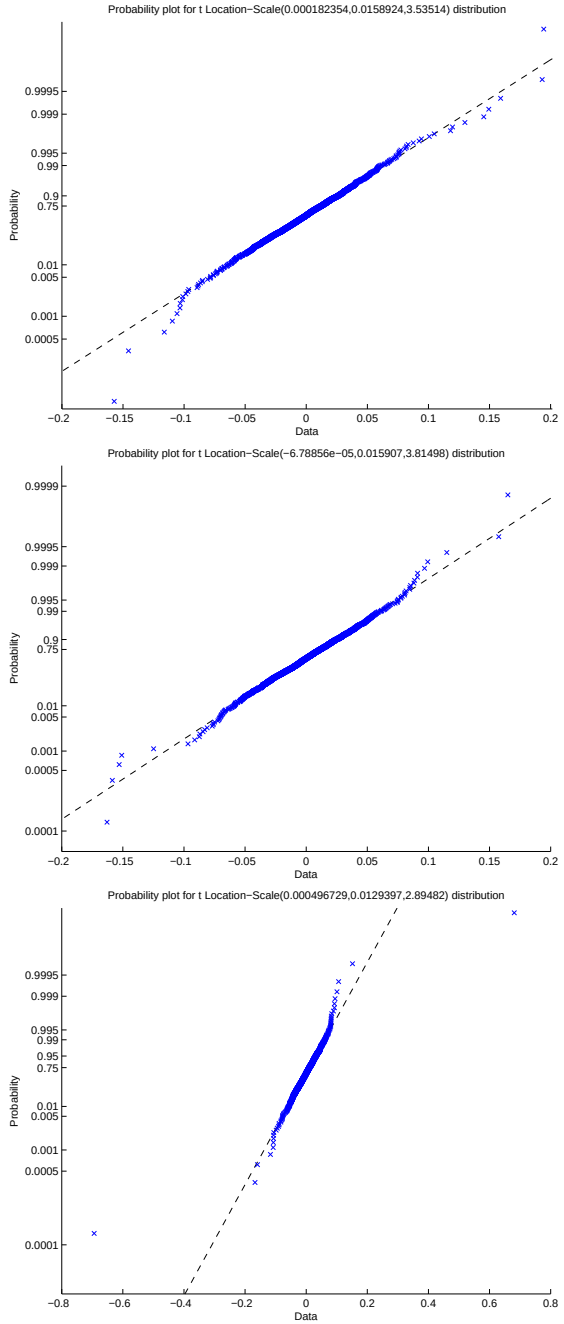

Figure 6.7: The marginal distributions of daily stock returns from Daimler, Lufthansa and Merck ($N = 4105$) plotted against the quantiles of location scaled t-distributions.

	(G,C)	(G,sG)	(sC,C)	(sC,sG)	(F,C)	(F,sG)	(sF,C)	(sF,sG)
q	0.258	0.258	0.341	0.246	0.441	0.263	0.263	0.184
θ_C	1.223	1.260	0.428	0.407	1.914	2.441	2.099	2.113
θ_P	0.994	1.584	0.982	1.599	0.972	1.480	0.893	1.514
AIC	-1837	-1814	**-1848**	-1755	-1742	-1744	-1754	-1678
BIC	-1818	-1795	**-1829**	-1736	-1723	-1725	-1735	-1659
M	0.302	0.239	0.430	0.361	0.249	**0.215**	0.218	0.237
K	2.002	1.951	2.287	2.376	1.974	**1.623**	1.720	1.834
$M^{\mathcal{R}}$	0.275	0.243	0.265	0.255	0.241	0.275	**0.227**	0.296
$K^{\mathcal{R}}$	2.300	2.279	2.446	2.528	2.219	2.289	2.096	**2.092**

Table 6.3: Fitted 3-dimensional Archimedean panic copulas in comparison, using the data from Table 6.2. The best value in each line is printed bold. The notation for the model is (calm distribution, panic distribution) with "C" for Clayton, "G" for Gumbel, "F" for Frank and "s" for the respective survival copula. Compared to the copula models from Table 6.2, they provide a very poor fit.

we will use two different models that have been mentioned as good models. We will then choose the model that matches the empirical distribution better. Fitting a location scaled Student-t distribution (with 3 parameters for each margin), we obtain good p-values for the Kolmogorov-Smirnov statistic in all components. Fitting a mixture of three normal distributions (with 8 parameters for each margin) using EM (see [45]), we get a better fit for the first two components, and a significantly worse fit for the third component. But the distribution hypotheses are still both contained in a 95% confidence interval. We choose the best fitting copula model to be "ML-inhom.", being the model with the best BIC. As for the marginal distributions, we choose the better fit according to the BIC as well. Therefore, the location scaled t-distribution is the distribution we choose, as it surpasses the mixed normal distribution in every component. We now take a look at a portfolio of $-2 \cdot S^{(1)} + 5 \cdot S^{(2)} + S^{(3)}$ with the stock prices of 29/4/2016, when the portfolio is worth €27.44. Looking at a time frame of 1 day, the VaR and ES of the loss are shown in Table 6.4 for the various models.

For comparison, the values calculated assuming a multivariate Gaussian distribution of the logarithmic stock returns are also shown in the table. As expected, it becomes apparent that assuming X to be multivariate Gaussian is not an adequate assumption. This was already known, as the tail dependence of the empirical data is not modelled in an assumed Gaussian distribution. The group of fitted t-panic copulas and the t-copula don't show much variation among themselves. As can be seen in Table 6.4, the model that we chose as the best one, "ML-inhom." with location scaled t-distributed margins, fits the empirical VaR quite well. It is marked by a star on the left side. The larger differences in the very high quantiles might stem from few data points actually realising beyond this quantile. The same

VaR_α	**0.95**	**0.98**	**0.99**	**0.995**	**0.999**	**0.9995**	**0.9999**
Empiric	4.1491	5.8229	7.6031	8.8634	13.7349	*17.8384*	*38.2686*
Gaussian	4.5981	5.7558	6.4963	7.2245	8.7290	9.4074	10.6049
t-margins							
∗ ML-inhom.	4.1371	5.9275	7.4572	9.6046	15.7677	18.8761	33.3190
ML-hom.	4.2055	6.0510	7.6914	9.6176	15.9619	19.6826	30.7414
PW-inhom.	4.0737	5.8808	7.4728	9.4595	15.8099	20.0573	30.5614
PW-hom.	4.1982	6.0498	7.6346	9.5082	14.9763	17.9812	26.4161
t-copula	4.0314	5.7857	7.2609	8.9393	14.6477	17.8036	34.9361
(sC,C)	4.4478	6.1230	7.5077	9.1892	14.4637	16.8472	26.2728
(F,sG)	4.5954	6.5002	8.0743	9.9603	16.9376	20.6442	34.4824
Mixed n.							
ML-inhom.	4.1297	5.7704	7.3641	9.2617	15.8352	19.8899	34.9253
ML-hom.	4.2321	6.0070	7.5501	9.4552	15.5133	19.8079	37.7194
PW-inhom.	4.1140	5.8660	7.3346	9.1374	15.6143	20.3083	32.4279
PW-hom.	4.1883	5.9355	7.4830	9.1739	15.7804	21.0979	34.9133
t-copula	4.0866	5.7406	7.1537	8.8813	14.1476	18.7571	32.8964
(sC,C)	4.4911	6.0592	7.3314	8.7405	13.8472	18.3718	34.0645
(F,sG)	4.5706	6.3827	7.9735	9.9230	16.7628	20.6833	35.6060

ES_α	**0.95**	**0.98**	**0.99**	**0.995**	**0.999**	**0.9995**	**0.9999**
Empiric	6.3573	8.4962	10.4702	12.7985	22.2100	*30.3390*	-
Gaussian	5.7786	6.7816	7.4832	8.1514	9.5446	10.0802	11.0321
t-margins							
∗ ML-inhom.	6.4543	8.8570	11.1196	13.8656	22.8783	28.5797	48.5438
ML-hom.	6.5632	8.9618	11.1667	13.8560	22.4701	27.3782	43.5599
PW-inhom.	6.4074	8.7811	10.9994	13.6518	22.5541	27.6539	43.6252
PW-hom.	6.4852	8.7560	10.7900	13.1661	20.4314	24.6473	38.4472
t-copula	6.2181	8.4169	10.4132	12.8551	21.1526	26.0178	40.0929
(sC,C)	6.5253	8.5826	10.4311	12.6317	19.1228	22.6874	35.6810
(F,sG)	7.0061	9.4511	11.7133	14.5425	24.0126	29.2575	46.8836
Mixed n.							
ML-inhom.	6.3551	8.6698	10.8789	13.5990	23.0888	28.7235	43.4935
ML-hom.	6.5262	8.8770	11.0890	13.8435	23.4350	29.5465	42.2528
PW-inhom.	6.3625	8.6496	10.8090	13.5090	22.6615	27.9550	39.3457
PW-hom.	6.4515	8.7721	10.9386	13.6752	23.4931	28.8952	41.6581
t-copula	6.1978	8.3537	10.3643	12.8501	21.7761	27.4933	40.9945
(sC,C)	6.4652	8.4284	10.2524	12.5545	21.3092	27.1075	38.7166
(F,sG)	6.9274	9.3255	11.5957	14.3904	23.5435	28.5541	40.8189

Table 6.4: VaR and ES of a portfolio on different confidence levels for the copula models from Tables 6.2 and 6.3 and different margins. For comparison, the empirical values and the corresponding values assuming a multivariate Gaussian distribution of X are shown. The values are calculated by simulating $100,000$ cases.

	ML-inhom.	ML-hom.	PW-inhom.	PW-hom.	t-copula
$\hat{q}$	0.1137	0.1103	0.1017	0.1017	-
$\hat{\nu}_C$	4	4	4	4	4
$\hat{\nu}_P$	6	8	7	7	-
AIC	**-4696.8**	-4696.2	-4692.0	-4692.6	-4679.6
BIC	-4639.9	-4652.0	-4635.2	-4648.4	**-4654.4**
M	0.1494	0.1092	0.0917	**0.0687**	0.1439
$p(M)$	0^{***}	0.004^{**}	0.005^{**}	0.028^{*}	0^{***}
K	1.0106	0.8857	0.9326	**0.8389**	1.4164
$p(K)$	0.012^{*}	$0.050^{\cdot}$	0.031^{*}	0.103	0^{***}
$M^{\mathcal{R}}$	0.1114	0.1275	**0.0973**	0.1165	0.2122
$K^{\mathcal{R}}$	**1.1012**	1.4132	1.1719	1.2832	1.6877

Table 6.5: Fitted 3-dimensional t-panic copulas in comparison with the t-copula model, using the daily stock returns from Deutsche Bank, Deutsche Telekom and Siemens with a sample size $N = 4105$. The p-values are computed by parametric bootstrapping with $K = 1000$ iterations. Significance levels in the p-values are marked with stars. The best value in each line is printed bold.

reason might cause the differences in the ES to be larger. In fact, no reasonable estimate is possible for the empirical $ES_{0.9995}$ because the base of data for it is only 2 data points. The empirical estimates which are based on too few data points are printed in cursive. Interestingly, the Archimedean panic copula model (sC,C) fits the empirical estimates of VaR and ES quite well in spite of the low number of 3 parameters in that model. However, the goodness-of-fit measures of the Archimedean panic copula do not compare to those of the t-panic copula, which makes their use unjustifiable.

In Table 6.5, we apply the same copula fitting scheme to the stocks of Deutsche Bank, Deutsche Telekom and Siemens. In this case, the BIC favours the t-copula. Nonetheless, the p-values for the t-copula are zero. Since all of the t-panic copulas also have very low p-values, we can not confidently choose one model and discard the others. The picture is not clear, as the single values contradict each other.

In Table 6.6, the copula models for the stocks of Continental, Fresenius and Adidas are compared. We see that even the hypothesis of a t-copula holds on a confidence level of 95%. The BIC also favours this distribution. Interestingly, the estimates for the panic threshold q in the t-panic copula models is very low, which makes the estimates for the correlation matrix $\Sigma^{(P)}$ rather unstable, with $\hat{\Sigma}^{(P)}_{12}$ being even lower than $\hat{\Sigma}^{(C)}_{12}$. This goes against our initial assumption that the correlation in the panic distribution should be higher than the correlation in the calm market distribution. However, this is more a cosmetic problem that does not cause any trouble. The very

	ML-inhom.	ML-hom.	PW-inhom.	PW-hom.	t-copula
$\hat{q}$	0.0059	0.0031	0.0059	0.0059	-
$\hat{\nu}_C$	5	5	5	5	5
$\hat{\nu}_P$	5	7	5	5	-
AIC	-1404.7	**-1407.5**	-1404.3	-1404.0	-1401.6
BIC	-1347.8	-1363.2	-1347.4	-1359.8	**-1376.4**
M	0.0911	0.0517	**0.0441**	0.0686	0.0503
$p(M)$	0.001^{***}	$0.091^{\cdot}$	0.139	0.010^{**}	$0.060^{\cdot}$
K	0.9795	**0.7727**	0.7883	0.9327	0.8312
$p(K)$	0.016^{*}	0.249	0.176	0.037^{*}	0.104
$M^{\mathcal{R}}$	0.0426	0.0424	0.0404	**0.0392**	0.0429
$K^{\mathcal{R}}$	0.7803	0.7634	0.7704	**0.7262**	0.7940

Table 6.6: Fitted 3-dimensional t-panic copulas in comparison with the t-copula model, using the daily stock returns from Continental, Fresenius and Adidas with a sample size $N = 4104$. The p-values are computed by parametric bootstrapping with $K = 1000$ iterations. Significance levels in the p-values are marked with stars. The best value in each line is printed bold.

low estimates for q can be seen as evidence for the fit of the t-copula. That is, the panic copula does not provide any significant advantages against the simpler model in this case.

It is also worth noting that if $\nu_P > \nu_C$ in the t-panic copula model, then the model will have $\lambda_L < \lambda_U$ by construction, as can be seen from formula (5.11) and (5.10). This is exactly the opposite of what we want to achieve and poses a significant flaw in the model. It is worth considering the addition of the restriction $\nu_P \leq \nu_C$ in the estimation process to prevent this from happening. Without it, the maximum likelihood estimates $\hat{\nu}_P$ tend to be larger than $\hat{\nu}_C$. Adding the restriction would force the lower tail dependence of the model to be larger than the upper tail dependence. We have implemented this in the next section with mixed results. The fit of the model decreased significantly and the lower tail dependence was, generally speaking, too high compared to the empirical estimates. This can be seen in the following section.

What is important to notice in these examples is that the difference between fitting the three-dimensional model using the pseudo-log-likelihood function and fitting it using the function l^* from equation (5.12) is not very large. This leads us the the conclusion that in the higher-dimensional cases where the likelihood function of a panic copula is not that easily computed anymore, the approach of using the sum l^* of the pairwise likelihood functions is a good way of fitting the model.

	PW-inhom.	PW-hom.	t-copula
# Parameters	33	19	16
$\hat{q}$	0.1066	0.1040	-
$\hat{\nu}_C$	5	4	5
$\hat{\nu}_P$	7	7	-
l^*	**10126.6**	10042.4	10004.1
M	**0.1592**	0.2450	0.2449
K	**1.1854**	1.4010	1.4315

Table 6.7: Fitted 6-dimensional t-panic copulas in comparison with the t-copula model, using the same daily stock returns from Tables 6.2 and 6.5 with a sample size $N = 4105$. The best value in each line is printed bold.

6.4 Fitting the Panic Copula to Higher-Dimensional Data

We have seen fitting the panic copula models to three-dimensional data. In the case of three dimensions, the likelihood function can still be dealt with. While Theorem 5.3 gives us the explicit formula (5.4) for the distribution function of the panic copula model in any dimension, it becomes nearly impossible to compute it because of the high-dimensional multivariate t-distributions involved. We resort to looking only at the pairwise two-dimensional margins of the data. This is essentially a generalisation of what we do with a t-copula. When fitting a t-copula to data, we compute the correlation matrix using Kendall's tau. This only takes into account the pairwise dependence. Then we estimate the degrees of freedom ν using ML estimation.

The process of fitting a t-panic copula to higher-dimensional data is straight-forward. We just use l^* in all cases where we would use the log-likelihood function. But during the evaluation of the fitted model, we lose important key figures like AIC, BIC and everything that relies on the Rosenblatt transform, because we want to avoid computing distribution functions of multivariate t-distributions in high dimensions. Parametric bootstrapping is still possible, but becomes more runtime-intensive. The general behaviour of the statistics M and K varies from dimension to dimension. Therefore, we cannot even give any rough assessment regarding the p-value of a certain statistic without parametric bootstrapping. But then again, the procedure of parametric bootstrapping becomes impractical for the t-panic copula in high dimensions, as the fitting of the model becomes more time-intensive with every added dimension. So we are left to compare the raw values of M and K among the different models and choose the best one.

In Tables 6.7 and 6.8, the data used in the section before is used again

	PW-inhom.	PW-hom.	t-copula
# Parameters	33	19	16
$\hat{q}$	0.1169	0.1163	-
$\hat{\nu}_C$	6	6	6
$\hat{\nu}_P$	11	12	-
l^*	**5799.5**	5738.0	5667.0
M	**0.0804**	0.0880	0.2291
K	**1.0151**	1.0562	1.6760

Table 6.8: Fitted 6-dimensional t-panic copulas in comparison with the t-copula model, using the same daily stock returns from Tables 6.2 and 6.6 with a sample size $N = 4105$. The best value in each line is printed bold.

to fit 6-dimensional t-panic copulas. In general, the only three additional parameters in the model "PW-hom." can already make a big difference compared to the t-copula. It is apparent that in high dimensions, it is not reasonable to employ the t-panic copula with inhomogeneous panic correlation, as the amount of parameters basically doubles without a large increase in l^*. Therefore, a homogeneous matrix $\Sigma^{(P)}$ is advisable. But we might think that assets of different industry sectors should not have the same panic correlation that closely related assets have. The stock price of a bank will react rather strongly in case of an extreme drop in the stock price of another bank company. A loss of an industrial company, on the other hand, will have less influence on the stock price of the bank. Therefore, we can propose the following approach.

We can group the assets into business sectors and restricting the correlation matrix in the following way: each group G_k has homogeneous correlation ρ^G_{kk} among their members, each two assets in different groups G_k, G_l have a correlation ρ^G_{kl} depending only on their group. That is,

$$\Sigma^{(P)}_{ij} = \rho^G_{kl}, \text{ if } i \in G_k, j \in G_l,$$

where ρ^G is a symmetric matrix containing the correlations of the group members. This reduces the number of additional parameters stemming from the panic correlation to a quadratic function in the number of groups. If there are k groups, the number of parameters in this model is then $\binom{n}{2} + \binom{k+1}{2} + 3$. This is just one possibility to restrict the model into having less parameters while preserving the inherent structure of the empirical data. This borrows the basic idea from the factor copula models. Further details on factor copula models can be found in [29].

This construction brings us to a new model fitting scheme that we call "PW-grouped", that comprises both the inhomogeneous and the homogeneous model as special cases. In fact, they are nested models with the

	PW-inhom.	PW-hom.	PW-grouped	t-copula
# Parameters	45	25	30	22
$\hat{q}$	0.2039	0.0006	0.2039	-
$\hat{\nu}_C$	5	4	5	5
$\hat{\nu}_P$	5	5	5	-
l^*	**12377.7**	12315.8	12334.6	12212.8
M	**0.0490**	0.3418	0.0631	0.2419
K	1.1442	2.6844	**1.0883**	1.9574

Table 6.9: Fitted 7-dimensional t-panic copulas in comparison with the t-copula model, using the daily stock returns from 7 German companies with a sample size $N = 4166$. The best value in each line is printed bold.

inhomogeneous one being the most general and the homogeneous one being the most restrictive.

Let us look at daily stock return data from 7 German companies: Volkswagen, BMW, Daimler, Commerzbank, Deutsche Bank, Merck and Fresenius. We make 3 groups: car companies, banks and medical companies. We then compare the fitted models. The results can be seen in Table 6.9. It is apparent that the small number of additional parameters makes a big difference to the model. The model "'PW-hom." does not seem to be an adequate model since the panic threshold virtually converges to 0. In the grouped model, we have a correlation of 0.8440 in the car sector, 0.7655 between car companies and banks and 0.8759 in the bank sector, and so on. This can be seen in the matrix $\hat{\rho}^G$:

$$\hat{\rho}^G = \begin{pmatrix} 0.8440 & 0.7655 & 0.5151 \\ 0.7655 & 0.8759 & 0.4889 \\ 0.5151 & 0.4889 & 0.4780 \end{pmatrix}$$

From the statistics M and K we can conclude that the model "PW-group." is indeed a reasonably good fit, with 15 parameters less than the "full" t-panic copula model.

For an example on what can be done in a reasonable time frame with the t-panic copula model, we take $N = 3896$ daily returns from 17 German companies listed in the DAX stock market index. As before, the data is harvested using Yahoo Finance.

The assets are partitioned into 10 groups. As can be seen in Table 6.10, the 135 additional parameters of the inhomogeneous t-panic copula model in comparison to the homogeneous one yield an improvement of the function l^* that is comparable to the improvement obtained by the 3 additional parameters of homogeneous t-panic copula model in comparison to the t-copula. The model "'PW-group." provides a good fit and the increase in

	PW-inhom.	PW-hom.	PW-group.*	PW-group.	t-copula
# Param.	275	140	194	194	137
$\hat{q}$	0.1321	0.1056	0.1286	0.1254	-
$\hat{\nu}_C$	5	5	5	5	7
$\hat{\nu}_P$	7	9	5	7	-
l^*	**74054.3**	73330.6	73695.0	73884.1	72318.8
M	**0.0763**	0.0908	0.1263	0.0815	0.2796
K	2.4857	**2.1237**	2.3921	2.1424	3.1695

Table 6.10: Fitted 17-dimensional t-panic copulas in comparison with the t-copula model, using the daily stock returns from 17 German companies with a sample size $N = 3896$. The best value in each line is printed bold.

l^* is moderate. Judging from the statistics M and K, the grouped model has no significant advantages over the homogeneous model, in spite of the large number of groups. "PW-group*" has the additional constraint that $\hat{\nu}_P \leq \hat{\nu}_C$. It has the worst fit of the four t-panic copulas, but it preserves the wanted property that the lower tail dependence be higher than the upper tail dependence, even though the values tend to be too high when compared to the empirical estimates.

We conclude that all of the four t-panic copula models provide a significantly better fit than the t-copula. The homogeneous model is very minimalistic in the number of additional parameters, the grouped model has more parameters and a comparable fit. The model "PW-inhom." does not provide a good enough fit considering the very high number of parameters. In this scope, "PW-hom." provides a good enough fit.

Using the Kolmogorov-Smirnov statistic, we have seen that the hypothesis of a location scaled t-distribution for the margins holds, with p-values typically over 0.5 and the lowest one being 0.22. We compute the VaR and ES for an approximately equally weighted portfolio of the 17 stocks with initial worth of €3089.8. The results can be seen in Table 6.11. We can see that the model "PW-hom." matches the empirical VaR rather well for confidence levels smaller than 0.995. While the results for the VaR are comparable with those of the t-copula model, the t-copula doesn't match the empirical estimates of the VaR quite as good. The ES estimated by the t-panic copula models is significantly larger than the empirical one even on the level 0.99. The t-copula provides slightly closer values. The Gaussian distribution that is on the list for comparison completely fails to fit the empirical estimates. Note that the model "PW-group.*" has significantly higher values for both the VaR and ES because of the higher lower tail dependence. We have seen that the theoretical values of λ_L in this restricted model match the empirical ones worse than in the other models, though. They overestimate the lower tail dependence.

VaR_α	**0.95**	**0.98**	**0.99**	**0.995**	**0.999**	**0.9995**	**0.9999**
Empiric	76.01	102.39	135.97	163.22	223.70	*235.09*	*238.25*
Gaussian	80.01	100.00	112.62	123.58	146.94	156.81	178.21
PW-inhom.	72.29	104.42	134.08	169.81	274.25	341.20	486.14
PW-hom.	72.50	105.87	134.01	166.04	263.60	316.60	466.80
PW-group.*	71.48	105.87	136.75	174.01	302.89	360.33	571.16
PW-group.	72.49	105.47	135.75	172.23	278.04	335.34	504.86
t-copula	70.59	101.18	129.04	159.57	246.50	302.20	510.96

ES_α	**0.95**	**0.98**	**0.99**	**0.995**	**0.999**	**0.9995**	**0.9999**
Empiric	109.07	142.68	169.35	194.00	234.38	*237.11*	-
Gaussian	99.78	116.63	127.63	137.80	159.97	168.76	187.38
PW-inhom.	113.65	155.51	193.82	238.30	373.09	443.39	637.05
PW-hom.	113.32	153.78	189.56	231.75	364.29	441.76	685.17
PW-group.*	115.69	161.32	203.34	254.03	416.77	504.95	782.49
PW-group.	114.63	157.16	196.10	241.15	375.90	448.47	667.32
t-copula	109.35	148.50	183.47	224.71	357.93	441.07	780.15

Table 6.11: VaR and ES of a portfolio with initial value of €3089.8 on different confidence levels for the copula models from Table 6.10 and location scaled t-distributed margins. For comparison, the empirical values and the corresponding values assuming a multivariate Gaussian distribution of X are shown. The values are calculated by simulating $100,000$ cases.

It is also worth noting that the VaR and ES of the panic copula models can be higher or lower than the empirical estimates. But as has already been said, the data on which the estimated ES is grounded is very scarce. Therefore, the empirical values are not reliable for very high quantiles.

In theory, there is no limit to the number of assets that can be used in this t-panic copula model. In practice, the runtime is $\Omega(d^2)$ and fitting the model on the discussed 17-dimensional data is already a matter of minutes. Therefore, the applicability of the fitting procedure is questionable in dimensions larger than 25.

6.5 Conclusion

In this chapter, we explored the practical ramifications of the panic copulas introduced before. We mainly concentrated on the most promising model, the t-panic copula. It addresses the issue of specifically modelling the downside of the copula.

The construction of the t-panic copula is based on a simple principle. We constructed the random vector X using two independent Student-t dis-

tributed random vectors Y and Z. If a component of Z reaches below a panic threshold, its value is used. Otherwise, we use the corresponding value of Y. This random vector X is very simple to simulate. We then use the separation step (Algorithm 1) of the CM-algorithm introduced by Meucci in [35] to extract the copula of X. This copula is called the t-panic copula. Its main advantage over the t-copula is the possibility of modelling the lower part of the copula differently than the upper part. The panic correlation matrix $\Sigma^{(P)}$ plays a central role among the additional parameters. It represents the correlation of the components near the lower tail. Archimedean panic copulas are constructed similarly by using Archimedean copulas instead of the Student-t distribution in the process.

The panic copula models we have developed provide a good alternative to the widespread static copula models. In small dimensions, the Archimedean panic copula can provide additional flexibility compared to the Archimedean copulas while still being very easy to compute. In medium-sized dimensions, the t-panic copula can improve the results attainable by the t-copula with a small number of additional parameters. The panic copulas are non-symmetric, which is a big advantage over the t-copula. The panic copulas are easily simulated from by using the CM-algorithm.

While the simulation is very straightforward, the major difficulty lies in the fitting procedure. We apply the semiparametric CML estimation for the parameters of the copula, i.e. the margins are estimated non-parametrically. The likelihood function of the panic copulas exhibits discontinuities with respect to one parameter, i.e. the panic threshold. Therefore, the maximisation of the likelihood function during the CML estimation is cumbersome. We use a hill-climbing routine, which by its nature only finds local optima. It is tedious to verify whether the local optimum is a global one. After the estimation, most computations on the fitted panic copula can easily be done using Monte Carlo simulation. The evaluation of the goodness of fit for any model can be carried out using approximate p-values for the distribution hypothesis of a certain copula model if we employ the parametric bootstrapping algorithm (Algorithm 3).

After evaluating the newly developed models on real market data by using various goodness-of-fit measures, we come to the following conclusions: Archimedean panic copulas provide an interesting alternative to other copulas in low dimensions. Because of their low number of parameters, they do not fit real market data as well as the t-panic copula does. In dimensions $d > 3$ they completely fail to keep up with other models. Their fitting procedure is unproblematic, albeit a lot more runtime-intensive than fitting a plain Archimedean copula.

The t-panic copula fits real data significantly better than the t-copula in many cases. This is to be expected because the t-copula is a special case of the t-panic copula model. We can model asymmetrical dependence well and achieve different values for the upper and the lower tail dependence.

Compared to the t-copula, we can also reach a significantly larger part of the τ-ρ_S-region. However, the difference between the models only becomes really apparent with larger sample sizes of at least $N > 600$. Fitting a panic copula to samples that are smaller than that is not advisable, as can be seen from Table 6.1. For such cases the t-copula suffices as it is the simpler model which should be preferred when the advantages of the more complex model are not evident.

The t-panic copula model can be applied to medium sized dimensions by employing a modified likelihood function that only incorporates pairwise dependence between the components. In practice, we applied it to up to 17 assets. The number of parameters for models of such data can become excessively high, including up to twice the number of parameters of a t-copula. To counteract this, we can make use of the idea of factor copula models. By partitioning the assets into groups with the same panic correlation, we limit the amount of additional parameters compared to the t-copula. With only few additional parameters, the goodness of fit is improved considerably with respect to the t-copula.

When fitting the panic copula to real data, it becomes apparent that we actually fail to implement the very idea of finding a "panic" distribution to model *only* the lower tail of the copula. Strictly speaking, what we achieve in our fitted models is something slightly different. The estimated values of the panic threshold quantile $\hat{q}$ are often higher than 0.1. As mentioned in Section 5.2, this means that not only the lower tail is influenced by the panic distribution Z, but basically the whole lowest fifth of the copula is modelled by it. This is of course contrary to the very name of the *panic* copulas because an event at quantile 0.19 cannot be described as panic in any sense. This makes the name misleading in general applications. However, the copula models can still be used for the specific task of stress testing by restricting the panic threshold q to be a lower quantile. But this restriction cannot be justified during the estimation process. Furthermore, because the estimation of the model relies on observing rare lower tail events, estimation of the panic correlation $\Sigma^{(P)}$ becomes less reliable if $\hat{q}$ is very small, as we have already discussed in the previous chapter.

After fitting a parametric model to the marginal distributions, we can use the copula models to calculate the Value-at-Risk and expected shortfall of a portfolio. To do so, we apply Monte Carlo simulation. We found the values we calculated by using the t-panic copula models to be comparable with the empirical estimates. Compared with the t-copula, the estimated values for the VaR and ES in the t-panic copula models tend to be slightly higher, which conforms with the idea we had for the construction. We wanted the high correlation in the lower part of the distribution to be more accurately reflected in the copula model. We cannot say the values obtained with the t-panic copula model fit the empirical estimates better in all cases. Depending on the data, they might also be farther away than the t-copula

values.

This result puts us into a dilemma because the resulting similarly well-fitting VaR and ES figures conflict with the significantly better fit of the copula model as evaluated by established measures like the Cramér-von Mises and Kolmogorov-Smirnov statistics. In the three-dimensional case, the hypothesis of a t-copula can be discarded on virtually any confidence level α for most asset sets. But the hypothesis of a t-panic copula often holds on a confidence level of $\alpha = 0.95$. Yet, the improved fit of the copula itself does not translate into better approximations of the empirical estimates of VaR and ES. We do know, however, that the empirical estimates of VaR and ES should be taken with a grain of salt, as especially the empirical ES is known to be unstable in the ambit of insufficient sample sizes.

The central idea of this thesis is applying the very flexible CM-algorithm to construct new copulas in an unconventional way. This thought can be continued further. With the means provided by the CM-algorithm, the development of new copulas with interesting and different structures has become possible. The copula of compound distributions becomes readily available. It is thinkable to construct a compound distribution where the variance of a multivariate normal distribution is determined by another positive random variable, say an exponential distribution. The copula of this distribution can then be easily separated by employing the CM-algorithm. Other mixed or compound distributions that might carry valuable properties immediately come to mind. The only problem with these kinds of constructions is that there is no straightforward way of estimating the parameters and every model has to be looked at separately. The copula transformation mentioned in Section 4.2.4 is also a promising approach that could be examined further. Given suitable transformation functions, the results should be very useful in modifying existing copula models. Finding such functions is difficult, though. With the tools presented in this thesis, that is the CM-algorithm and the parametric bootstrapping algorithm, new copula models can easily be developed and evaluated.

Bibliography

[1] Carol Alexander. Principal component models for generating large garch covariance matrices. *Economic Notes*, 31(2):337–359, 2002.

[2] Felipe M. Aparicio and Javier Estrada. Empirical distributions of stock returns: scandinavian securities markets, 1990-95. 1996.

[3] Eric Bouyé, Valdo Durrleman, Ashkan Nikeghbali, Gaël Riboulet, and Thierry Roncalli. Copulas for finance - a reading guide and some applications. *Available at SSRN 1032533*, 2000.

[4] Kenneth P. Burnham and David R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[5] Cyril Caillault and Dominique Guegan. Empirical estimation of tail dependence using copulas: application to asian markets. *Quantitative Finance*, 5(5):489–501, 2005.

[6] George Casella and Roger L. Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

[7] Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula Methods in Finance*. The Wiley Finance Series. Wiley, 2004.

[8] Barbara Choroś, Rustam Ibragimov, and Elena Permiakova. Copula estimation. In *Copula theory and its applications*, pages 77–91. Springer, 2010.

[9] Paul Deheuvels. Non parametric tests of independence. In *Statistique non paramétrique asymptotique*, pages 95–107. Springer, 1980.

[10] Jadran Dobrić and Friedrich Schmid. A goodness of fit test for copulas based on rosenblatt's transformation. *Computational Statistics & Data Analysis*, 51(9):4633–4642, 2007.

[11] Fabrizio Durante, Rachele Foschi, and Peter Sarkoci. Distorted copulas: constructions and tail dependence. *Communications in Statistics—Theory and Methods*, 39(12):2288–2301, 2010.

[12] Valdo Durrleman, Ashkan Nikeghbali, and Thierry Roncalli. A simple transformation of copulas. *Available at SSRN 1032543*, 2000.

[13] Paul Embrechts, Rüdiger Frey, and Alexander McNeil. Quantitative risk management. *Princeton Series in Finance, Princeton*, 10, 2005.

[14] Paul Embrechts, Filip Lindskog, and Alexander McNeil. Modelling dependence with copulas. 2001.

[15] Elzbieta Ferenstein and Mirosław Gasowski. Modelling stock returns with ar-garch processes. *SORT*, 28(1):55–68, 2004.

[16] Jean-David Fermanian, Dragan Radulovic, Marten Wegkamp, et al. Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860, 2004.

[17] Gabriel Frahm, Markus Junker, and Alexander Szimayer. Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286, 2003.

[18] Maurice J. Frank. On the simultaneous associativity of f(x,y) and x+y-f(x,y). *Aequationes mathematicae*, 19(1):194–226, 1979.

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

[20] Christian Genest and Johanna Nešlehová. A primer on copulas for count data. *Astin Bulletin*, 37(02):475–515, 2007.

[21] Christian Genest, Jean-François Quessy, and Bruno Rémillard. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366, 2006.

[22] Christian Genest and Bruno Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 1096–1127, 2008.

[23] Christian Genest, Bruno Rémillard, and David Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213, 2009.

[24] Nicholas J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA journal of Numerical Analysis*, 22(3):329–343, 2002.

[25] Harry Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

[26] Sam Jones. The formula that felled wall st. http://www.ft.com/cms/s/0/912d85e8-2d75-11de-9eba-00144feabdc0.html, 2009.

[27] Clark H. Kimberling. A probabilistic interpretation of complete monotonicity. *Aequationes mathematicae*, 10(2):152–164, 1974.

[28] Stanley J. Kon. Models of stock returns—a comparison. *The Journal of Finance*, 39(1):147–165, 1984.

[29] Pavel Krupskii and Harry Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101, 2013.

[30] H.O. Lancaster. Correlation and complete dependence of random variables. *The Annals of Mathematical Statistics*, 34(4):1315–1321, 1963.

[31] Susan Lee. Formula from hell. http://www.forbes.com/2009/05/07/gaussian-copula-david-x-li-opinions-columnists-risk-debt.html, 2009.

[32] Erich Leo Lehmann, George Casella, and George Casella. *Theory of point estimation*. Wadsworth & Brooks/Cole Advanced Books & Software, 1991.

[33] David X. Li. On default correlation: A copula function approach. *Available at SSRN 187289*, 1999.

[34] C. H. Ling. Representation of associative functions. 1965.

[35] Attilio Meucci. A new breed of copulas for risk and portfolio management. *Risk*, 24(9):110, 2011.

[36] Patricia Mariela Morillas. A method to obtain new copulas from a given one. *Metrika*, 61(2):169–184, 2005.

[37] Roger B. Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 1999.

[38] Murray Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Statist.*, 23(3):470–472, 09 1952.

[39] Michael Roth. On the multivariate t distribution. Technical report, Tech. Rep., Division of automatic control, Linköping university, 2013.

[40] Peter J. Rousseeuw and Geert Molenberghs. Transformation of non positive semidefinite correlation matrices. *Communications in Statistics–Theory and Methods*, 22(4):965–984, 1993.

[41] Felix Salmon. Recipe for disaster: the formula that killed wall street. *Wired*, 17(3), 2009.

[42] Friedrich Schmid, Rafael Schmidt, Thomas Blumentritt, Sandra Gaißer, and Martin Ruppert. Copula-based measures of multivariate association. In *Copula theory and its applications*, pages 209–236. Springer, 2010.

[43] Manuela Schreyer, Roland Paulin, and Wolfgang Trutschnig. On the exact region determined by kendall's tau and spearman's rho. *arXiv preprint arXiv:1502.04620*, 2015.

[44] Robert J. Serfling. *Approximation Theorems of Mathematical Statistics (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 1980.

[45] Carlo Tomasi. Estimating gaussian mixture densities with em – a tutorial. *Duke University*, 2004.

[46] Emiliano A. Valdez and Yugu Xiao. On the distortion of a copula and its margins. *Scandinavian Actuarial Journal*, 2011(4):292–317, 2011.

[47] Aad W. Van Der Vaart and Jon A. Wellner. *Weak Convergence*. Springer, 1996.

[48] Mark Whitehouse. How a formula ignited market that burned some big investors. `http://www.ft.com/cms/s/0/912d85e8-2d75-11de-9eba-00144feabdc0.html`, 2005.