Characterizations of Multivariate Tail Dependence

On theory and inference to assess extremal dependence structures

by

Carina van der Zee

in partial fulfillment of the requirements for the degree of

Master of Science
in Applied Mathematics,
with specialization Probability and Statistics,

at the Delft University of Technology,
Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS),
to be defended publicly on Wednesday July 25, 2018 at 2:00 PM.

Student number:	4150325	
Thesis committee:	Dr. D. Kurowicka,	TU Delft (supervisor)
	Dr. J.J. Cai,	TU Delft (supervisor)
	Dr. P. Cirillo,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft Delft University of Technology

Acknowledgments

First and foremost, I would like to thank Dorota Kurowicka and Juan-Juan Cai for supervising this thesis project during the past nine months and for introducing me to the world of copulas and extreme value theory. Both of you have given me the tools and freedom to pursue my interests and provided me with valuable directions and feedback when needed. I have enjoyed discussing theory and statistics during our meetings, and I am grateful for everything I have learned about multivariate tail dependence.

I am also grateful for Pasquale Cirillo, being a member of my thesis committee, for giving inspiring lectures during the minor finance. Not long after my encounter with tail risk and extreme events in risk management, I decided to switch from studying Architecture to Mathematics. As this transition was quite unusual, I am very grateful to Martin van Gijzen for granting me the opportunity to take this step. I have to credit Bas Borgdorff for helping me take this decision and for supporting me during the long hours I needed to complete my studies and, in particular, this thesis. I also want to thank my parents for their continuous support during my various studies both in Rotterdam and Delft.

Besides researching multivariate tail dependence, I've enjoyed working at the FX Derivatives desk of the ING bank during the past year. In particular, I would like to thank Tijmen van Paasen and Alexander Schreuder Goedheijt for a fantastic time. You have taught me different aspects of banking as well as many, many, other things, and I enjoyed the opportunity to apply my knowledge of statistics to practical problems.

Last but not least, I have to thank my father, Rien van der Zee, Bas Borgdorff and Iris Koks for meticulously proofreading this thesis.

Carina van der Zee
Delft, July 2018

Abstract

This thesis gathers, develops and evaluates several characterizations of multivariate tail dependence. It is established that the stable tail dependence function (STDF) is a suitable copula-based dependence function that fully captures the multivariate extremal dependence structure in all dimensions $d \geq 2$ and can be used to visualize the tail dependence structure for bivariate and trivariate problems. Based on the STDF, we propose a multivariate tail dependence coefficient (TDC) as an extension of the well-known bivariate TDC. Importantly, we show that the proposed measure can identify tail independence in all dimensions $d \geq 2$, similar to its bivariate variant. The performance of nonparametric estimators for the STDF and, inherently, the multivariate TDC, is assessed with an extensive simulation study, including smoothed and bias-corrected versions of the empirical STDF. Based on the estimators for the STDF and the multivariate TDC, test statistics under the null hypothesis of tail independence are developed and evaluated in another simulation study. The STDF-based estimation and testing procedures are applied to foreign exchange (FX) data to characterize the tail dependence structure between three European FX rates and five worldwide FX rates.

Keywords: extreme value copula, stable tail dependence function, tail dependence coefficient, tail dependence testing

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Extreme events are by definition very rare but have a substantial impact when they do occur, especially if they occur together. The financial crisis of 2008 has become a textbook example of such a doom scenario: financial markets all suffered massive losses, and major financial institutions collapsed (or received financial aid to prevent such a collapse). Also during less apocalyptic times, adverse shocks to financial asset returns tend to co-occur (Longin and Solnik, 2001). Increasing globalization and regulation are both forces that contribute to such interlinkages amongst financial institutions and financial markets, thus increasing the likelihood of joint downturns. Because these joint tail events may have such a high impact on a company, an investor, or the economy as a whole, the dependence of large values in a stochastic process is an essential topic in risk management, the insurance industry, and general finance, and will be the topic of this thesis.

1.1 The impact of tail dependence

To understand the impact of tail dependence, we consider an example. Let $\mathbf{X} = (X_1, \dots, X_d)$ represent $d \geq 2$ risk factors to which an entity is exposed. The specification of the risk factors can vary by application. In credit risk, the risk factors can be represented by the yields on bonds that may downgrade or default; a professional investor might consider them to be the return series for the assets in his portfolio; and for a multinational company, the risk factors can be represented by foreign exchange (FX) rates that affect its profit and loss statement. The aggregate loss associated with the risk factors is given by the sum of its marginal components, $S_d = X_1 + \cdots + X_d$, and is often used to determine the aggregate risk associated with the portfolio. An example where this mathematical framework is applied is the so-called Internal Capital Adequacy Assessment Process (ICAAP), in which banks are required to aggregate all relevant risk types in a top-level joint portfolio of losses (BCBS, 2010). This assessment is, of course, also critical to consider for companies or investors that do not have to comply with these regulations.

The distribution of the aggregate losses S_d depends on the margins of its components, denoted by $F_1, \dots, F_d$, and on the dependence structure between the components. The joint distribution F of $(X_1, \dots, X_d)$ captures both of these factors fully and is defined as follows,

$$F(x_1, \dots, x_d) = \mathbb{P}\left(X_1 \leq x_1, \dots, X_d \leq x_d\right).$$

The typical situation is that banks or insurance companies possess information and models concerning the standalone risks rather than the joint behavior of the risks (Puccetti et al., 2017). It might, therefore, be desirable to model the marginal behavior and the dependence structure separately, which is pre-eminently done with copulas. A copula is a dependence function on the d-dimensional unit cube that fully specifies the dependence structure between the random vector $(X_1, \dots, X_d)$. The seminal theorem of Sklar (1959) shows that the joint distribution F can be decomposed into the copula function C_F and the marginal distributions $F_1, \dots, F_d$ as follows,

$$F(x_1, \dots, x_d) = C_F\left(F_1(x_1), \dots, F_d(x_d)\right).$$

More information on copulas can be found in Chapter 2. Different specifications of the dependence structure between risk factors can lead to very different conclusions on the amount of risk that is induced by the given risk factors. In particular, the strength of *tail* dependence affects the *tail*

risk that is incurred. Tail dependence concerns the degree to which two or more random variables move together in the tails of their distributions and captures the degree to which extreme values of two or more random variables are likely to occur simultaneously. When two random variables are tail dependent, extreme realizations can be observed together and are dependent on each other. If two random variables are tail independent, it is nearly impossible to observe extreme values simultaneously due to the unlikely nature of extreme values and their independent occurrences. The tail dependence structure is not necessarily the same as the general dependence structure. It can therefore be instructive to consider the tail dependence structure separately from the general dependence structure.

The influence of tail dependence on tail risk is quantifiable by its impact on tail risk measures, such as the Value-at-Risk (VaR) and the Expected Shortfall (ES) of the aggregate loss. The VaR is defined as a quantile of the loss distribution over a given timeframe for a given significance level $\alpha \in (0, 1)$,

$$\mathrm{VaR}_\alpha(S_d) = \inf\left\{x \in \mathbb{R} : \mathbb{P}\left(S_d \leq x\right) > \alpha\right\},$$

and the ES is defined as the conditional expected loss given that the loss is larger than the VaR,

$$ES_\alpha(S_d) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\zeta(S_d)\mathrm{d}\zeta.$$

Usual values of the significance level include $\alpha = 0.95, 0.99, 0.995, 0.999$. To illustrate the effect of tail dependence on these tail risk measures, $N = 2500$ observations are simulated for two variables with standard t-margins with 5 degrees of freedom, giving the margins heavy tails. Either strong upper tail dependence with a Gumbel copula ($\theta = 1.54$) or tail independence with a Frank copula ($\theta = 3.45$) is imposed. The parameters of both copulas are chosen as that they yield the same correlation of $\rho = 0.5$, by which the comparison is between tail dependencies, not dependence in general. The simulation results are shown in Figure 1.1.1. Positive values correspond to losses.

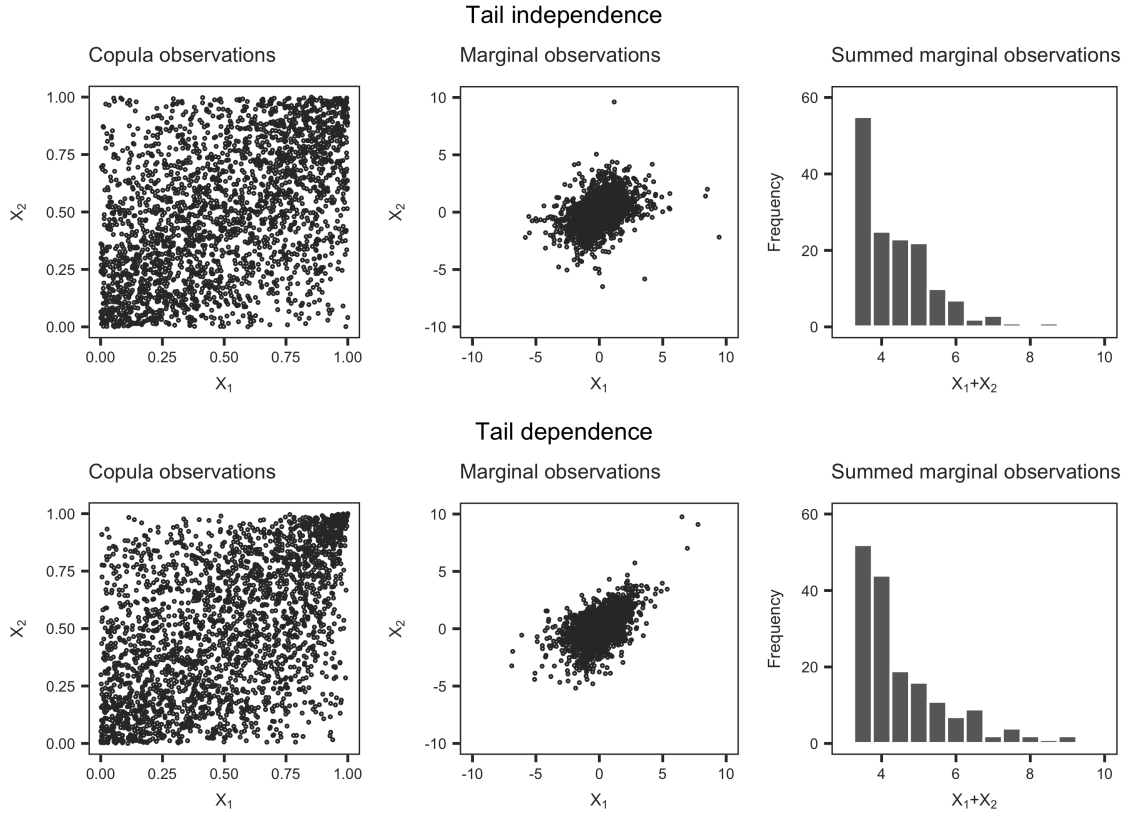

Figure 1.1.1: Simulation results for $N = 2500$ realizations from a distribution with standard student t margins with 5 degrees of freedom and a Frank copula with parameter $\theta = 3.45$ for the tail independent case (top) and a Gumbel copula with parameter $\theta = 1.54$ for the tail dependent case (bottom).

From the simulated marginal observations, it is clear that in case of tail dependence, extremely high values are more likely to occur simultaneously than in case of tail independence. For copula observations this can be observed by the cluster of observations in the upper right corner for tail dependence. The histograms presented in Figure 1.1.1 focus on the upper right tail of the distribution of the sums, i.e., $S_2 = X_1 + X_2$ in this case. It is visible that the tail of the distribution of the sum for tail dependence is fatter than for tail independence. That it, it contains more observations. To quantify these findings, tail risk measures for these two distributions are determined based on $N = 100,000$ simulated observations; the number of simulations is increased in order to get a more accurate estimate. The 99% VaR for the tail independent case equals 5.34 versus 6.24 for the tail dependent case. The difference between the ES estimations is even more pronounced: the 99% ES is estimated to be 6.72 for tail independence versus 8.33 for tail dependence. Of course, it depends on the scale of these numbers whether these differences are economically significant. Since exposures for large financial institutions are substantial, the scale will typically be in millions or even billions. The difference matters. This effect becomes even stronger when we repeat this simulation experiment for higher dimensions, see Table 1.1.1.

Table 1.1.1: Tail risk measures for tail independent and tail dependent data.

		$d = 2$	$d = 3$	$d = 5$	$d = 7$	$d = 10$
99% VaR	*Tail independent*	5.34	7.33	11.15	15.00	20.72
	Tail dependent	6.25	9.17	15.15	20.86	29.68
99% ES	*Tail independent*	6.72	8.88	13.15	17.28	23.57
	Tail dependent	8.33	12.23	20.11	27.85	39.73

The tail risk measures are computed based on $N = 100,000$ simulated observations from a distribution with standard student t margins with 5 degrees of freedom and a Gumbel copula with parameter $\theta = 1.54$ for the tail dependent case and a Frank copula with parameter $\theta = 3.45$ for the tail independent case for dimensions $d \in \{2, 3, 5, 7, 10\}$.

The fact that tail dependence between risk factors is essential to consider when an aggregate risk exposure has to be assessed has been widely recognized since the past financial crisis. The question arises how information on tail dependence should be incorporated in the aggregate risk analysis.

1.2 Problem statement

Recently, it has been investigated how to take possible dependence structures between individual risk factors into account when assessing tail risk. An approach introduced by Cont (2006) is to consider the best- and worst-case scenarios over the set of all possible dependence structures between the individual risk factors. Usually, this leads to bounds on risk measures that entail too much uncertainty, making them unattractive to use, both for financial institutions as well as regulators. See, e.g., Puccetti and Rüschendorf (2012). More information on the tail dependence is required to improve the accuracy of the bounds. For example, information on whether particular groups of risk factors are tail dependent or tail independent already strongly reduces the possible dependence structures (Puccetti et al., 2017). Additional information on how strong the tail dependence is, or, even better, information on the entire structure of the multivariate tails, might be able to subsequently further reduce uncertainty.

The evaluation of the multivariate tail dependence structure can, however, be a daunting exercise. A scarcity of data in the tails and the need to extrapolate stochastic behavior beyond historically observed extremes are factors that already complicate inference on the univariate tails of a distribution. These complications become worse for bivariate tails and even more problematic for higher dimensions $d \geq 3$. Unfortunately, it was concluded that the importance of an accurate characterization of the tail dependence structure grows in higher dimensions (see Table 1.1.1). Therefore, this study focuses on gathering, comparing, and, when possible, extending characterizations of high-dimensional tail dependencies. Consequently, the focus is on the following general research question:

How can multivariate tail dependence be characterized?

1.2.1 Main findings

To answer the above question, the aim is to collect, connect and complement theoretical characterizations of multivariate tail dependence and to assess or, where necessary, develop statistical methods to estimate those characterizations of multivariate tail dependence that are best suited for practical use. Three types of characterizations of multivariate tail dependence are distinguished: (1) the full tail dependence structure; (2) a summary measure of the strength of the tail dependence; and (3) a classification into either tail dependence or tail independence. These three types of characterizations contain a decreasing amount of information on the tail dependence structure but can all be used to gain insights into the tail dependence structure of multivariate data and to reduce uncertainty in the bounds of tail risk measures. It depends on the situation and type of analysis to be conducted which characterization is most appropriate to use.

The literature on multivariate extremes presents a rich number of alternatives to fully characterize multivariate tail dependence structures, including, for example, the multivariate extreme value distribution, the spectral measure, and the tail copula. It is argued here that as a full characterization of multivariate tail dependence, the stable tail dependence function (STDF) is best suited to be used in multivariate extreme value analysis. Intuitively, the STDF captures the probability that at least one of the risk factors is extreme. It is based on a limiting value of the copula C_F that describes the global dependence structure between d risk factors and is denoted by ℓ. The STDF can be estimated nonparametrically with the empirical STDF or with smoothed and bias-corrected versions thereof. Based on the STDF a coefficient to summarize multivariate tail dependence is proposed to characterize the strength of high dimensional tail dependence. The coefficient of multivariate tail dependence is a rescaled version of the STDF evaluated in the point $(1, \ldots, 1)$. Estimation procedures for the STDF can therefore be employed to estimate the multivariate TDC $\Lambda(C_F)$. Finally, both the STDF and the multivariate TDC can be used to classify (groups of) risk factors as either tail dependent or tail independent. The vector of risk factors is tail independent if its copula C_F belongs to the maximum domain of attraction of the independence copula Π, which is denoted as $C_F \in MDA(\Pi)$.

1.2.2 Main contributions

In analyzing the different ways to characterize multivariate tail dependence, both theoretically and statistically, several contributions to the existing literature on multivariate tail dependence are added in this thesis. In summary, the main contributions of this thesis are the following.

1. **Multivariate tail dependence coefficient**
   In the bivariate case, the tail dependence coefficient $\lambda(C_F)$ is often used as a summary measure for the strength of tail dependence between two random variables. Based on its relationship with the STDF, we propose a multivariate extension of the coefficient, denoted by $\Lambda(C_F)$, as follows,

   $$\lambda(C_F) = 2 - \ell(1, 1), \qquad \Lambda(C_F) = \frac{d - \ell(1, \ldots, 1)}{d - 1}.$$

   The multivariate tail dependence coefficient Λ satisfies most desirable properties of a tail dependence measure. Most importantly, we show that the multivariate measure Λ can identify tail independence,

   $$\Lambda(C_F) = 0 \Leftrightarrow C_F \in MDA(\Pi).$$

2. **Testing tail independence**
   New procedures to test whether multivariate data are tail dependent or independent are introduced and evaluated. The tests are based on either the STDF or the multivariate TDC, $\Lambda(C_F)$, because both can identify tail independence. Since this issue has not been addressed in the literature before for dimensions $d \geq 3$, we consider it as one of the major contributions to the existing body of knowledge on inference for multivariate extremes.

3. **Estimation of the STDF**
   Since the STDF characterizes multivariate tail dependence fully and it is the corner stone of both the multivariate TDC and the test statistics for testing tail independence, it is important to have a good estimator for the ℓ-function. Therefore, several nonparametric estimators are

discussed and evaluated in an extensive simulation study that is currently lacking in the literature, especially for higher dimensions $d \geq 3$.

1.2.3 Thesis structure

The structure of the remainder of this thesis is as follows. Chapter 2 addresses several preliminaries regarding copula theory. Chapter 3 presents the multivariate extreme value distribution along with an overview of the most important characterizations of its dependence structure. Important properties and results for several known copula distributions are also given. Several tail dependence measures are presented in Chapter 4. Extensions of bivariate dependence measures to higher dimensions are considered and compared. Chapter 5 deals with inference for multivariate extremal dependence. Estimation procedures for the stable tail dependence function and tail dependence summary measures are described and compared with an extensive simulation study. Furthermore, several test statistics to test the null hypothesis of tail independence are developed and evaluated in another simulation study. An application of the characterizations of multivariate tail dependence measures to FX data is considered in Chapter 6. Finally, Chapter 7 presents conclusions and provides directions for further research.

Chapter 2

Copulas

The aim of this thesis is to characterize the multivariate tail dependence structure between $d \geq 2$ random variables. Before introducing results for tail dependence, however, this chapter presents several necessary preliminaries on copulas required later on in the thesis. In particular, the next sections discuss several properties of copulas and introduce two main classes of copulas: elliptical and Archimedean copulas. Although these copulas are defined for higher dimensions, the pair-copula construction for multivariate vine models yields much more flexible models and is therefore also introduced. The chapter concludes with the empirical copula, a nonparametric estimator of the copula, and smoothed versions thereof.

2.1 The joint distribution function

Let the risk portfolio be represented by a vector of continuous random variables $\mathbf{X} = (X_1, \ldots, X_d)$ on a standard non-atomic probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Throughout this thesis, this setup will be employed unless mentioned otherwise. The dependence structure between real-valued random variables can be completely characterized by their joint distribution function, which is defined as follows.

Definition 2.1. *The joint distribution function $F : \mathbb{R}^d \to [0,1]$ of a d-dimensional vector of random variables $\mathbf{X} = (X_1, \ldots, X_d)$ is given by*

$$F(x_1, \ldots, x_d) = \mathbb{P}(\{X_1 \leq x_1\} \cap \cdots \cap \{X_n \leq x_d\}), \quad x_1, \ldots, x_d \in \mathbb{R}.$$

The one-dimensional distribution function $F_j(x) = \mathbb{P}(X_j \leq x)$ of the component X_j, $1 \leq j \leq d$, is called a marginal or margin of the joint distribution function of the random vector $(X_1, \ldots, X_d)$. If the components of the random vector are distributed independently of each other, the vector is said to be independent. In this case, the joint distribution function is the product of the marginals, as shown by the following definition.

Definition 2.2. *The d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is independent if for any intervals $I_1, \ldots, I_d \subseteq \mathbb{R}^d$,*

$$\mathbb{P}\left(\{X_1 \in I_1\} \cap \cdots \cap \{X_d \in I_d\}\right) = \prod_{j=1}^{d} \mathbb{P}(X_j \in I_j).$$

If random variables are not independent, they are dependent. Intuitively, the dependence structure describes the extent to which two or more random variables move together. That is, positive dependence indicates that two random variables move together in the same direction, whereas negative dependence indicates that two random variables move together in opposite directions.

Although the dependence structure between two or more random variables can be described entirely by their joint distribution function F, this joint distribution function also contains information regarding the marginal distributions of the random variables. By the theorem of Sklar (1959), copulas can be used to separate dependence structures from marginal behavior.

Theorem 2.1. *(Sklar, 1959). The d-dimensional continuous random vector* $\mathbf{X} = (X_1, \ldots, X_d)$ *is joint by the unique copula function* $C_F : [0,1]^d \to [0,1]$ *if its joint distribution* F *can be written as*

$$F(x_1, \ldots, x_d) = C_F(F_1(x_1), \ldots, F_d(x_d)), \quad x_1, \ldots, x_d \in \mathbb{R}, \tag{2.1.1}$$

where $F_1, \ldots, F_d$ *are the marginal distributions of* $X_1, \ldots, X_d$.

Since it is often hard to estimate the joint distribution in practice, especially for high dimensions, copulas can simplify the problem by disentangling marginal and joint behavior. Moreover, the separation between margins and the dependence structure can be desirable if separate models on individual risk factors have to be combined in order to get an aggregate risk model, as mentioned before in Chapter 1.

2.1.1 Copula characteristics

As can be deduced from Theorem 2.1, a copula is a d-dimensional distribution function defined on the unit hypercube, $[0,1]^d$, that describes the dependence between a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$. The copula captures the entire dependence structure and is not influenced by the behavior of the marginal distributions. Formally, the copula function is defined as follows.

Definition 2.3. *A function* $C_F : [0,1]^d \to [0,1]$ *is called a copula if there is a random vector* $(U_1, \ldots, U_d)$ *such that each component* $U_j, 1 \leq j \leq d$, *has a uniform distribution on* $[0,1]$ *and*

$$C_F(u_1, \ldots, u_d) = \mathbb{P}\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right), \quad u_1, \ldots, u_d \in [0,1].$$

A comprehensive introduction to copulas can be found in, e.g., Nelsen (2006) and Kurowicka and Cooke (2006). Expositions on using copulas in the context of financial modeling are presented in Cherubini et al. (2004) and Mai and Scherer (2014). From these references, several essential properties of copulas are taken to be discussed. First of all, a copula can also be characterized by the following theorem.

Theorem 2.2. *The function* $C_F : [0,1]^d \to [0,1]$ *is called a d-dimensional copula if and only if the following properties hold:*

1. C_F *is grounded:* $C_F(u_1, \ldots, u_d) = 0$ *if there exists an* $j \in \{1, \ldots, d\}$ *such that* $u_j = 0$.
2. C_F *has normalized margins:* $C_F(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$, *for* $u_j \in [0,1]$, $j \in \{1, \ldots, d\}$.
3. C_F *is d-increasing: for each d-dimensional rectangle* $[a_1, b_1] \times \cdots \times [a_d, b_d] \subseteq [0,1]^d$ *the following holds:*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1 + \cdots + i_n} C_F(x_{1i_1}, \ldots, x_{ni_n}) \geq 0,$$

where $x_{j1} = a_j$ *and* $x_{j2} = b_j$ *for all* $j \in \{1, \ldots, d\}$. *This corresponds to the fact that the probability of* $(U_1, \ldots, U_d) \sim C_F$ *falling in the rectangle in question must be nonnegative.*

By the definition of a copula, the joint density f of random variables $X_1, \ldots, X_d$ can be written in terms of the copula density c and the marginal densities as follows,

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d), \tag{2.1.2}$$

given that the density exists. Here, $F_1, \ldots, F_d$ denote the marginal distribution functions and $f_1, \ldots, f_d$ the marginal densities. Consequently, the copula density can be formulated as follows,

$$c(F_1(x_1), \ldots, F_d(x_d)) = \frac{f(x_1, \ldots, x_d)}{f_1(x_1) \cdots f_d(x_d)}. \tag{2.1.3}$$

Special cases of the copula function are presented by independence and perfect positive and negative dependence. The independence copula is given by $\Pi(u_1, \ldots, u_d) = u_1 \cdots u_d$ and is also called the product copula. It is clear to see that random variables are independent if and only if they are joint by the independence copula,

$$\begin{aligned} F(x_1, \ldots, x_d) &= \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = \mathbb{P}(X_1 \leq x_1) \cdots \mathbb{P}(X_d \leq x_d) \\ &= F_1(x_1) \cdots F_d(x_d) = \Pi(F_1(x_1), \ldots, F_d(x_d)). \end{aligned}$$

For the bivariate copula, the Fréchet-Hoeffding bounds provide a lower and upper bound for the copula function. Specifically, for any bivariate copula C_F, we have $W \leq C_F \leq M$ with

$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0),$$
$$M(u_1, u_2) = \min(u_1, u_2).$$

The copulas M and W describe completely positive and negative dependence, respectively. The upper bound copula is also called the comonotonicity copula and can be easily extended to the multivariate case as $M_d(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d)$. This multivariate upper bound is itself a copula which is also referred to as the comonotonicity copula. A random vector $(U_1, \ldots, U_d)$ has M as joint distribution function if and only if $U_1 = \cdots = U_d$ almost surely. Although the lower Fréchet-Hoeffding bound can be extended to $W_d = \max(u_1 + \cdots + u_d - (d-1), 0)$ for higher dimensions, this is not a proper copula for dimensions $d > 2$. In the bivariate case the lower bound is a proper copula which is referred to as the countermonotonicity copula. A random vector (U_1, U_2) has W as copula if and only if $U_1 = 1 - U_2$ holds with probability one.

Since the copula captures the dependence structure of random variables, it is connected to several dependence measures. For example, the Spearman rank correlation ρ can be written in terms of the bivariate copula C_F as follows,

$$\rho(U_1, U_2) = 12 \int_0^1 \int_0^1 C_F(u_1, u_2) \mathrm{d}u_1 \mathrm{d}u_2 - 3. \tag{2.1.4}$$

This relationship can be used to find copula parameters that correspond to a given Spearman correlation.

Survival copulas

Usually, the focus is on the joint distribution function. However, sometimes the quantity of interest is the survival function, i.e., $\overline{F}(x_1, \ldots, x_d) = \mathbb{P}(X_1 > x_1, \ldots, X_d > x_d)$. The copula of the joint marginal survival functions is called the survival copula and is defined as

$$\overline{F}(x_1, \ldots, x_d) = \hat{C}_F\left(\overline{F}_1(x_1), \ldots, \overline{F}_d(x_d)\right), \quad x_1, \ldots, x_d \in \mathbb{R}, \tag{2.1.5}$$

where $\overline{F}_j$ denotes the survival function of X_j, $1 \leq j \leq d$. Note that the survival copula is linked to the joint survival function of the copula of the joint distribution function as follows:

$$\begin{aligned}
\hat{C}_F(u_1, \ldots, u_d) &= \mathbb{P}\left(\overline{F}_1(x_1) \leq u_1, \ldots, \overline{F}_d(x_d) \leq u_d\right) \\
&= \mathbb{P}(1 - F_1(x_1) \leq u_1, \ldots, 1 - F_d(x_d) \leq u_d) \\
&= \mathbb{P}\left(F_1(x_1) > 1 - u_1, \ldots, F_d(x_d) > 1 - u_d\right) \\
&= \overline{C}_F(1 - u_1, \ldots, 1 - u_d).
\end{aligned}$$

A copula and its survival copula are related through the principle of inclusion and exclusion. That is, the probability of a union of events $A_1, \ldots, A_d \in \mathcal{F}$ can be computed from probabilities involving only intersections of the same events as

$$\mathbb{P}\left(\bigcup_{j=1}^{d} A_j\right) = \sum_{j=1}^{d} (-1)^{j+1} \sum_{1 \leq i_1 \leq \cdots \leq i_j \leq d} \mathbb{P}\left(\bigcap_{k=1}^{j} A_{i_k}\right). \tag{2.1.6}$$

For the bivariate case this simplifies to $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. This principle can be used to express the survival copula $\hat{C}_F$ in terms of the general copula C_F:

$$\hat{C}_F(u_1, \ldots, u_d) = 1 + \sum_{j=1}^{d} (-1)^j \sum_{1 \leq i_1 < \cdots < i_j \leq d} C_{F;i_1,\ldots,i_j}(1 - u_{i_1}, \ldots, 1 - u_{i_j}),$$

where $C_{F;i_1,\ldots,u_j}$ is the j-dimensional marginal copula of C_F which is obtained from C_F by plugging the components $i_1, \ldots, i_j$ of $(u_1, \ldots, u_d) \in [0,1]^d$ into the respective arguments of C_F and setting

all other arguments to one. This can be seen as follows.

$$\begin{aligned}
1-\hat{C}_F(u_1,\ldots,u_d) &= 1-\mathbb{P}\left(U_1>1-u_1,\ldots,U_d>1-u_d\right) = \mathbb{P}\left(\{\exists i\in\{1,\ldots,d\}: U_i\leq 1-u_i\}\right)\\
&= \mathbb{P}\left(\bigcup_{j=1}^{d}\{U_j\leq 1-u_j\}\right) = \sum_{j=1}^{d}(-1)^{j+1}\sum_{1\leq i_1\leq\cdots\leq i_j\leq d}\mathbb{P}\left(\bigcap_{k=1}^{j}\{U_{i_k}\leq 1-u_{i_k}\}\right)\\
&= \sum_{j=1}^{d}(-1)^{j+1}\sum_{1\leq i_1\leq\cdots\leq i_j\leq d}C_{F;i_1,\ldots,i_j}(1-u_{i_1},\ldots,1-u_{i_j})\\
&= \sum_{\emptyset\neq S\subseteq\{1,\ldots,d\}}(-1)^{|S|+1}C_{F;S}(1-u_i, i\in S).
\end{aligned}$$

If a copula is equal to its own survival copula, i.e., $\hat{C}_F = C_F$, the copula is called radial symmetric. This implies that the random vector $(U_1,\ldots,U_d)\sim C_F$ has the same distribution as the random vector $(1-U_1,\ldots,1-U_d)\sim\hat{C}_F$.

2.2 Elliptical copulas

The first class of copulas introduced here are elliptical copulas. Elliptical copulas are copulas of elliptical distributions. Following the exposition of Fang et al. (1992) on the subject, we first define spherical distributions, which are a special type of elliptical distributions.

Definition 2.4. *Let* $\mathbf{X}$ *be a d-dimensional random vector. Then* $\mathbf{X}$ *is said to be spherically distributed if*

$$\mathbf{OX}\overset{d}{=}\mathbf{X}$$

for every orthogonal matrix $\mathbf{O}\in\mathbb{R}^{d\times d}$.

It can be shown that $\mathbf{X}$ belongs to the class of spherical distributions if and only if its characteristic function $\varphi:\mathbb{R}^d\to\mathbb{R}$, defined by $\varphi(\mathbf{t})=\mathbb{E}\left(e^{i\mathbf{t}\mathbf{X}}\right)$, has the form

$$\varphi(\mathbf{t}) = h_\varphi(\mathbf{t}'\mathbf{t}) = h_\varphi\left(t_1^2+\cdots+t_n^2\right),$$

for some function $h_\varphi:\mathbb{R}\to\mathbb{R}$, called the characteristic generator. This implies that if $\mathbf{X}$ has a density $f(\mathbf{x}) = f(x_1,\ldots,x_d)$ then it is equivalent to $f(\mathbf{x}) = g(\mathbf{x}'\mathbf{x}) = g(x_1^2+\cdots+x_d^2)$ for some function $g:\mathbb{R}_+\to\mathbb{R}_+$. Spherical distributions can therefore be interpreted as distributions with density functions that are constant on spheres (Embrechts et al., 1999). The observation that spherical distributions are induced by their characteristic generator h_φ leads to the stochastic representation presented in Theorem 2.3 (Schoenberg, 1938). The proof can be found in Fang et al. (1992).

Theorem 2.3. *Suppose that* $\mathbf{X}$ *is spherically distributed with characteristic generator* h_φ. *Then* $\mathbf{X}$ *has the representation*

$$\mathbf{X}\overset{d}{=}R\mathbf{U},$$

where the random variable $R\geq 0$ *is independent of the d-dimensional random vector* $\mathbf{U}$ *which is uniformly distributed on the unit sphere in* $\mathbb{R}^d$.

This representation will turn out to be very useful in the studying the tail behavior of spherical distributions. The random variable R is also called the generating variable and its distribution function is called the generating distribution. Note that this representation implies that spherical distributions are a mixture of uniform distributions on spheres in $\mathbb{R}^d$ with varying radii. For the standard multivariate Normal distribution, the generating variable is distributed as $R\sim\sqrt{\chi^2_d}$, where χ^2_d denotes the χ^2-distribution with d degrees of freedom. For the multivariate t-distribution with ν degrees of freedom, the generating variable is defined by $R^2/d\sim F(d,\nu)$, where $F(d,\nu)$ denotes the F-distribution with d and ν degrees of freedom (Embrechts et al., 1999).

Spherical distributions are a special class of the more general elliptical distributions. Elliptical distributions can be seen as an extension of the multivariate spherical distribution with mean vector μ and covariance matrix $\mathbf{\Sigma}$. The definition is as follows.

Definition 2.5. *Let* $\mathbf{X}$ *be a* d*-dimensional random vector. Then* $\mathbf{X}$ *is called elliptically distributed with parameters* $\mu \in \mathbb{R}^d$ *and* $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ *if*

$$\mathbf{X} \overset{d}{=} \mu + \mathbf{A}'\mathbf{Y},$$

where $\mathbf{Y}$ *is an* m*-dimensional spherically distributed random vector,* $\mathbf{A} \in \mathbb{R}^{m\times d}$ *with* $\mathbf{A}'\mathbf{A} = \mathbf{\Sigma}$*, and* $rank(\mathbf{\Sigma}) = m$*.*

It can be shown that if $\mathbf{Y}$ has a density $f(\mathbf{y}) = g(\mathbf{y}'\mathbf{y})$ then $\mathbf{X} = \mathbf{A}\mathbf{Y} + \mu$ has density

$$f(\mathbf{x}) = \frac{1}{\sqrt{\det(\mathbf{\Sigma})}} g\left((\mathbf{x}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right).$$

Hence, the contours of equal density are now ellipsoids (Embrechts et al., 1999).

Both the Normal (Gaussian) copula and the t-copula are two examples of elliptical copulas which are used frequently in financial modeling. Historically, the Gaussian copula has been the standard for modeling purposes. The main difference between these two copulas is the fact that the Gaussian copula has independent tails, whereas the t-copula does possess tail dependence. These properties will be discussed extensively in later chapters. Scatterplots of random samples and Normal contour plots of the copulas are shown in Figure 2.2.1 and Figure 2.2.2. One problem of elliptical copulas is that laborious algebraic expressions complicate employment for some applications. Moreover, radial symmetry can present a problem when asymmetric dependence has to be modeled.

Gaussian: Denoting by Φ the univariate standard Normal distribution function, the bivariate Gaussian copula is given by

$$C(u_1,u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right) \mathrm{d}x\mathrm{d}y, \quad u_1,u_2 \in [0,1],$$

with $-1 < \rho < 1$ the correlation coefficient. Perfect positive and negative dependence are achieved as $\rho \to 1$ and $\rho \to -1$, respectively. Exact independence is given by $\rho = 0$. The multivariate extension is given by

$$C(u_1,\ldots,u_d) = \Phi_{\mathbf{\Sigma}}\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)\right), \quad u_1,\ldots,u_d \in [0,1],$$

where $\Phi_{\mathbf{\Sigma}}$ is the multivariate Normal distribution function with zero mean vector and $d \times d$-dimensional covariance matrix $\mathbf{\Sigma}$. Figure 2.2.1 further illustrates the behavior of the Normal copula.

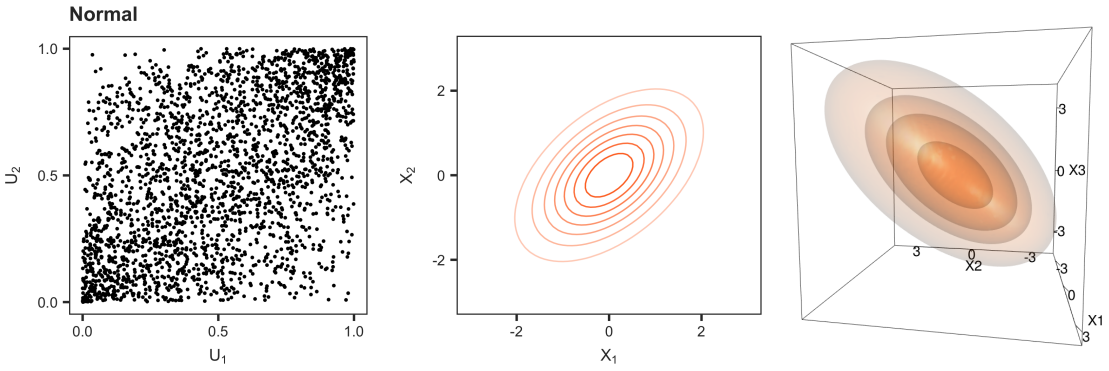

Figure 2.2.1: Scatterplot of ($N = 2500$) random samples and Normal contour plots for the bivariate and trivariate Normal copula. The parameters are chosen such that $\rho = 0.5$ for all bivariate margins.

t: Denoting by T_ν the univariate standard t-distribution function with ν degrees of freedom, the bivariate t-copula is given by

$$C(u_1,u_2)=\int_{-\infty}^{T_\nu^{-1}(u_1)}\int_{-\infty}^{T_\nu^{-1}(u_2)}\frac{1}{\pi\nu|\boldsymbol{\Sigma}|^{1/2}}\frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}}{\nu}\right)^{-\nu/2+1}\mathrm{d}\mathbf{x},\quad u_1,u_2\in[0,1],$$

where $\boldsymbol{\Sigma}$ is the 2×2 covariance matrix. The multivariate extension is given by

$$C(u_1,\ldots,u_d)=T_{\nu,\boldsymbol{\Sigma}}\left(T_\nu^{-1}(u_1),\ldots,T_\nu^{-1}(u_d)\right),\quad u_1,\ldots,u_d\in[0,1],$$

where $T_{\nu,\boldsymbol{\Sigma}}$ denotes the multivariate t-distribution function with ν degrees of freedom, zero mean vector and $d\times d$-dimensional covariance matrix $\boldsymbol{\Sigma}$. Figure 2.2.2 further illustrates the behavior of the t-copula. Relative to the Normal copula, the t-copula has a higher density for joint extreme values.

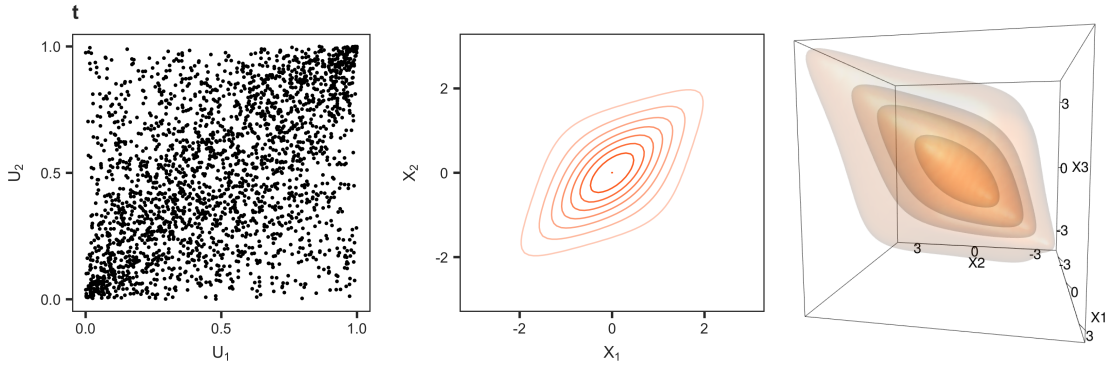

Figure 2.2.2: Scatterplot of ($N=2500$) random samples and Normal contour plots for the bivariate and trivariate t-copula. The parameters are chosen such that $\rho=0.5$ for all bivariate margins.

2.3 Archimedean copulas

Archimedean copulas are the most frequently used alternative to elliptical copulas. Advantages of Archimedean copulas over the elliptical variant are their ability to model asymmetric dependence and their analytic tractability. The class of Archimedean copulas was introduced by Genest and Mackay (1986) and Schweizer and Sklar (1983) as follows.

Definition 2.6. *The copula $C_F(u_1,\ldots,u_d)$ is called a d-dimensional Archimedean copula with generator $\psi:[0,\infty)\to[1,0)$ if it can be represented as,*

$$C_F(u_1,\ldots,u_d)=\psi\left(\psi^{-1}(u_1)+\cdots+\psi^{-1}(u_d)\right),\quad u_1,\ldots,u_d\in[0,1],\tag{2.3.1}$$

and if the generator ψ is a d-monotone, continuous and strictly decreasing function with $\psi(0)=1$ and $\lim_{t\to\infty}\psi(t)=0$.

If $\psi(t)>0$ for all $t\geq 0$ and if $\lim_{t\to\infty}\psi(t)=0$, the generator is said to be strict. In this case, the generalized inverse of the generator coincides with the inverse ψ^{-1}. In this thesis, generators are assumed to be strict generators, unless it is specified otherwise.

A convenient property of the Archimedean copula is that if $(X_1,\ldots,X_d)$ are joint by a d-dimensional Archimedean copula with generator ψ, then any subset of variables of $(X_1,\ldots,X_d)$, are joint by the Archimedean copula with the same generator ψ as the d-dimensional copula. For example, the marginal copula for X_1 and X_2 can be retrieved as follows,

$$\begin{aligned}C_F(u_1,u_2)&=C_F(u_1,u_2,1,\ldots,1)=\psi\left(\psi^{-1}(u_1)+\psi^{-1}(u_2)+\psi^{-1}(1)+\cdots+\psi^{-1}(1)\right)\\&=\psi\left(\psi^{-1}(u_1)+\psi^{-1}(u_2)+0+\cdots+0\right)=\psi\left(\psi^{-1}(u_1)+\psi^{-1}(u_2)\right).\end{aligned}$$

Note the fact that ψ is at least d-monotone implies that $(-1)^j\psi^{(j)} \geq 0$ for $j \in \{1, \ldots, d\}$. This asserts that the density function is nonzero on $(0,1]^d$, if it exists. A special case arises if ψ is infinitely many times differentiable and $(-1)^j\psi^{(j)} \geq 0$ for every $j \geq 1$. In this case ψ is said to be completely monotone and can be represented as a Laplace transform (LT) of a positive random variable,

$$\psi(t) = \int_0^\infty e^{-xt} F_X(\mathrm{d}x),$$

where F_X is the distribution function of a positive random variable X. Next to this Laplace-based stochastic representation of the generator function, McNeil and Nešlehová (2009) derive a stochastic representation for Archimedean copulas that resembles the one encountered for the class of elliptical copulas. The result is captured by the following theorem.

Theorem 2.4. *(McNeil and Nešlehová, 2009). Suppose that $\mathbf{X} = (X_1, \ldots, X_d)$ is distributed such that its survival copula is Archimedean. Then $\mathbf{X}$ has the representation*

$$\mathbf{X} \stackrel{d}{=} R\mathbf{S},$$

where the random variable $R \geq 0$ is independent of the d-dimensional random vector $\mathbf{S}$ which is uniformly distributed on the unit simplex $\Delta_d = \{\mathbf{x} \in \mathbb{R}_+^d : ||\mathbf{x}||_1 = 1\}$.

The random vector $\mathbf{X}$ is in this case said to follow a ℓ_1-norm symmetric distribution and the random variable R is referred to as the radial part of $\mathbf{X}$. Moreover, it can be shown that the generator ψ of the Archimedean copula equals both the survival function of the components of $\mathbf{X}$ and the Williamson d-transform of the distribution function of R,

$$\psi(t) = \mathscr{M}_d F_R(x) = \int_x^\infty \left(1 - \frac{x}{t}\right)^{d-1} \mathrm{d}F_R(t).$$

If $\psi = \mathscr{M}_d F_R$, then for $x \in [0, \infty)$, $F_R(x) = \mathscr{M}_d^{-1}\psi(x)$, where the inverse Williamson d-transform $\mathscr{M}_d^{-1}$ can be written down explicitely. These results are formalized in the following theorem.

Theorem 2.5. *(McNeil and Nešlehová, 2009). Let $\mathbf{U}$ be distributed according to the d-dimensional Archimedean copula C with generator ψ. Then $\left(\psi^{-1}(U_1), \ldots, \psi^{-1}(U_d)\right)$ has an ℓ_1-norm symmetric distribution with survival copula C and radial distribution F_R satisfying $F_R = \mathscr{M}_d^{-1}\psi$. Conversely, if $\mathbf{X}$ has a d-dimensional ℓ_1-norm symmetric distribution with radial distribution F_R satisfying $F_R(0) = 0$, then $\mathbf{X}$ has an Archimedean survival copula with generator $\psi = \mathscr{M}_d F_R$.*

The section finishes with several examples of frequently used parametric Archimedean copulas or copulas related to Archimedean copulas. Scatterplots of random samples and Normal contour plots of a selection of the introduced copulas are shown in Figure 2.3.2 and Figure 2.3.1. Compared to the behavior of the elliptical copulas illustrated in Figure 2.2.1 and 2.2.2, the Archimedean copulas are more flexible in their ability to model asymmetric dependence.

Clayton: The Archimedean copula with generator function $\psi(t) = (1 + \theta t)^{-1/\theta}$ and inverse generator function $\psi^{-1}(t) = \frac{1}{\theta}(t^{-\theta} - 1)$ is called the Clayton copula. In the bivariate case, the distribution function of this copula is given by

$$C(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}, \quad u_1, u_2 \in [0,1],$$

with $\theta \geq -1$ and $\theta \neq 0$. The multivariate extension is given by

$$C(u_1, \ldots, u_d) = \left(1 + \sum_{i=1}^d u_i^{-\theta} - d\right)^{-1/\theta}, \quad u_1, \ldots, u_d \in [0,1]. \tag{2.3.2}$$

Gumbel: The Archimedean copula with generator function $\psi(t) = \exp(-t^{1/\theta})$ and inverse generator function $\psi^{-1}(t) = (-\log t)^{\theta}$ is called the Gumbel copula. In the bivariate case, the distribution function of this copula is given by

$$C(u_1, u_2) = \exp\left(-\left((-\log u_1)^{\theta} + (-\log u_2)^{\theta}\right)^{1/\theta}\right), \quad u_1, u_2 \in [0,1],$$

with $\theta > 1$. Exact independence is given by $\theta = 1$. Increasing θ increases dependence. Perfect dependence is achieved as $\theta \to \infty$. The multivariate extension is given by

$$C(u_1, \ldots, u_d) = \exp\left(-\left(\sum_{i=1}^{d}(-\log u_i)^{\theta}\right)^{1/\theta}\right), \quad u_1, \ldots, u_d \in [0,1].$$

Figure 2.3.1 further illustrates the behavior of the Gumbel copula.

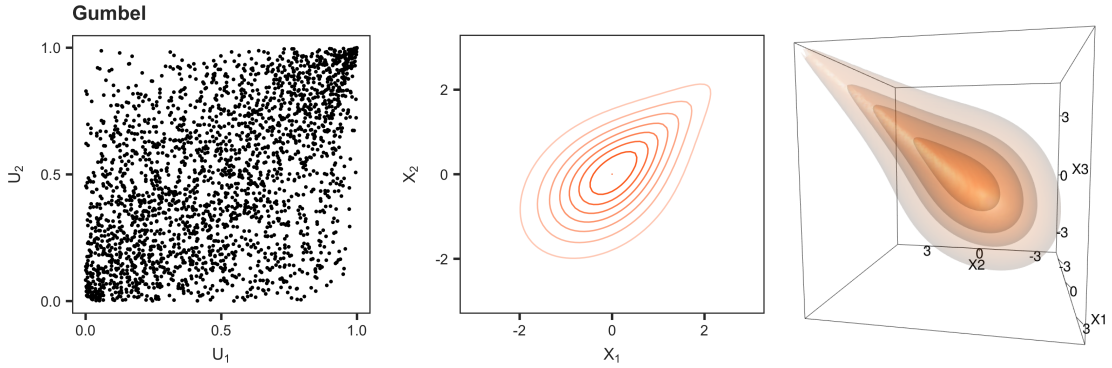

Figure 2.3.1: Scatterplot of ($N = 2500$) random samples and Normal contour plots for the bivariate and trivariate Gumbel copula. The parameters are chosen such that $\rho = 0.5$ for all bivariate margins.

Joe: The Archimedean copula with generator function $\psi(t) = 1 - (1 - \exp(-t))^{1/\theta}$ and inverse generator function $\psi^{-1}(t) = -\log\left(1 - (1-t)^{\theta}\right)$ is called the Joe copula. In the bivariate case, the distribution function of this copula is given by

$$C(u_1, u_2) = 1 - \left((1-u_1)^{\theta} + (1-u_2)^{\theta} - (1-u_1)^{\theta}(1-u_2)^{\theta}\right)^{1/\theta}, \quad u_1, u_2 \in [0,1],$$

with $\theta \geq 1$. Exact independence is given by $\theta = 1$. Increasing θ increases dependence. Perfect dependence is achieved as $\theta \to \infty$. The distribution is a special case of BB6 ($\theta = 1$), BB7 (as $\theta \to 1$) and of BB8 ($\theta = 1$) (Joe, 1997). The multivariate extension is given by

$$C(u_1, \ldots, u_d) = 1 - \left(1 - \prod_{i=1}^{d}\left(1 - (1-u_i)^{\theta}\right)\right)^{1/\theta}, \quad u_1, \ldots, u_d \in [0,1].$$

BB1: The bivariate two-parameter BB1-copula introduced by Joe and Hu (1996) is given by

$$C(u_1, u_2) = \left(1 + \left[\left(u_1^{-\theta} - 1\right)^{\delta} + \left(u_2^{-\theta} - 1\right)^{\delta}\right]^{1/\delta}\right)^{-1/\theta}, \quad u_1, u_2 \in [0,1],$$

with $\theta > 0$ and $\delta \geq 1$. Dependence increases as δ and θ increase. The independence copula is obtained as $\theta \to 0$ and $\delta \to 1$. The comononotic copula is obtained as $\theta \to \infty$ or $\delta \to \infty$. An attractive property of the BB1-copula is that every tail dependence coefficient in $(0,1)^2$ can be realized (Joe, 2015). It can be shown that the BB1-copula is an Archimedean copula with inverse generator function $\psi^{-1}(t) = \left(1 + t^{1/\delta}\right)^{-1/\theta}$. The generator is then given by $\psi(t) = \left(t^{-\theta} - 1\right)^{\delta}$. The multivariate extension is given by

$$C(u_1, \ldots, u_d) = \left(1 + \left[\sum_{i=1}^{d}\left(u_i^{-\theta} - 1\right)^{\delta}\right]^{1/\delta}\right)^{-1/\theta}, \quad u_1, \ldots, u_d \in [0,1].$$

Frank: The Archimedean copula with generator function $\psi(t) = -\frac{1}{\theta}\log\left(1+\exp(-t)\left(\exp(-\theta)-1\right)\right)$ and inverse generator function $\psi^{-1}(t) = -\log\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$ is called the Frank copula. Its bivariate distribution function is given by

$$C(u_1,u_2) = -\frac{1}{\delta}\log\left(\left\{1-e^{-\delta}-\left(1-e^{\delta u_1}\right)\left(1-e^{-\delta u_2}\right)\right\}/\left(1-e^{-\delta}\right)\right), \quad u_1,u_2\in[0,1],$$

with $0 \leq \delta \leq \infty$. Exact independence is achieved as $\delta \to 0$. Increasing δ increases dependence. Perfect dependence is achieved when $\delta \to \infty$ (Joe, 1997). The multivariate extension is given by

$$C(u_1,\ldots,u_d) = -\frac{1}{\theta}\log\left(1+\left(e^{-\theta}-1\right)^{1-d}\prod_{i=1}^{d}\left(e^{-\theta u_i}-1\right)\right), \quad u_1,\ldots,u_d\in[0,1].$$

The Frank copula is the only radially symmetric copula in the Archimedean family. Figure 2.3.2 further illustrates the behavior of the Frank copula.

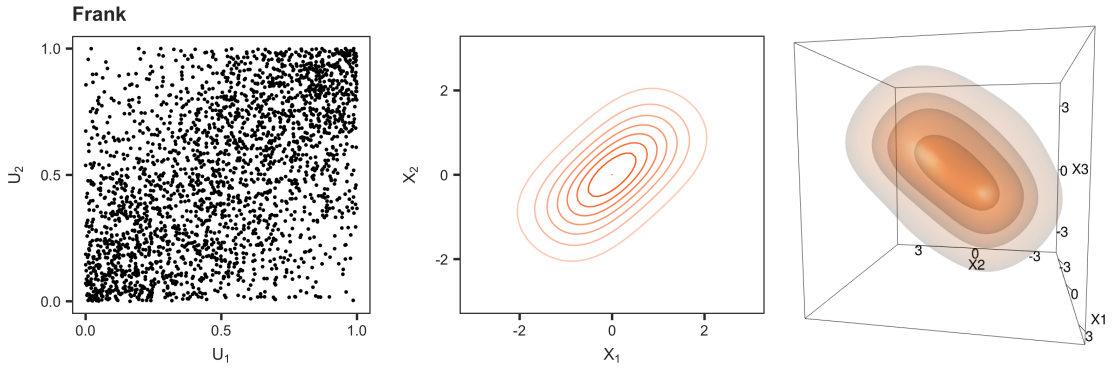

Figure 2.3.2: Scatterplot of ($N = 2500$) random samples and Normal contour plots for the bivariate and trivariate Frank copula. The parameters are chosen such that $\rho = 0.5$ for all bivariate margins.

Although the final two copulas presented below are related to Archimedean copulas, they are not Archimedean itself. In Section 3.3 it will be discussed that the Galambos copula is the extreme value copula of the survival function of Archimedean copulas. The BB4-copula is an extension hereof. The definitions are as follows.

Galambos: The bivariate Galambos copula is given by

$$C(u_1,u_2) = u_1u_2\exp\left(\left[(-\log u_1)^{-\delta}+(-\log u_2)^{-\delta}\right]^{-1/\delta}\right), \quad 0\leq\delta<\infty.$$

The dependence increases as δ increases. The independence copula is obtained for $\delta \to 0$, whereas the comononotic copula is obtained as $\delta \to \infty$ (Joe, 2015).

BB4: The bivariate BB4-copula introduced by Joe and Hu (1996) is given by

$$C(u_1,u_2) = \left(u_1^{-\theta}+u_2^{-\theta}-1-\left[\left(u_1^{-\theta}-1\right)^{-\delta}+\left(u_2^{-\theta}-1\right)^{-\delta}\right]^{-1/\delta}\right)^{-1/\theta}, \quad \theta\geq 0,\delta>0.$$

It is a gamma power mixture of the Galambos copula. The dependence increases as δ and θ increase. The comononotic copula is obtained as $\theta \to \infty$ or $\delta \to \infty$. The Galambos family is obtained as $\theta \to 0$ (Joe, 2015).

2.4 Vine copulas

The elliptical and Archimedean copulas introduced above present a wide range of different dependence structures in the bivariate case. The multivariate extensions of the models, however, offer less flexibility. For example, for the Archimedean copula, all pairwise dependencies are by definition the same. Elliptical copulas offer more flexibility since pairwise correlation parameters can be specified, but still impose a similar symmetrical dependence structure between all pairs that is either tail dependent (t-copula) or tail independent (Normal copula). A more flexible approach that has gained a lot of popularity in the last decade is the vine copula construction that was introduced in Bedford and Cooke (2001) and Bedford and Cooke (2002) and has been further developed in Aas et al. (2009) and Czado (2010). In the vine copula modeling approach, bivariate copulas are used to specify pairwise and conditional pairwise dependencies. This way, a d-dimensional copula is specified in terms of $d(d-1)/2$ pairwise copulas. The idea stems from a decomposition of the copula density into pairwise densities and conditional densities following a dependence tree structure. For example, consider the following decomposition for a 3-dimensional density f,

$$\begin{aligned}
f(x_1,x_2,x_3) &= f_{12|3}(x_1,x_2|x_3)f_3(x_3) \\
&= c_{12|3}\left(F_{1|3}(x_1|x_3),F_{2|3}(x_2|x_3);x_3\right) f_{1|3}(x_1|x_3)f_{2|3}(x_2|x_3)f_3(x_3) \\
&= c_{12|3}\left(F_{1|3},F_{2|3};x_3\right)\frac{f_{13}(x_1,x_3)}{f_3(x_3)}\frac{f_{23}(x_2,x_3)}{f_3(x_3)}f_3(x_3) \\
&= c_{12|3}\left(F_{1|3},F_{2|3};x_3\right)\frac{c_{13}\left(F_1,F_3\right)f_1f_3}{f_3}\frac{c_{23}\left(F_2,F_3\right)f_2f_3}{f_3}f_3 \\
&= c_{12|3}\left(F_{1|3},F_{2|3};x_3\right)c_{13}\left(F_1,F_3\right)c_{23}\left(F_2,F_3\right)f_1f_2f_3,
\end{aligned}$$

where for subsets $S_1, S_2 \subset \{1,2,3\}$, $F_{S_1|S_2}$ denotes the conditional distribution function, $f_{S_1|S_2}$ the corresponding conditional density, $c_{S_1|S_2}$ the corresponding copula density, and where F_j, $j=1,2,3$ denotes the marginal distribution function. Note that by conditioning other variables, different decompositions can be fashioned. In general, the conditional copulas (in the example: $c_{12|3}$) can vary for different values of the conditioning variables (in the examle: x_3).

By assuming that conditional copulas are constant for different values of the conditioning variables, the model simplifies significantly. This assumption is called the simplifying assumption. Although some distributions can be characterized with a simplified vine copula model, the simplifying assumption is not always appropriate. A multivariate copula can be represented by a vine if the bivariate linking copulas in level 2 to $d-1$ are approximately constant over the conditioning variables. Vine copulas include multivariate Gaussian and t-copula as special cases since these have linking copulas that are constant over the conditioning variables.Stöber et al. (2013) discuss classes of copulas where the simplifying assumption is satisfied. Discussions on the simplifying assumption can be found in Hobæk Haff et al. (2010) and Killiches et al. (2017).

Given the simplifying assumption, the pairwise copula densities and the conditional pairwise copula densities can be interpreted as a vine structure which leads to an attractive expression for the joint density in terms of products of bivariate copula densities. Formally, a vine is defined as follows.

Definition 2.7. *$\mathcal{V}$ is a regular vine on d elements if*

1. *$\mathcal{V} = \{T_1,\dots,T_{d-1}\}$;*
2. *T_1 is a tree with nodes $N_1 = \{1,\dots,d\}$, and edges E_1, and for $2 \le i \le d-1$, T_i is a tree with nodes $N_i = E_{i-1}$;*
3. *If for $2 \le i \le d-1$, $\{a,b\} \in E_i$ and $a = \{a_1,a_2\}$, $b = \{b_1,b_2\}$, then exactly one of the a_i is equal to one of the b_i (regularity).*

The general decomposition of the density of a copula vine is given by the following theorem. The decomposition derived above can be seen as a special case.

Theorem 2.6. *Let $\mathcal{V} = \{T_1, \ldots, T_{d-1}\}$ be a regular vine on d elements. For each edge $e(j,k) \in T_i, 1 \leq i \leq d-1$, with the conditioned set $\{j,k\}$ and the conditioning set D_e we write the corresponding copula $C_{jk|D_e}$ and its density $c_{jk|D_e}$. Furthermore, marginal distributions F_i with densities f_i, $1 \leq i \leq n$ are specified. Then the unique vine dependent distribution has a density given by*

$$f_{1\ldots d} = f_1 \ldots f_d \prod_{i=1}^{n-1} \prod_{e(j,k)\in E_i} c_{jk|D_e}\left(F_{j|D_e}, F_{k|D_e}\right)$$

where $f_i > 0$, $1 \leq i \leq d$.

Depending on the types of the trees, various vine copulas can be constructed. Two boundary cases are D-vines and C-vines. For the d-dimensional D-vine, the pairs at level 1 (also referred to as the baseline level) are $i, i+1$, for $1 \leq i \leq d-1$, and for the higher levels $l, 2 \leq l \leq d$, the conditional pairs are $i, i+l|i+1, \ldots, i+l-1$ for $1 \leq i \leq d-l$. That is, for the D-vine, conditional copulas are specified for variables i and $i+l$ given the variables indexed in between them. For the d-dimensional C-vine, the pairs at level 1 are $1, i$ for $2 \leq i \leq d$, and for level $l, 2 \leq l \leq d$, the conditional pairs are $l, i|1, \ldots, l-1$ for $l+1 \leq i \leq d$. For the C-vine, conditional copulas are specified for variables l and i given those indexed as 1 to $l-1$. Figure 2.4.1 shows the structure of a 4-dimensional D-vine.

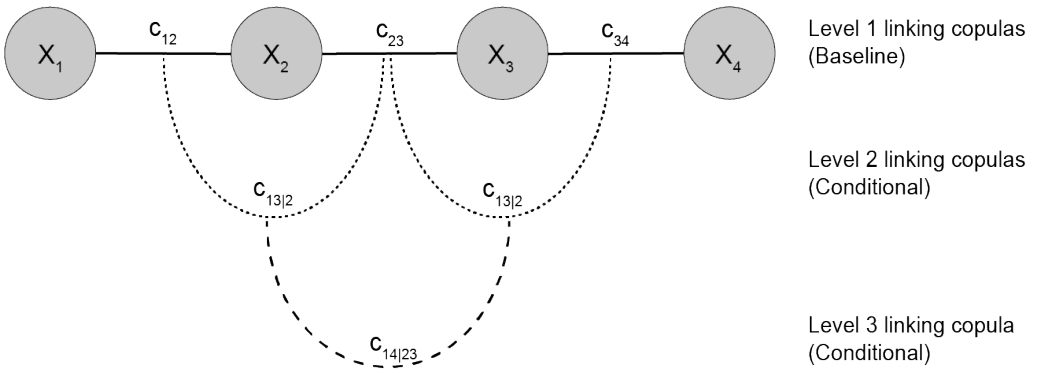

Figure 2.4.1: D-vine model for 4 variables.

In this thesis, we will consider several examples for the D-vine. As shown by Aas et al. (2009), the density of a d-dimensional D-vine is given by

$$f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}\Big(F_{i|i+1,\ldots,i+j-1}(x_i|x_{i+1}, \ldots, x_{i+j-1}),$$
$$F_{i+j|i+1,\ldots,i+j-1}(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})\Big),$$

where the index j denotes the level in the tree, while i runs over the edges in each level. Similarly, the density of a d-dimensional C-vine copula is given

$$f(x_1, \ldots, x_d) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+1|1,\ldots,j-1}\Big(F_{j|1,\ldots,j-1}(x_j|x_1, \ldots, x_{j-1}),$$
$$F_{j+1|1,\ldots,j-1}(x_{j+i}|x_1, \ldots, x_{j-1})\Big).$$

Vine copula models have lots of flexibility and members of the class should provide good approximations when trivariate and higher-order margins have conditional distributions with copulas that do not vary much over different values of the conditioning variables (Joe, 2015). To illustrate the flexibility of vines relative to the other parametric copula models, Figure 2.4.2 and Figure 2.4.3 show density contour plots for two 3-dimensional vines. Compared to the 3D-contour plots for the elliptical and Archimedean copulas, the vine copulas exhibit quite exotic density shapes.

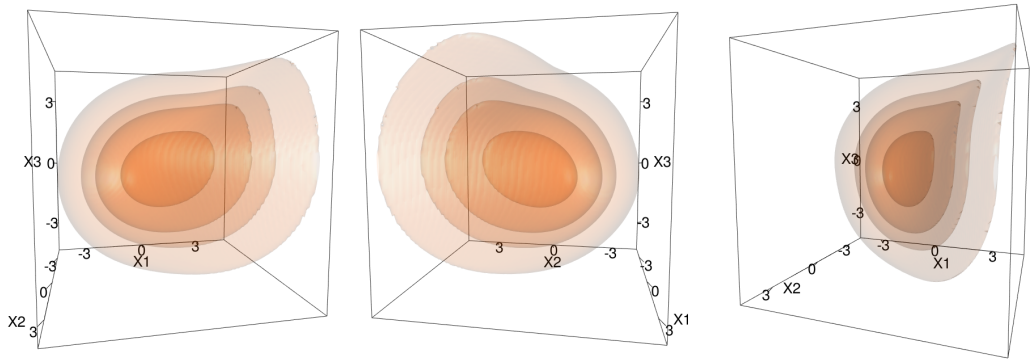

Figure 2.4.2: Trivariate Normal contour plots for a D-vine with the first baseline linking copula equal to a Gumbel copula ($\rho = 0.2$), the second baseline linking copula equal to a Frank copula ($\rho = 0.5$) and the conditional copula equal to the Gumbel copula ($\rho = 0.8$).

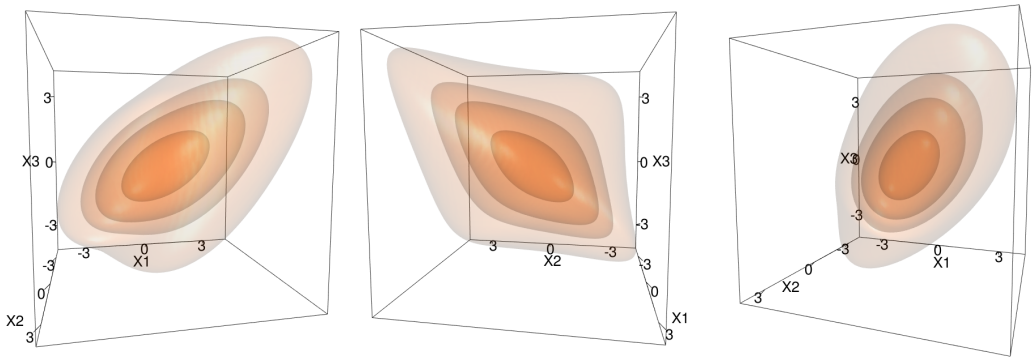

Figure 2.4.3: Trivariate Normal contour plots for a D-vine with the first baseline linking copula equal to a t-copula ($\rho = 0.5$) and both the second baseline linking copula and the conditional copula equal to the Normal copula ($\rho = 0.5$).

2.5 Empirical copulas

As a final preliminary, this section briefly addresses the issue of estimating copulas based on empirical data. To estimate the copula function, it is necessary to estimate both the marginal distributions as well as the copula distribution, since the margins are required to transform observations to the copula unit scale:

$$F(x_1, \ldots, x_d) = C_F(F_1(x_1), \ldots, F_d(x_d)), \quad x_1, \ldots, x_d \in \mathbb{R}.$$

The estimation problem can be approached entirely parametrically by assuming a parametric model for both the marginals and the copula distribution. Maximum likelihood procedures can be employed in this approach (see for example Oakes (1982)), taking either a two-step approach by first fitting the margins and, next, the copula, or a one-step approach by estimating the parameters of the margins and the copula simultaneously. Other parametric estimation methods such as the method of moments or Bayesian techniques can be used as well. Alternatively, Genest et al. (1995) have proposed a semiparametric estimation procedure where marginal distributions are fitted nonparametrically but where a parametric model is specified for the copula. Obviously, the fit of a parametric copula model is influenced heavily by observations that occur frequently and not so much by the observations that are rarely observed. Since the aim of this thesis is to assess the dependence structure between extreme events, parametric copula models are not appropriate. The estimation problem can also be approached fully nonparametrically, where both the margins and the copula function are estimated nonparametrically. This approach offers the greatest flexibility and will be discussed below.

The empirical distribution function

Assume that we have n independently and identically (iid) data observations from a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with joint distribution F and copula C_F. Data are denoted by $(X_{11}, \ldots, X_{1d}), \ldots, (X_{n1}, \ldots, X_{nd})$. The univariate marginal distributions can be estimated nonparametrically with the empirical distribution function. This estimator is defined as follows.

Definition 2.8. *Let $(X_{1j}, \ldots, X_{nj})$ be iid observations of the random variable X_j with distribution function F_j, $1 \leq j \leq d$. The empirical distribution function is given by*

$$\hat{F}_{nj}(x_j) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}\{X_{ij} \leq x_j\}. \tag{2.5.1}$$

Here $\mathbb{1}_A(x)$ is the indicator function which is equal to 1 if $x \in A$ and equal to 0 otherwise. The sum is divided by $n+1$ instead of n in order to avoid problems at the boundaries. The univariate empirical distribution functions can be used to transform data to pseudo-copula observations of the random vector $(U_1, \ldots, U_d) = (F_1(X_1), \ldots, F_d(X_d))$, as follows:

$$\left(\tilde{U}_{i1}, \ldots, \tilde{U}_{id}\right) = \left(\hat{F}_{n1}(X_{i1}), \ldots, \hat{F}_{nd}(X_{id})\right), \quad i = 1, \ldots, n. \tag{2.5.2}$$

Related to the empirical distribution function is the empirical quantile function, which can be interpreted as the generalized inverse of $\hat{F}_{nj}$:

$$\hat{F}_{nj}^{-1}(u) = \inf\{x \in \mathbb{R} : \hat{F}_{nj}(x) \geq u\} = \begin{cases} X_{k:n,j}, & \text{if } (k-1)/n < u \leq k/n, \\ -\infty, & \text{if } u = 0. \end{cases}$$

The multivariate empirical distribution function can be defined analogous to the univariate case as follows.

Definition 2.9. *Let $(X_{11}, \ldots, X_{n1}), \ldots, (X_{1d}, \ldots, X_{nd})$ be iid observations of the random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with joint distribution function F. The empirical distribution function is given by*

$$\hat{F}_n(x_1, \ldots, x_d) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}\{X_{i1} \leq x_1, \ldots, X_{id} \leq x_d\}. \tag{2.5.3}$$

The empirical copula

The empirical copula is the standard nonparametric estimator of the copula. The copula function,

$$C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right), \quad u_1, \ldots, u_d \in [0, 1],$$

is in this case estimated by employing the empirical distributions for both the joint distribution F and the marginal distributions $F_1, \ldots, F_d$. The definition is given as follows.

Definition 2.10. *Let $\left(\tilde{U}_{11}, \ldots, \tilde{U}_{n1}\right), \ldots, \left(\tilde{U}_{1d}, \ldots, \tilde{U}_{nd}\right)$ be iid pseudo-copula observations (Equation 2.5.2) of the random vector $\mathbf{X} = (X_1, \ldots, X_d)$. The empirical copula function is given by*

$$\hat{C}_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\tilde{U}_{i1} \leq u_1, \ldots, \tilde{U}_{id} \leq u_d\}. \tag{2.5.4}$$

The empirical copula depends only on the ranks of the observations $(X_{i1}, \ldots, X_{id})$; not on the specific values. The empirical copula process has been investigated by many authors in the context of process convergence. See, e.g., Deheuvels (1979), Stute (1984), Van der Vaart and Wellner (1996), Fermanian et al. (2004), and Segers (2012a). The last three references establish weak convergence of the empirical copula process to a Gaussian process under independent and dependent marginal distributions. Although the empirical copula is very flexible, converges to the true copula distribution, and can accurately approximate the observed distribution of a given sample, this nonparametric estimator is not a proper copula function because of discontinuities in the estimated copula function.

The kernel-smoothed empirical copula

A disadvantage of the empirical copula is that it is discontinuous and therefore not a proper copula. Several smoothing procedures have been proposed in the literature in order to mitigate this problem. Kernel smoothing methods are commonly employed in nonparametric statistics to estimate distribution functions. A kernel K is a positive integrable function. It is usually taken to be a symmetric unimodal probability density such that $\int_{-\infty}^{\infty} K(u)\mathrm{d}u = 1$. In order to smooth standard nonparametric estimates, the kernel function can be employed as a weighing function for data points. Specifically, the kernel-smoothed empirical distribution function is given by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{2.5.5}$$

The bandwidth h is the parameter that controls the smoothness of the estimate. Kernel smoothing methods for the empirical copula have been discussed in Gijbels and Mielniczuk (1990) and Fermanian and Scaillet (2003), for example. However, since the copula is defined on the compact unit cube, $[0,1]^d$, kernel smoothing methods suffer from severe boundary bias. Chen and Huang (2007) propose a kernel copula estimator based on local linear kernels and a simple mathematical correction that removes the boundary bias. However, their approach is not easily extended to higher dimensions. Geenens (2014) and Wen and Wu (2018) discuss the idea of applying the kernel-smoother to a transformed probit distribution that does not have a bounded support. However, the implementation of this approach is not straightforward.

The empirical beta copula

Alternatively, Segers et al. (2017) have introduced the empirical beta copula as smoothed version of the empirical copula. It is defined as follows,

$$\hat{C}_n^{\beta}(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} F_{n,R_{i,j}^{(n)}}(u_j), \tag{2.5.6}$$

where, for $u \in [0,1]$ and $r \in \{1,\ldots,d\}$,

$$F_{n,r}(u) = \mathbb{P}(U_{r:n} \leq u) = \sum_{s=r}^{n}\binom{n}{s}u^s(1-u)^{n-s} \tag{2.5.7}$$

is the cumulative distribution function of $\mathrm{Beta}(r, n+1-r)$, describing the distribution of the r-th order statistic of an independent random sample of size n from the uniform distribution of $[0,1]$. Intuitively, the beta copula smooths empirical estimates by incorporating the probability of observing the particular ranks into the estimate. The beta copula is a special case of the empirical Bernstein copula which was introduced in Sancetta and Satchell (2004). In particular, Segers et al. (2017) show that the empirical Bernstein copula is a copula if and only if all the polynomial degrees $m_1,\ldots,m_d$ are divisors of the sample size n. If all polynomial degrees are equal to the sample size n the beta copula is retrieved. Advantages of the empirical beta copula are that it is a proper copula and that it does not require the choice of a smoothing parameter (Segers et al., 2017). The asymptotic distribution of the empirical beta copula is the same as that of the empirical copula, but in small samples, it performs better both in terms of bias and variance.

Chapter 3

Tail Dependence Structures

Whereas the previous chapter covered theory regarding copulas to describe the overall dependence structure of a vector of random variables, this chapter focuses on the dependence between the tails of these variables. A first full characterization of the multivariate tail dependence structure is the multivariate extreme value distribution (MEVD) which can be seen as an extension to the univariate extreme value distribution. The MEVD models both the marginal and joint behavior of extreme values. The bivariate extreme value distribution, $d = 2$, has been known since the contributions of De Oliviera (1958), Geffroy (1958) and Sibuya (1960). Similar results for the multivariate case, $d > 2$, were obtained by De Haan and Resnick (1977) and Pickands (1981). The MEVD and its properties are discussed in Section 3.1. Since the MEVD is usually quite cumbersome to work with in practice, several alternative representations of the MEVD are discussed in Section 3.2.

Although the MEVD or one of its equivalent characterizations can theoretically be used to describe the entire multivariate tail dependence structure, this is usually not the best solution in practice because the joint distribution also contains information on marginal behavior, making it difficult to interpret dependence structures only. A copula-based dependence function that only captures dependence in the tails of distributions can solve this problem. To this end, the extreme value copula and several related dependence functions are introduced (Section 3.3-3.4). The different characterizations are applied to several copula models to get a better sense of the properties of the tail dependence characterizations *and* to assess the tail dependence structure induced by several commonly used copula models.

3.1 Extreme value distributions

Extreme value theory is concerned with assessing the behavior of extreme events. To this end, only the tails of a distribution have to be modeled because this is where extreme events occur by definition. The theory of extremes started with the work of Fisher and Tippett (1928) and Gnedenko (1943) who showed that the univariate extreme value distribution (EVD) can be divided into three classes that are all characterized by the same parameter: the extreme value index γ. The extension of the univariate EVD to higher dimensions is less straightforward to parametrize than the univariate case, since there are infinitely many ways to couple the marginal EVDs to attain an appropriate multivariate extreme value distribution. The univariate EVD and its multivariate extension are discussed in the next sections.

3.1.1 The univariate extreme value distribution

Let $X_1, \ldots, X_n$ be a sequence of identically and independently distributed (iid) observations of a random variable X with distribution function $F : \mathbb{R} \to [0, 1]$. We are interested in the upper tail of the distribution, i.e., we are interested in the distribution of extremely large values. Therefore, define the maximum of these observations as $M_n = \max(X_1, \ldots, X_n)$. It is sufficient to consider the theory for maxima only, because minima and maxima are related through the following relationship: $\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n)$. In order to find the distribution function of this maximum M_n, a normalizing sequence is needed to avoid the degenerate limiting distribution

function that occurs for the unnormalized maximum, i.e.,

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) = F^n(x) \to \begin{cases} 0 & \text{if } F(x) < 1 \\ 1 & \text{if } F(x) = 1 \end{cases} \qquad x \in \mathbb{R},$$

as $n \to \infty$. Hence, suppose there exist sequences of constants $(a_n)_{n\geq 1} > 0$ and $(b_n)_{n\geq 1} \in \mathbb{R}$ such that for a non-degenerate distribution function G the following convergence holds,

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \quad x \in \mathbb{R}. \tag{3.1.1}$$

The distribution function $G : \mathbb{R} \to [0, 1]$ is called the extreme value distribution (EVD) and F is said to be in the maximum domain of attraction of G ($F \in MDA(G)$). The EVD G takes a specific parametric form, as identified by Fisher and Tippett (1928), depending only on one parameter (Theorem 3.1).

Theorem 3.1. *The class of extreme value distribution functions is $G(c_1 x + c_2)$ with $c_1 > 0$, $c_2 \in \mathbb{R}$, where*

$$G(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \quad 1 + \gamma x > 0, \tag{3.1.2}$$

with $\gamma \in \mathbb{R}$ and where for $\gamma = 0$ the right hand side is read as $\exp(-\exp(-x))$.

The parameter γ is called the extreme value index. It is a key parameter in extreme value theory and it characterizes the heaviness of the tail of a distribution. The scale-location family that is retrieved by this extreme value distribution, is the so-called generalized extreme value distribution, given by

$$G(x) = \exp\left(-\left(1 + \gamma\frac{x - \mu}{\sigma}\right)^{-1/\gamma}\right), \quad 1 + \gamma(x - \mu)/\sigma > 0, \tag{3.1.3}$$

with $\mu, \gamma \in \mathbb{R}$ and $\sigma > 0$. The parameters γ, μ, and σ are referred to as the shape, location, and scale parameter, respectively. Based on the value of the extreme value index, three particular forms of G can be distinguished:

- $\gamma > 0$: the Fréchet distribution (Φ_α). If $F \in MDA(\Phi_\alpha)$ then it is characterized by a heavy upper tail with an infinite endpoint. Only moments up to $1/\gamma$ exist due to the heaviness of the tails. Examples are Cauchy, Pareto, and F distributions, and the Student t-distribution with a small degrees of freedom.

- $\gamma = 0$: the Gumbel distribution (Λ). If $F \in MDA(\Lambda)$ then it is characterized by a light upper tail with either a finite or an infinite endpoint. All moments exist. Examples are the Normal, Gamma, and Exponential distributions.

- $\gamma < 0$: the Weibull distribution (Ψ_α). If $F \in MDA(\Psi_\alpha)$ then it is characterized by a short upper tail with a finite endpoint. All moments exist. An example is the uniform distribution.

The result that the normalized maximum converges to one of these three types of EVDs is powerful to use in statistical analyses. However, the speed of convergence of a distribution F to its extreme value distribution G depends on the particular form of F and can vary significantly. The Normal distribution has a notoriously slow convergence rate, for example. Although the asymptotic result remains valid regardless of the convergence speed, statistical results for finite samples might be impaired by slow rates of convergences, especially if the sample size is small.

A difficulty with this convergence result is that it can be quite challenging to find suitable sequences of normalizing constants. The following Theorem 3.2 provides alternative conditions that can sometimes be checked more easily to ascertain that a particular distribution belongs to some maximum domain of attraction (MDA).

Theorem 3.2. *(Gnedenko, 1943). The real-valued variable X with distribution function F is in the domain of attraction of*

- *a Fréchet distribution, hence $\gamma > 0$, if and only if its survival function $\overline{F} = 1 - F$ is regularly varying, i.e.,*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}.$$

- *a Weibull distribution, hence $\gamma < 0$, if and only if*

$$\lim_{t\downarrow 0} \frac{1 - F(x^* - tx)}{1 - F(x^* - t)} = x^{-1/\gamma}$$

  for all $x > 0$, with a finite endpoint $x^* \in \mathbb{R}$.

- *the Gumbel distribution, hence $\gamma = 0$, if and only if*

$$\lim_{t\to x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x}$$

  for all $x \in \mathbb{R}$, where f is a suitable positive function. The endpoint x^* can be either finite or infinite.

The first two conditions are based on the concept of regular variation, which is a key concept in extreme value analysis. Intuitively, regular variation indicates that a function behaves as a power function asymptotically. Some key results on regular variation can be found in Appendix A. Besides these three conditions to identify the extreme value distribution of F, de Haan and Ferreira (2006) present a range of alternative conditions that can be used.

3.1.2 The multivariate extreme value distribution

Following a similar approach as for the univariate extreme value distribution, the multivariate extreme value distribution (MEVD) will now be derived. Let $\mathbf{X_1}, \ldots, \mathbf{X_d}$ be independent and identically distributed (iid) random vectors in $\mathbb{R}^n$ with joint distribution function $F : \mathbb{R}^d \to [0, 1]$, marginal distribution functions $F_1, \ldots, F_d : \mathbb{R} \to [0, 1]$, and copula $C_F : [0, 1]^d \to [0, 1]$. The aim is to assess the joint distribution of the vector of componentwise maxima of each of the variables $X_1, \ldots, X_d$,

$$\mathbf{M_n} = (M_{n1}, \ldots, M_{nd}), \quad M_{nj} = \max\left(X_{1j}, \ldots, X_{nj}\right), \quad j = 1, \ldots, d. \tag{3.1.4}$$

Note that also for the multivariate case, it is sufficient to consider the theory for maxima only because of the relationship between maxima and minima. Similarly to the univariate case, the componentwise maxima need to be normalized to find a non-degenerate joint distribution of the vector of maxima $\mathbf{M_n}$. If for some choice of vectors $(\mathbf{a_n})_{n\geq 1} = \{(a_{n1}, \ldots, a_{nd})\}_{n\geq 1} \in (0, \infty)^d$ and $(\mathbf{b_n})_{n\geq 1} = \{(b_{n1}, \ldots, b_{nd})\}_{n\geq 1} \in \mathbb{R}^d$, the following convergence relationship holds,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(\frac{M_{n1} - b_{n1}}{a_{n1}} \leq x_1, \ldots, \frac{M_{nd} - b_{nd}}{a_{dn}} \leq x_d\right) \\ = \lim_{n\to\infty} F^n(a_{n1}x_1 + b_{n1}, \ldots, a_{nd}x_d + b_{nd}) = G(x_1, \ldots, x_d), \quad (x_1, \ldots, x_d) \in \mathbb{R}^d,\end{aligned} \tag{3.1.5}$$

where $G : \mathbb{R}^d \to [0, 1]$ is a non-degenerate distribution function, we say that F belongs to the maximum domain of attraction of the distribution function G ($F \in MDA(G)$). The distribution function G is in this case a multivariate extreme value distribution (MEVD). The marginal distributions $G_1, \ldots, G_d : \mathbb{R} \to [0, 1]$, must be univariate extreme value distributions. That is,

$$G_j(x) = \exp\left((1 + \gamma_j x)^{-1/\gamma_j}\right),$$

for all $j \in \{1, \ldots, d\}$, where this specific parametric form is implied by Theorem 3.1. The example below illustrates how a bivariate extreme value distribution can be derived from the convergence relationship (Equation 3.1.5).

Example 3.1. *We consider a standard bivariate exponential survival function* $\overline{F} : \mathbb{R}_+^2 \to [0,1]$,

$$\overline{F}(x_1, x_2) = \left(e^{x_1} + e^{x_2} - 1\right)^{-1}, \quad x_1, x_2 \in \mathbb{R},$$

following an example covered in Mardia (1970), Galambos (1978), and Marshall and Olkin (1983). The corresponding bivariate distribution function $F : \mathbb{R}_+^d \to [0,1]$ *is given by*

$$F(x_1, x_2) = 1 - \overline{F}_1(x_1) - \overline{F}_2(x_2) + \overline{F}(x_1, x_2) = 1 - e^{-x_1} - e^{-x_2} + \left(e^{x_1} + e^{x_2} - 1\right)^{-1},$$

where the inclusion-exclusion principle (Equation 2.1.6) was used. Note that the marginal distributions are univariate standard exponential distributions, since

$$F_1(x_1) = F(x_1, \infty) = 1 - e^{-x_1}, \qquad F_2(x_2) = F(\infty, x_2) = 1 - e^{-x_2}.$$

The convergence relationship (Equation 3.1.5) implies that the bivariate extreme value distribution G *to which the distribution* F *converges can be identified as follows,*

$$\begin{aligned} G(x_1, x_2) &= \lim_{n\to\infty} F^n(a_{n1}x_1 + b_{n1}, a_{n2}x_2 + b_{n2}) \\ &= \lim_{n\to\infty} \left(1 - e^{-(a_{n1}x_1 + b_{n1})} - e^{-(a_{n2}x_2 + b_{n2})} + \left(e^{a_{n1}x_1 + b_{n1}} + e^{a_{n2}x_2 + b_{n2}} - 1\right)^{-1}\right)^n, \end{aligned}$$

given that there exist sequences $\{(a_{n1}, a_{n2})\}_{n\geq 1} \in (0,\infty)^2$ *and* $\{(b_{n1}, b_{n2})\}_{n\geq 1} \in \mathbb{R}^2$ *such that this limit exists. Taking* $a_{n1} = a_{n2} = 1$ *and* $b_{n1} = b_{n2} = \log n$*, we find that*

$$\begin{aligned} &\lim_{n\to\infty} F^n(a_{n1}x_1 + b_{n1}, a_{n2}x_2 + b_{n2}) \\ &\quad= \lim_{n\to\infty} \left(1 - e^{-(x_1 + \log n)} - e^{-(x_2 + \log n)} + \left(e^{x_1 + \log n} + e^{x_2 + \log n} - 1\right)^{-1}\right)^n \\ &\quad= \lim_{n\to\infty} \left(1 - \frac{e^{-x_1}}{n} - \frac{e^{-x_2}}{n} + \left(ne^{x_1} + ne^{x_2} - 1\right)^{-1}\right)^n \\ &\quad= \lim_{n\to\infty} \left(\exp\left(-\frac{e^{-x_1}}{n} - \frac{e^{-x_2}}{n} + \frac{1}{ne^{x_1} + ne^{x_2} - 1}\right)\right)^n \\ &\quad= \exp\left(-\left(e^{-x_1} + e^{-x_2} - \left(e^{x_1} + e^{x_2}\right)^{-1}\right)\right), \end{aligned}$$

where we used $e^{-x} \sim 1 - x$ *for* $x \to 0$ *(Taylor expansion). It follows that the bivariate extreme value distribution* G *to which* F *converges is given by*

$$G(x_1, x_2) = \exp\left(-\left(e^{-x_1} + e^{-x_2} - \left(e^{x_1} + e^{x_2}\right)^{-1}\right)\right).$$

The MEVD contains all information on the marginal behavior of the componentwise maxima $\mathbf{M}_n$ *and* on the dependence structure between the componentwise maxima of $\mathbf{X_1}, \ldots, \mathbf{X_d}$. The stronger the dependence between the extreme values of $\mathbf{X}$, the more likely it is to observe tail events simultaneously. On the other hand, if the componentwise maxima of $\mathbf{X}$ are independent, it is very improbable to observe tail events simultaneously. The concept of tail independence is formally defined as follows.

Definition 3.1. *Let* G *be a multivariate extreme value distribution with marginal distributions* $\exp(-(1+\gamma_j x)^{-1/\gamma_j})$ *for* $j = 1, \ldots, d$*. If* $G(x_1, \ldots, x_d) = \prod_{j=1}^d \exp(-(1+\gamma_j x_j)^{-1/\gamma_j})$*, then the random variables are said to be asymptotically independent or tail independent. Otherwise, the random variables are said to be asymptotically dependent or tail dependent.*

An interesting property of the MEVD is that it cannot contain negative dependence. That is, all multivariate extreme value distributions are positive orthant dependent, hence

$$G(x_1, \ldots, x_d) \geq \prod_{j=1}^d G_j(x_j), \tag{3.1.6}$$

as shown by De Oliviera (1962). This property is called positive quadrant dependence for $d = 2$ (Lehmann, 1966). The weakest dependence structure between multivariate extremes is therefore independence. This will turn out to have important implications for dependence functions of the

extreme value distribution that will be discussed in the next sections. In particular, the dependence functions need to be bounded between independence and comonotonicity (perfect positive dependence).

A defining property of the multivariate extreme value distribution is max-stability. Intuitively, this means that taking maxima (i.e., multiplying the distribution function) only rescales the distribution function, but leaves it unchanged otherwise. Hence, the distribution is stable under taking maxima. Formally, this can be given as follows.

Definition 3.2. *A d-dimensional distribution function $G(x)$ is called max-stable if for $j = 1, \ldots, d$ and every $t > 0$ there exist functions $a_j(t) > 0, b_j(t)$ such that*

$$G^t(x_1, \ldots, x_d) = G(a_1(t)x_1 + b_1(t), \ldots, a_d(t)x_d + b_d(t)). \tag{3.1.7}$$

Proposition 3.1. *(Resnick, 1987). The class of multivariate extreme value distributions is precisely the class of max-stable distribution functions with nondegenerate marginals.*

Proof. Let G be a max-stable distribution. By the definition of max-stability, it is possible to write

$$\lim_{n\to\infty} G^n(a_1(1/n)x_1 + b_1(1/n), \ldots, a_d(1/n)x_d + b_d(1/n)) = G(x_1, \ldots, x_d)$$

for certain functions $a_j(t) > 0$, $b_j(t) \in \mathbb{R}$, $1 \leq j \leq d$. Therefore, it is clear that the class of max-stable distribution functions with nondegenerate marginals belongs to the class of multivariate extreme value distributions. Conversely, let G be a multivariate extreme value distribution. Then,

$$\begin{aligned} G^t(x_1, \ldots, x_d) &= \lim_{n\to\infty} F^{nt}(a_{n1}x_1 + b_{n1}, \ldots, a_{nd}x_d + b_{nd}) \\ &= \lim_{m\to\infty} F^m(a_{m1}x_1 + b_{m1}, \ldots, a_{md}x_d + b_{md}) = G(x_1, \ldots, x_d), \end{aligned}$$

where the normalizing sequences can be chosen in a way such that the convergences hold. Hence, G is max-stable and belongs to its own maximum domain of attraction. ■

Analogous to the univariate case, multivariate regular variation (MRV) can characterize a large class of multivariate extreme value distributions. Specifically, Resnick (1987) showed that a random vector $\mathbf{X} \geq \mathbf{0}$ with distribution function F belongs to the maximum domain of attraction of a multivariate extreme value distribution G if and only if marginal convergences hold *and* the following transformed distribution function of F,

$$F_*(\mathbf{x}) = F\left(F_1^{-1}\left(1 - \frac{1}{x_1}\right), \ldots, F_d^{-1}\left(1 - \frac{1}{x_d}\right)\right),$$

is multivariate regularly varying. The function $U(x) = F^{-1}(1 - 1/x)$ is known as the tail quantile function in extreme value theory and zooms in on the high quantiles of a distribution function (Beirlant et al., 2004). Multivariate regular variation can be defined in several equivalent ways, but the following definition will be used throughout this thesis.

Definition 3.3. *A random vector $\mathbf{X} \geq \mathbf{0}$ is multivariate regularly varying if there exist an index $\alpha > 0$ and a Radon probability measure μ (i.e., finite on compact sets) on $\mathbb{S} = \{\omega \in \mathbb{R}^d : ||\omega|| = 1\}$, the unit hypersphere with respect to norm $||\cdot||$, such that*

$$\lim_{t\to\infty} \frac{\mathbb{P}(||\mathbf{X}|| \geq tx, \mathbf{X}/||\mathbf{X}|| \in A)}{\mathbb{P}(||\mathbf{X}|| \geq t)} = x^{-\alpha}\mu(A), \tag{3.1.8}$$

for every $x > 0$ and Borel set $A \subseteq \mathbb{S}$ with $\mu(\partial A) = 0$.

If $\mathbf{X}$ is multivariate regularly varying with index $\alpha > 0$, this is denoted by $\mathbf{X} \in MRV_\alpha$. The definition of multivariate regular variation is independent of the specification of the norm $||\cdot||$ since all norms on $\mathbb{R}^d$ are equivalent (Hult and Lindskog, 2002). Alternative definitions of multivariate regular variation and more information on this property can be found in Appendix A. Intuitively, multivariate regular variation implies that the distribution of $\mathbf{X}$ can be decomposed into a radial part and an angular part. The radial part corresponds to the conditional probability that the norm of $\mathbf{X}$ is large. The distribution of the norm of the random vector $\mathbf{X}$ is assumed to be regularly varying at infinity, which shows a clear parallel with Theorem 3.2 for the univariate case. The angular part describes how the standardized components of the random vector $\mathbf{X}$ are distributed given that the sum of the components of $\mathbf{X}$ is large. The angular part has a clear connection with the spectral measure which will be introduced in Section 3.2.2.

Tail densities

If the multivariate regular variation convergence condition in Equation 3.1.8 holds at the density level, it gives rise to the so-called tail density. The tail density $q(\cdot)$ was introduced by De Haan and Omey (1983) and De Haan and Resnick (1987) and can be seen as a density-based convergence condition. Such a condition can be convenient to assess the extremal behavior of distributions that are (most easily) defined by their density. Elliptical distributions and vine copula models are examples of distributions that are defined based on their density and where the cumulative distribution function cannot be determined explicitly. To introduce the tail density, we assume that the marginals $F_1, \ldots, F_d$ have equivalent tails. The margins of the random vector $\mathbf{X} = (X_1, \ldots, X_d)$ are said to be right-tail equivalent if

$$\frac{1 - F_j(x)}{1 - F_1(x)} \to 1, \quad x \to \infty, \tag{3.1.9}$$

for all $j = 1, \ldots, d$. Formally, the tail density is defined as follows.

Definition 3.4. *Let $\mathbf{X} \geq \mathbf{0}$ be a right-tail equivalent multivariate regularly varying random vector with index $\alpha > 0$. Denote the joint distribution function by F and suppose that it has a continuous and positive density f. If for some positive function q and some positive function V, regularly varying at infinity with negative index $-\alpha$, we have*

$$\lim_{t \to \infty} \frac{f(t\mathbf{z})}{t^{-d} V(t)} = q(\mathbf{z}), \tag{3.1.10}$$

on $\overline{\mathbb{R}}_+^d \backslash \{0\}$ and uniformly on $\{\mathbf{z} > 0 : ||\mathbf{z}|| = 1\}$, then q is said to be the tail density of $\mathbf{X}$.

Note that the fact that the marginals are assumed to be right-tail equivalent is not a limitation, since the marginal distributions can be transformed to the same scale in such a way that this condition is satisfied. Since V is assumed to be regularly varying at infinity with negative index $-\alpha$, this function can be represented as $V(t) = \mathscr{L}(t) t^{-\alpha}$, where $\mathscr{L}$ is a slowly varying function (see Appendix A). Therefore, the tail density can also be represented as

$$\lim_{t \to \infty} \frac{f(t\mathbf{z})}{t^{-\alpha - d} \mathscr{L}(t)} = q(\mathbf{z}).$$

In the opposite direction, it can be shown that if Equation 3.1.10 holds, then for any $\mathbf{z} \in \mathbb{R}_+^d \backslash \{0\}$,

$$\lim_{t \to \infty} \frac{1 - F(t\mathbf{z})}{V(t)} = \int_{[\mathbf{0}, \mathbf{z}]^c} q(\mathbf{y}) \mathrm{d}\mathbf{y}, \tag{3.1.11}$$

with homogeneous property that $q(s\mathbf{z}) = s^{-\alpha - d} q(\mathbf{z})$ for $s > 0$. Hence, if the tail density is homogeneous of order $-\alpha - d$, then F is multivariate regularly varying with index α. Furthermore, it can be shown that given that Equation 3.1.10 converges uniformly on the unit sphere of $\mathbb{R}_+^d$, then the regular variation of the density implies multivariate regular variation of its cumulative distribution function (De Haan and Resnick, 1987). Specifically, if $f \in MRV_\alpha$, then $F \in MRV_{\alpha + d}$. This can be seen as a multivariate extension of Karamata's theorem (Theorem A.1) with an additional condition. The next example illustrates how the tail density can be identified.

Example 3.2. *Consider the bivariate Cauchy distribution with density $f : \mathbb{R}^2 \to \mathbb{R}$ given by,*

$$f(x, y) = \frac{1}{2\pi} \left(1 + x^2 + y^2\right)^{-3/2}, \quad (x, y) \in \mathbb{R}^2.$$

The aim is to find the tail density q of this distribution. Taking $V(t) = t^{-1}$, we find,

$$\begin{aligned} \lim_{t \to \infty} \frac{f(tx, ty)}{t^{-2} V(t)} &= \lim_{t \to \infty} \frac{1}{2\pi} \left(1 + t^2 x^2 + t^2 y^2\right)^{-3/2} \frac{1}{t^{-3}} = \lim_{t \to \infty} \frac{1}{2\pi} \frac{t^3}{\left(1 + t^2 x^2 + t^2 y^2\right)^{3/2}} \\ &= \lim_{t \to \infty} \frac{1}{2\pi} \frac{1}{\left(t^{-2} + x^2 + y^2\right)^{3/2}} = \frac{1}{2\pi} \left(x^2 + y^2\right)^{-3/2} =: q(x, y). \end{aligned}$$

Since

$$q(tx, ty) = \frac{1}{2\pi} \left(t^2 \left(x^2 + y^2\right)\right)^{-3/2} = t^{-3} q(x, y),$$

q is homogeneous of order 3 and, hence, F is multivariate regularly varying with index $\alpha = 1$.

3.2 Characterizations of the MEVD

Besides the joint distribution function, there are other ways to characterize the multivariate extreme value distribution. These characterizations can be useful because the convergence-based definition of the extreme value distribution (Equation 3.1.5) is cumbersome to use in practice, especially for higher dimensions. In this section, the exponent measure and the spectral measure will be covered following the exposition of Beirlant et al. (2004) on the subjects. Similar to the MEVD, these characterizations fully describe the tail dependence structure but also contain information on the marginal behavior of extremes.

3.2.1 Exponent measure

The construction of the exponent measure starts with the observation that the convergence relation given in Equation 3.1.5,

$$\lim_{n\to\infty} F^n(a_{1n}x_1 + b_{1n}, \ldots, a_{dn}x_d + b_{dn}) = G(x_1, \ldots, x_d),$$

is equivalent to,

$$\lim_{n\to\infty} n\left(1 - F(a_{1n}x_1 + b_{1n}, \ldots, a_{dn}x_d + b_{dn})\right) = -\log G(x_1, \ldots, x_d), \tag{3.2.1}$$

since $\log(x) \sim x - 1$ for $x \to 1$ (Taylor expansion). This implies that there exists a measure ν on $\mathbb{R}_+^d$ such that

$$\nu\left([0,\infty]^d \backslash [\mathbf{0}, \mathbf{x}]\right) = -\log G(\mathbf{x}), \tag{3.2.2}$$

due to the fact that G is the distribution function of a probability measure (De Haan and Resnick, 1977). The measure ν is called the exponent measure and assigns finite values to Borel sets bounded away from the origin. Since the measure ν is related to the MEVD G by the relationship,

$$G(\mathbf{x}) = \exp\left(-\nu\left([0,\infty]^d \backslash [\mathbf{0}, \mathbf{x}]\right)\right), \quad \mathbf{x} \in \mathbb{R}^d,$$

it is called the exponent measure. The definition of the exponent measure can also be formulated in terms of vague convergence of measures on $[0,\infty]^d \backslash \{\mathbf{0}\}$,

$$n\mathbb{P}\left(\frac{\mathbf{X} - \mathbf{b}_n}{\mathbf{a}_n} \in \cdot\right) \to \nu(\cdot), \tag{3.2.3}$$

(De Haan and Resnick, 1993). Vague convergence concerns the convergence of measures that are finite on bounded sets and coincides with weak convergence in certain spaces (Basrak and Planinić, 2018). Alternatively, an equivalent definition of the exponent measure can be retrieved based on the convergence of a sequence of point processes to a limiting Poisson process (De Haan, 1984).

An important property of the exponent measure is that it is homogeneous of order one, i.e., $\nu(s\cdot) = s^{-1}\nu(\cdot)$. In order to see that this is in fact the case, we use the max-stability property of the extreme value distribution (Proposition 3.1). For rectangles bounded away from the origin this implies that,

$$\nu\left([0,\infty]^d \backslash [\mathbf{0}, s\mathbf{x}]\right) = -\log G(s\mathbf{x}) = -\log G^{1/s}(\mathbf{x}) = -s^{-1}\log G(\mathbf{x}) = s^{-1}\nu\left([0,\infty]^d \backslash [\mathbf{0}, \mathbf{x}]\right).$$

Since the homogeneity holds for all rectangles in $\mathbb{R}_+^d \backslash \{\mathbf{0}\}$, it can be shown that it holds for all Borel subsets in $\mathbb{R}_+^d \backslash \{\mathbf{0}\}$. Hence, $\nu(s\cdot) = s^{-1}\nu(\cdot)$. Because of this property, the exponent measure can be used to scale a region to another region where observations are available (Figure 3.2.1). This is useful for estimating the probability of a set of rare events for which no or little observations are available.

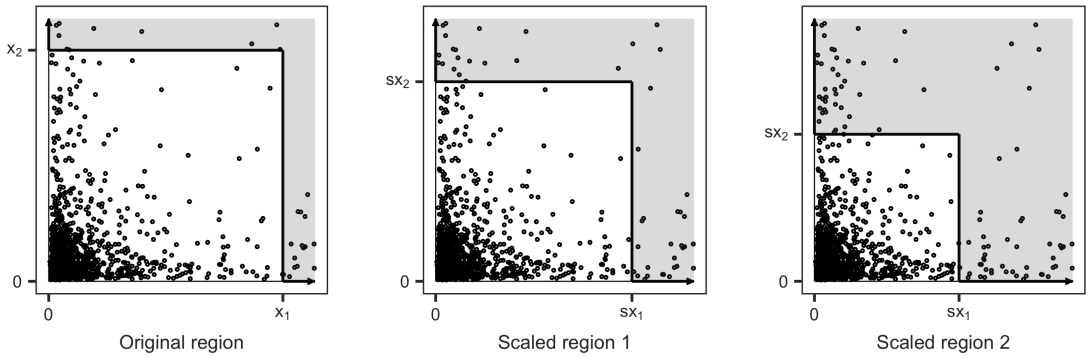

Figure 3.2.1: Scaling extreme regions using the homogeneity property of the exponent measure ν in $\mathbb{R}^2_+$ for the bivariate case.

Other properties of the exponent measure are related to the bounds of the dependence structure of the MEVD G. Recall from Section 3.1.2 that the MEVD is bounded between comonotonicity and independence and in these boundary cases the exponent measure takes special forms. To show these results, first note that by the relationship with the MEVD G, the marginal exponent measure of a subset $S \subset \{1, \ldots, d\}$, denoted by ν_S, can be retrieved by letting arguments of components indexed by numbers not belonging to S go to infinity. For example, the marginal exponent measure between the variables X_1 and X_2, ν_{12}, can be retrieved from the exponent measure ν as follows:

$$\begin{aligned}\nu_{12}\left([0,\infty]^2 \backslash [\mathbf{0}, (x_1, x_2)]\right) &= -\log G_{12}(x_1, x_2) = -\log G(x_1, x_2, \infty, \ldots, \infty) \\ &= \nu\left([0,\infty]^d \backslash [\mathbf{0}, (x_1, x_2, \infty, \ldots, \infty)]\right).\end{aligned}$$

Similarly, the univariate exponent measures can be defined as

$$\begin{aligned}\nu_j\left((x_j, \infty]\right) &= -\log G_j(x_j) = (1 + \gamma_j x_j)^{-1/\gamma_j} = -\log G(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty) \\ &= \nu\left([0,\infty]^d \backslash [\mathbf{0}, (\infty, \ldots, \infty, x_j, \infty, \ldots, \infty)]\right).\end{aligned}$$

Now we can deduce that when the MEVD G is independent, the exponent measure is equal to the sum of the univariate marginal contributions to the exponent measure,

$$\begin{aligned}\nu\left([0,\infty]^d \backslash [\mathbf{0}, \mathbf{x}]\right) &= -\log G(\mathbf{x}) = -\log \mathbb{P}\left(X_1 \leq x_1, \ldots, X_d \leq x_d\right) = -\log\left(\prod_{j=1}^{d} \mathbb{P}\left(X_j \leq x_j\right)\right) \\ &= \sum_{j=1}^{d}\left(-\log \mathbb{P}\left(X_j \leq x_j\right)\right) = \sum_{j=1}^{d}\left(-\log G_j(x_j)\right) = \sum_{j=1}^{d} \nu_j\left([0,\infty] \backslash [0, x_j]\right). \quad (3.2.4)\end{aligned}$$

Hence, in the case of tail independence, the exponent measure is concentrated on the axes in $\mathbb{R}^d_+$, and, therefore, does not assign mass to regions that do not contain values on the axes:

$$\nu\left([\mathbf{x}, \infty]\right) = 0, \quad (3.2.5)$$

with $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d_+$ such that $x_j > 0$ for all $j = 1, \ldots, d$. In contrast, if the MEVD G is tail comonotonic, the exponent measure is concentrated on the main diagonal,

$$\begin{aligned}\nu\left([0,\infty]^d \backslash [\mathbf{0}, \mathbf{x}]\right) &= -\log G(\mathbf{x}) = -\log \mathbb{P}\left(X_1 \leq x_1, \ldots, X_d \leq x_d\right) = -\log \min_{1 \leq j \leq d} \mathbb{P}\left(X_j \leq x_j\right) \\ &= \max_{1 \leq j \leq d}\left(-\log G_j\left(x_j\right)\right) = \max_{1 \leq j \leq d} \nu_j\left((x_j, \infty]\right) = \nu\left(\left(\min_{1 \leq j \leq d} x_j, \infty\right]\right), \quad (3.2.6)\end{aligned}$$

where for the last step it has to be assumed that the margins are right-tail equivalent. Figure 3.2.2 illustrates these regions for the bivariate case.

These observations can be employed to show the result that pairwise tail independence implies multivariate tail independence. This is an important finding since this is not true for general dependence structures. The result is captured by Theorem 3.3.

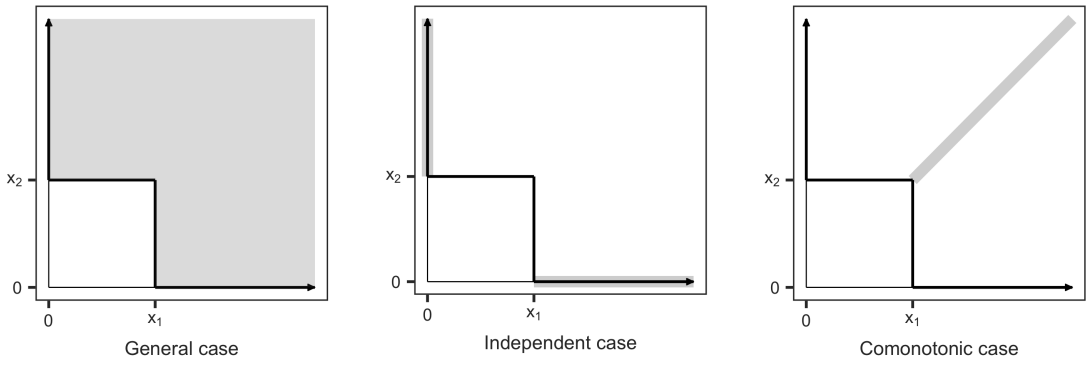

Figure 3.2.2: Regions in $\mathbb{R}^2_+$ where the exponent measure ν is concentrated for the bivariate case in general (left) in the independent case (middle) and in the comonotonic case (right).

Theorem 3.3. *(Based on Theorem 6.2.3 in de Haan and Ferreira (2006)). Let $(X_1, \ldots, X_d)$ be a vector of d random variables. All pairs X_i and X_j, $1 \leq i < j \leq d$, are tail independent if and only if $(X_1, \ldots, X_d)$ are tail independent. That is, pairwise asymptotic independence implies multivariate asymptotic independence, and vice versa,*

$$\forall\ 1 \leq i < j \leq d : (X_i, X_j) \in MDA(\Pi) \Leftrightarrow (X_1, \ldots, X_d) \in MDA(\Pi).$$

Proof. (Based on the proof in de Haan and Ferreira (2006) and the original proof of Berman (1961)). The proof hinges on the notion that the exponent measure equals the sum of univariate marginal exponent measures for tail independent variables (Equation 3.2.4) and therefore assigns zero mass to sets that do not contain values on the axes in $\mathbb{R}^d_+$ (Equation 3.2.5). If the vector $\mathbf{X} = (X_1, \ldots, X_d)$ is tail independent it follows immediately that Equation 3.2.5 holds for all bivariate pairs, and, therefore, that all bivariate pairs are tail independent. Conversely, if all pairs are tail independent, marginal bivariate exponent measures are concentrated on the axes (Equation 3.2.4), and, hence, equal zero on regions bounded away from the axes (Equation 3.2.5). We will show that this implies $\nu([\mathbf{x}, \infty]) = 0$ for all $\mathbf{x} \in \mathbb{R}^d_+$ such that $x_j > 0$ for all $j = 1, \ldots, d$ and, therefore, multivariate tail independence.

Let $\mathbf{x}^* \in \mathbb{R}^d_+$ be such that $\mathbf{x}^*_j > 0$ for all $j = 1, \ldots, d$. The exponent measure evaluated on the set where all components of $\mathbf{x}^*$ are large, i.e., $[\mathbf{x}^*, \infty]$, is smaller than (or equal to) the exponent measure evaluated on the set where at least one component of $\mathbf{x}^*$ is large, $[0, \infty]^d \backslash [\mathbf{0}, \mathbf{x}^*]$, *minus* the exponent measure evaluated on the axes included in this region:

$$\nu([\mathbf{x}^*, \infty]) \leq \nu\left([0, \infty]^d \backslash [\mathbf{0}, \mathbf{x}^*]\right) - \sum_{j=1}^{d} \nu_j\left((x^*_j, \infty]\right).$$

By the inclusion-exclusion principle (Equation 2.1.6), we may write

$$\begin{aligned}
\nu\left([0, \infty]^d \backslash [\mathbf{0}, \mathbf{x}^*]\right) &= \nu\left(\bigcup_{j=1}^{d} [x^*_j, \infty] \times (0, \infty]^{d-1}\right) \\
&= \sum_{j=1}^{d} (-1)^{j+1} \sum_{1 \leq i_1 \leq \cdots \leq i_j \leq d} \nu\left(\bigcap_{k=1}^{j} [x^*_{i_k}, \infty] \times (0, \infty]^{d-1}\right) \\
&= \sum_{j=1}^{d} \nu\left([x^*_j, \infty] \times (0, \infty]^{d-1}\right) - \sum_{i \neq j} \nu\left(\left([x^*_i, \infty] \times (0, \infty]^{d-1}\right) \cap \left([x^*_j, \infty] \times (0, \infty]^{d-1}\right)\right) - \ldots \\
&= \sum_{j=1}^{d} \nu_j\left((x^*_j, \infty]\right) - \sum_{i \neq j} \nu\left(\left([x^*_i, \infty] \times (0, \infty]^{d-1}\right) \cap \left([x^*_j, \infty] \times (0, \infty]^{d-1}\right)\right) - \ldots.
\end{aligned}$$

For all bivariate intersections, a similar argument can be used to find that for any $1 \leq i < j \leq d$,

$$\begin{aligned}
&\nu\left(\left([x_i^*, \infty] \times (0, \infty]^{d-1}\right) \cap \left([x_j^*, \infty] \times (0, \infty]^{d-1}\right)\right) \\
&\quad \leq \nu\left([0, \infty]^d \setminus \left([0, x_i^*] \times [0, x_j^*] \times [0, \infty]^d\right)\right) - \nu_i\left((x_i^*, \infty]\right) - \nu_j\left((x_j^*, \infty]\right) \\
&\quad = \nu_{ij}\left([0, \infty]^2 \setminus [\mathbf{0}, (x_i^*, x_j^*)]\right) - \nu_i\left((x_i^*, \infty]\right) - \nu_j\left((x_j^*, \infty]\right) = 0,
\end{aligned}$$

where the last equality holds by the assumption of independent bivariate margins and because of Equation 3.2.4. Since this is true for all bivariate intersections and since all higher order intersections are smaller than the maximum of the bivariate intersections, we can conclude that

$$\nu\left([0, \infty]^d \setminus [\mathbf{0}, \mathbf{x}^*]\right) = \sum_{j=1}^{d} \nu_j\left((x_j^*, \infty]\right).$$

Hence, $0 \leq \nu\left([\mathbf{x}^*, \infty]\right) \leq 0$, which remained to be shown. ■

3.2.2 Spectral measure

Another characterization of the multivariate extreme value distribution can be found based on the exponent measure by switching from the cartesian coordinate system to the polar coordinate system. It is convenient to standardize the exponent measure to the so-called simple exponent measure to derive this characterization. The simple exponent measure ν_* is defined as the exponent measure for the MEVD with standard Fréchet margins. Recalling that the marginal distributions of the extreme value distribution are of the form $G_j(x) = \exp\left(-(1 + \gamma_j x)^{-1/\gamma_j}\right)$, $1 \leq j \leq d$, and that the standard Fréchet distribution is given by $F(x) = \exp(-1/x)$, it follows that the simple MEVD, denoted by G_*, is given by

$$G_*(\mathbf{z}) = G\left(\frac{z_1^{\gamma_1} - 1}{\gamma_1}, \ldots, \frac{z_d^{\gamma_d} - 1}{\gamma_d}\right). \tag{3.2.7}$$

The exponent measure for the simple extreme value distribution is therefore defined as

$$\nu_*\left([0, \infty]^d \setminus [\mathbf{0}, \mathbf{z}]\right) = -\log G_*(\mathbf{z}) = -\log G\left((\mathbf{z}^\gamma - \mathbf{1})/\gamma\right). \tag{3.2.8}$$

After deriving expressions based on the simple MEVD, the results can be transformed easily to the case of a general MEVD. The polar coordinates are defined for a point $\mathbf{z} \in \mathbb{R}^d$ as $r = ||\mathbf{z}||_p$ and $\omega = \mathbf{z}/||\mathbf{z}||_q$ for any two norms $||\cdot||_p$ and $||\cdot||_q$ on $\mathbb{R}^d$. They are obtained with a transformation $T : \mathbb{R} \to (0, \infty) \times \mathbb{S}$, where $\mathbb{S} = \{\omega \in \mathbb{R}^d : ||\omega||_q = 1\}$ is the unit sphere with respect to norm $||\cdot||_q$ and where $T(\mathbf{z}) = (r, \omega)$. The spectral measure S is defined on $\Xi = [0, \infty)^d \cap \mathbb{S}$ by

$$S(B) = \nu_*(\{\mathbf{z} \in [0, \infty)^d : ||\mathbf{z}||_p \geq 1, \mathbf{z}/||\mathbf{z}||_q \in B\}), \tag{3.2.9}$$

for Borel subsets B of Ξ. Hence, it measures the rescaled probability that joint observations occur in a region B, given that the norm (or, more intuitively, the sum in case the L_1-norm is taken) of the observations if large. Note that the definition is based on the condition that $||\mathbf{z}||_p \geq 1$, which does not necessarily correspond to a "large" norm. However, by the positive homogeneity of the exponent measure, this condition can be adjusted to an appropriate large value as follows,

$$\nu_*(\{\mathbf{z} \in [0, \infty)^d : ||\mathbf{z}||_p \geq r, \mathbf{z}/||\mathbf{z}||_q \in B) = r^{-1} S(B).$$

The simple exponent measure therefore only depends nontrivially on the spectral measure S. Note that the specification of the norm determines which events are considered to be extreme (see Figure 3.2.3). Selecting different norms will lead to different representations but theoretically it does not matter what norms are chosen. However, when choosing a threshold in the process of statistical inference, it turns out that selecting a specific norm for the radius influences the sample that is selected which consequently influences the resulting estimates (Einmahl and Segers, 2009). The norms should be selected based on the goal of the analysis.

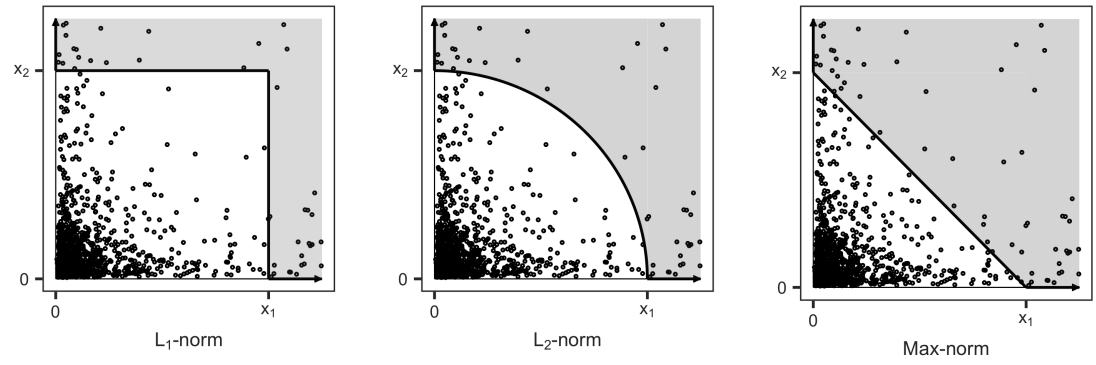

Figure 3.2.3: Extreme regions for the spectral measure for different norms.

Intuitively, the spectral measure considers the contribution of each of the d components to their norm ($\mathbf{z}/||\mathbf{z}||$), conditionally on the norm being large ($||\mathbf{z}|| \geq r$) (Kiriliouk et al., 2014). An example for the bivariate case is shown in Figure 3.2.4. The left-hand picture shows the points for which the sum of the two variables X_1 and X_2 is large for the L_1-norm. The middle panel shows the relationship between the contribution of the variable X_1 to the sum and the size of the sum. The right-hand figure shows the distribution of the contribution of X_1 to the sum $S = X_1 + X_2$. Strong extremal dependence leads to the ratio X_1/S being close to 0.5 whereas low extremal dependence leads to the ratio being close to either 0 or 1. Hence, the distribution of the ratios of the components to their sum, given that the sum is large, carries information about the dependence between large values. This distribution is called the angular or spectral measure.

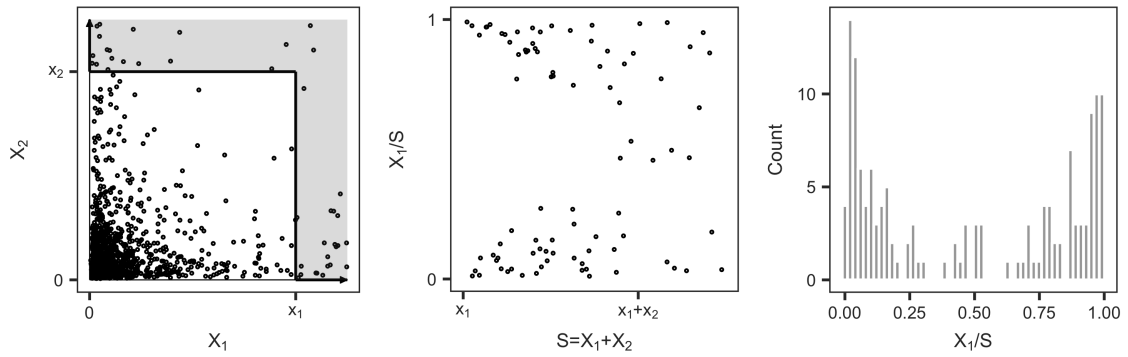

Figure 3.2.4: Spectral measure intuition.

The simple MEVD can be characterized in terms of the spectral measure as follows,

$$
\begin{aligned}
-\log G_*(\mathbf{z}) &= \nu_*([0,\infty]^d \backslash [\mathbf{0},\mathbf{z}]) \\
&= \int_{[0,\infty)^d \backslash \{\mathbf{0}\}} \mathbb{1}\{\mathbf{y} \in [0,\infty]^d \backslash [\mathbf{0},\mathbf{z}]\} \nu_*(\mathrm{d}\mathbf{y}) \\
&= \int_{[0,\infty)^d \backslash \{\mathbf{0}\}} \mathbb{1}\left\{ \max_{j=1,\dots,d} \frac{y_j}{z_j} > 1 \right\} \nu_*(\mathrm{d}\mathbf{y}) \\
&= \int_{\Xi} \int_0^{\infty} \mathbb{1}\left\{ \max_{j=1,\dots,d} \left(\frac{\omega_j}{||\omega||_p} \frac{r}{z_j} \right) > 1 \right\} \frac{1}{r^2} \mathrm{d}r S(\mathrm{d}\omega) \\
&= \int_{\Xi} \int_0^{\infty} \mathbb{1}\left\{ r > 1/\left(\max_{j=1,\dots,d} \left(\frac{\omega_j}{||\omega||_p} \frac{1}{z_j} \right) \right) \right\} \frac{1}{r^2} \mathrm{d}r S(\mathrm{d}\omega) \\
&= \int_{\Xi} \left[-\frac{1}{r} \right]_{1/\left(\max_{j=1,\dots,d} \left(\frac{\omega_j}{||\omega||_p} \frac{1}{z_j} \right) \right)}^{\infty} S(\mathrm{d}\omega) \\
&= \int_{\Xi} \max_{j=1,\dots,d} \left(\frac{\omega_j}{||\omega||_p} \frac{1}{z_j} \right) S(\mathrm{d}\omega),
\end{aligned}
$$

where $\mathbb{1}_A(x)$ is the indicator function which is equal to 1 if $x \in A$ and equal to 0 otherwise. Hence,

$$G_*(\mathbf{z}) = \exp\left(-\int_\Xi \max_{j=1,\ldots,d}\left(\frac{\omega_j}{||\omega||_p}\frac{1}{z_j}\right) S(\mathrm{d}\omega)\right), \tag{3.2.10}$$

and for the original multivariate extreme value distribution,

$$G(\mathbf{x}) = \exp\left(\int_\Xi \min_{j=1,\ldots,d}\left(\frac{\omega_j}{||\omega||_p}\log G_j(x_j)\right) S(\mathrm{d}\omega)\right). \tag{3.2.11}$$

Note that in case the two norms $||\cdot||_p$ and $||\cdot||_q$ are equal, the formulas can be simplified slightly, since $||\omega||_p = 1$ for $\omega \in \Xi$. Furthermore, since the margins of G_* are unit Fréchet, i.e., $G_{*j}(z) = \exp(-1/z)$, it follows that

$$-\log G_{*j}(z_j) = -\log G_*(\infty,\ldots,\infty,z_j,\infty,\ldots,\infty) = \int_\Xi \frac{\omega_j}{||\omega||_p}\frac{1}{z_j} S(\mathrm{d}\omega) = 1/z_j,$$

which leads to the condition

$$\int_\Xi \frac{\omega_j}{||\omega||_p} S(\mathrm{d}\omega) = 1 \quad \forall j \in \{1,\ldots,d\}. \tag{3.2.12}$$

Conversely, any positive measure S on Ξ satisfying these conditions is the spectral measure of the d-dimensional extreme value distribution $G_* = \exp(-\nu_*)$. The results are summarized in the following theorem.

Theorem 3.4. *G is a d-dimensional multivariate extreme value distribution if and only if there exists a spectral measure S on Ξ such that,*

$$G(\mathbf{x}) = \exp\left(\int_\Xi \min_{j=1,\ldots,d}\left(\frac{\omega_j}{||\omega||_p}\log G_j(x_j)\right) S(\mathrm{d}\omega)\right)$$

with constraints

$$\int_\Xi \frac{\omega_j}{||\omega||_p} S(\mathrm{d}\omega) = 1 \quad \forall j \in \{1,\ldots,d\}.$$

As an example, the bivariate case with both norms equal to the Euclidean L_2-norm is considered. This leads to the classical polar coordinates situation and was considered by De Haan and Resnick (1977), for example.

Example 3.3. *Consider two random variables X and Y with bivariate extreme value distribution function G to be evaluated in the point $\mathbf{x} = (x, y) \in \mathbb{R}^2$. When both norms for the polar coordinates transformation are taken to be the L_2-norm, the polar coordinates are defined by*

$$r = \sqrt{x^2 + y^2}, \quad \omega = (x/r, y/r)\,.$$

The spectral measure can be defined on Borel subsets of part of the unit circle,

$$B \subseteq \Xi = \mathbb{R}^2_+ \cap \mathbb{S} = \{(x,y) \in \mathbb{R}^2_+ : \sqrt{x^2+y^2} = 1\}$$

as,

$$S(B) = \nu_*(\{(x,y) \in \mathbb{R}^2_+ : r \geq 1, \omega \in B\}),$$

or, equivalently, for $\theta \in [0, \pi/2]$,

$$S(\theta) = \nu_*\left(\left\{(x,y) \in \mathbb{R} : \sqrt{x^2+y^2} \geq 1, y/x \leq \tan\theta\right\}\right).$$

Putting $x = r\cos\theta$ and $y = r\sin\theta$, we find that $\omega = (\cos\theta, \sin\theta)$ and therefore,

$$\begin{aligned} G_*(z_1, z_2) &= \exp\left(-\int_\Xi \max\left(\frac{\omega_1}{z_1}, \frac{\omega_2}{z_2}\right) S(\mathrm{d}\omega)\right) \\ &= \exp\left(-\int_0^{\pi/2} \max\left(\frac{\cos\theta}{z_1}, \frac{\sin\theta}{z_2}\right) S(\mathrm{d}\theta)\right), \end{aligned}$$

with,

$$\int_0^{\pi/2} \cos\theta S(\mathrm{d}\theta) = \int_0^{\pi/2} \sin\theta S(\mathrm{d}\theta) = 1.$$

Note that X and Y are tail independent if and only if the spectral measure S is concentrated on $\{0, \pi/2\}$. In that case, $S(\{0\}) = 1 = S(\{\pi/2\})$. Similarly, the variables X and Y are completely tail dependent if S is concentrated on $\{\pi/4\}$. In that case, $S(\{\pi/4\}) = ||(1,1)||$.

3.3 Copula approach

So far, the considered characterizations of the tail dependence structure between $d \geq 2$ random variables contained information on both the marginal and the joint behavior of the tails of these variables. The notion that the characterizations also reflect the marginal behavior is one of the factors that make them quite cumbersome to work with, especially when the primary interest is to evaluate tail dependence. In this section, we disentangle the tail dependence structure from the marginal distributions employing a copula-based approach. The copula of the multivariate extreme value distribution turns out to be a special type, being an extreme value copula. Generally, we denote the copula of the distribution F that belongs to the domain of attraction of a MEVD by C_F and the extreme value copula of the MEVD by C. We present several results for extreme value copulas and consider the asymptotic behavior of several commonly used copulas. See Gudendorf and Segers (2009) for a comprehensive overview of results regarding extreme value copulas.

3.3.1 Extreme value copulas

In Chapter 2.1 it was discussed that for each joint distribution function F of a vector of continuous random variables $X_1, \ldots, X_d$, there exists a copula function that both fully and solely captures the dependence structure of the random vector. For the multivariate extreme value distribution G the theorem of Sklar (1959) therefore asserts that G can be represented by a copula C as

$$G(x_1, \ldots, x_d) = C\left(G_1(x_1), \ldots, G_d(x_d)\right), \quad (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{3.3.1}$$

Since the joint and marginal distribution functions of $\mathbf{M_n}$ are given by F^n and $F_1^n, \ldots, F_d^n$ respectively, it follows that the copula C_n of the maxima $\mathbf{M_n}$ is related to the copula C_F as

$$C_n(u_1, \ldots, u_d) = C_F(u_1^{1/n}, \ldots, u_d^{1/n})^n, \quad (u_1, \ldots, u_d) \in [0,1]^d. \tag{3.3.2}$$

Note that copulas are invariant under monotone transformations and that the copula joining $\mathbf{M_n}$ is therefore the same as the copula joining the normalized maxima. The limit of the copulas C_n as $n \to \infty$ is given by C, an extreme value copula (Galambos, 1978).

Definition 3.5. *A copula C is called an extreme value copula if there exists a copula C_F such that*

$$C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)^n \to C(u_1, \ldots, u_d) \quad (u_1, \ldots, u_d) \in [0,1]^d \tag{3.3.3}$$

as $n \to \infty$. The copula C_F is said to be in the domain of attraction of C ($C_F \in MDA(C)$).

The maximum domain of attraction condition on the joint distribution function (Equation 3.1.5) implies the copula convergence (Equation 3.3.3), but is not reciprocal since the marginal distributions also have to converge to univariate extreme value distributions. The copula convergence condition together with the convergence of the margins to univariate EVDs implies the convergence of a multivariate distribution to a multivariate extreme value distribution. Although marginal convergences are required to achieve a proper MEVD, the extreme value copula C is solely determined by the copula C_F. The marginal distributions $F_1, \ldots, F_d$ influence the marginal distributions of the limiting MEVD but not the dependence structure.

Similar to the MEVD, another representation of the extreme value copula can be given with the concept of max-stability (see Proposition 3.1). For copulas, the max-stability is defined as follows (Leadbetter and Rootzén, 1988).

Definition 3.6. *A d-variate copula C is max-stable if it satisfies the relationship*

$$C(u_1, \ldots, u_d) = C\left(u_1^{1/m}, \ldots, u_d^{1/m}\right)^m \tag{3.3.4}$$

for every integer $m \geq 1$ and all $(u_1, \ldots, u_d) \in [0,1]^d$.

It follows immediately from this definition that an extreme value copula is max-stable and that a max-stable copula is in its own domain of attraction and must, therefore, be an extreme value copula. That is, a copula is an extreme value copula if and only if it is max-stable. The observation that the class of multivariate extreme value distributions coincides with the class of max-stable distributions is researched by De Haan and Resnick (1977). The fact that C is max-stable implies that $C^{1/k}$ is a distribution function for any integer k, that is, C is max-infinitely divisible (Balkema and Resnick, 1977).

The extreme value copula can be retrieved based on the MEVD but also based on the exponent measure or the spectral measure since these characterizations describe the multivariate dependence structure of the MEVD. Hence, the extreme value copula can be completely characterized by the exponent measure, as shown by the following relationship,

$$C\left(G_1(x_1), \ldots, G_d(x_d)\right) = G(\mathbf{x}) = \exp\left(-\nu\left([\mathbf{0}, \mathbf{x}]^c\right)\right),$$

or, setting $G_j(x_j) = u_j$ for $1 \leq j \leq d$,

$$C(u_1, \ldots, u_d) = \exp\left(-\nu\left(\left[\mathbf{0}, \left(G_1^{-1}(u_1), \ldots, G_d^{-1}(u_d)\right)\right]^c\right)\right).$$

The spectral measure representation can also be used to characterize extreme value copulas. Since

$$C(G_1(x_1), \ldots, G_d(x_d)) = G(\mathbf{x}) = \exp\left(\int_{\Xi} \min_{j=1,\ldots,d}\left(\frac{\omega_j}{||\omega||_1} \log G_j(x_j)\right) S(\mathrm{d}\omega)\right),$$

it follows that

$$C(u_1, \ldots, u_d) = \exp\left(\int_{\Xi} \min_{j=1,\ldots,d}\left(\frac{\omega_j}{||\omega||_1} \log u_j\right) S(\mathrm{d}\omega)\right). \tag{3.3.5}$$

Using the definition of the extreme value copula, it is now possible to investigate the maximum domain of attraction to which several known copulas belong. First, using the copula convergence, it is easy to see that the independence copula is in its own maximum domain of attraction.

Example 3.4. *The independence copula is given by $C_F(u_1, \ldots, u_d) = u_1 \cdots u_d$. By the copula convergence condition (Equation 3.3.3), it follows that*

$$\lim_{n\to\infty} C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)^n = \lim_{n\to\infty}\left(u_1^{1/n} \cdots u_d^{1/n}\right)^n = u_1 \cdots u_d = C(u_1, \ldots, u_d).$$

Hence, if two or more variables are joint by the independence copula, their maxima are also independent. Similarly, it can be shown that the positive comonotonicity copula (i.e., the upper Fréchet bound) is also in its own maximum domain of attraction.

Example 3.5. *The comonotonicity copula is given by $C_F(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d)$. By the copula convergence condition (Equation 3.3.3), it follows that*

$$\lim_{n\to\infty} C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)^n = \lim_{n\to\infty} \min\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)^n = \min(u_1, \ldots, u_d) = C(u_1, \ldots, u_d).$$

Hence, if two or more variables are joint by the comonotonicity copula, then this is also the case for their maxima. Furthermore, these results show that the independence copula and the comonotonicity copula are both extreme value copulas. A somewhat more complicated example of a copula that belongs to its own maximum domain of attraction is the Gumbel copula. In extreme value theory, the Gumbel copula is also known as the Gumbel-Hougaard copula or the logistic copula.

Example 3.6. *The Gumbel copula is the Archimedean copula defined by the generator* $\psi(t) = \exp(-t^{1/\theta})$ *(see Section 2.3). Its copula function is given by*

$$C_F(u_1,\ldots,u_d) = \exp\left(-\left[\sum_{i=1}^{d}(-\log u_i)^\theta\right]^{1/\theta}\right), \quad (u_1,\ldots,u_d)\in[0,1]^d,$$

with $\theta \geq 1$. *By the copula convergence condition (Equation 3.3.3), it follows that*

$$\begin{aligned}
\lim_{n\to\infty} C_F\left(u_1^{1/n},\ldots,u_d^{1/n}\right)^n &= \lim_{n\to\infty}\exp\left(-\left[\left(-\log u_1^{1/n}\right)^\theta+\cdots+\left(-\log u_d^{1/n}\right)^\theta\right]^{1/\theta}\right)^n \\
&= \lim_{n\to\infty}\exp\left(-n\left[\left(-\frac{1}{n}\log u_1\right)^\theta+\cdots+\left(-\frac{1}{n}\log u_d\right)^\theta\right]^{1/\theta}\right) \\
&= \exp\left(-\left[(-\log u_1)^\theta+\cdots+(-\log u_d)^\theta\right]^{1/\theta}\right) \\
&= C(u_1,\ldots,u_d).
\end{aligned}$$

Again, this indicates that if two or more variables are joint by the Gumbel copula, their maxima are also following the Gumbel copula. Furthermore, this example shows that the Gumbel copula is an extreme value copula. It can be shown that the Gumbel copula is the only copula that is both an Archimedean and an extreme value copula (Genest and Rivest, 1989). Moreover, for Archimedean copulas in general, it can be shown that if the copula C_F is Archimedean, it always belongs to the maximum domain of attraction of either the Gumbel copula, the comonotonicity copula, or the independence copula. For the lower tail a similar result holds with the Galambos copula instead of the Gumbel copula. For copulas that are not extreme value copulas and, hence, convergence to a different asymptotic distribution, the max-stability convergence definition (Equation 3.3.3) is difficult to employ to identify the limiting copula distribution. Other methods have therefore been developed. For Archimedean copulas, the behavior of the generator is usually analyzed to assess the tail dependence structure. However, the stochastic representation also provides an intuitive approach to interpret the tail dependence structure of Archimedean copulas. This is illustrated by the following theorem.

Theorem 3.5. *(Proposition 2 in Larsson and Nešlehová (2011)). Let* $\mathbf{X} = (X_1,\ldots,X_d)$ *be a* d*-dimensional random vector joint by an Archimedean copula* C_F *with* d*-monotone generator function* ψ. *If* $1-\psi(1/x)\in RV_{-\alpha}$, *for* $\alpha\in(0,1]$, *then* $C_F \in MDA(C^{Gu}_{1/\alpha})$, *where* C^{Gu}_θ *is the Gumbel copula with parameter* $\theta \geq 1$.

Proof. (Based on the proof shown in Larsson and Nešlehová (2011)). Since C_F is a d-dimensional Archimedean copula with generator ψ, the stochastic representation of McNeil and Nešlehová (2009) implies that C_F is the survival copula of a random vector $\mathbf{X} = (X_1,\ldots,X_d)$ that can be stochastically represented as $\mathbf{X} = R\mathbf{S}$ with $\mathbf{S}$ a random vector that is uniformly distributed on the unit simplex Δ_d and R a nonnegative random variable independent of $\mathbf{S}$ with distribution function equal to inverse Williamson d-transform of ψ (see Section 2.3). Since the survival function of each component of $\mathbf{X}$ is equal to ψ, the survival function of the random vector $\mathbf{X}$ is given by

$$\begin{aligned}
\mathbb{P}(X_1>x_1,\ldots,X_d>x_d) &= \overline{F}_{\mathbf{X}}(x_1,\ldots,x_d) \\
&= C_F\left(\overline{F}_1(x_1),\ldots,\overline{F}_d(x_d)\right) \\
&= C_F(\psi(x_1),\ldots,\psi(x_d)) = \psi(x_1+\cdots+x_d) = \psi(||\mathbf{x}||_1).
\end{aligned}$$

Now consider n iid replicates $\mathbf{X}_i = (X_{i1},\ldots,X_{id})$ of $\mathbf{X}$, $1\leq i\leq n$, and define $\mathbf{W}_n = (W_{n1},\ldots,W_{nd})$ to be the vector of componentwise minima, i.e.,

$$W_{nj} = \min(X_{1j},\ldots,X_{nj}), \quad 1\leq j\leq d.$$

The survival function of $\mathbf{W}$ is given by

$$\mathbb{P}\left(W_{n1}>x_1,\ldots,W_{nd}>x_d\right) = \prod_{i=1}^{n}\mathbb{P}\left(X_{i1}>x_1,\ldots,X_{id}>x_d\right) = \psi^n(||\mathbf{x}||_1).$$

In order to retrieve the distribution of the maxima, we now consider $-\mathbf{X}$ with the componentwise maxima given by $-\mathbf{W}_n$. Note that the margins of $-\mathbf{X}$ are given by $F_i(x_i) = \psi(-x_i)$, $1 \leq i \leq d$ since $\mathbb{P}(-X_i \leq x_i) = \mathbb{P}(X_i > -x_i)$. Because it is assumed that $1 - \psi(1/x) \in RV_{-\alpha}$, it follows that the margins of $-\mathbf{X}$ belong to the Weibull domain of attraction, i.e., there exists a sequence $(b_n) > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(-\frac{W_{n1}}{b_n} \leq x\right) = \lim_{n\to\infty} \psi^n(-b_n x) = \exp\left(-(-x)^{\alpha}\right),$$

for $x < 0$. This implies that $\psi^n(-b_n||\mathbf{x}||_1)$ converges to $\exp(-(-||\mathbf{x}||_1)^{\alpha})$, which can be recognized as the generator of the Gumbel copula with parameter $1/\alpha$. Hence, $-\mathbf{W}_n/b_n$ converges weakly to $-\mathbf{W}$, where $\mathbf{W}$ is an ℓ_1-norm symmetric random vector with survival function $\psi^{Gu}_{1/\alpha}(-||\mathbf{x}||_1)$ and survival copula $C^{Gu}_{1/\alpha}$. It follows that $C_F \in MDA(C^{Gu}_{1/\alpha})$. ■

Larsson and Nešlehová (2011) show that the assumption of the Theorem 3.5 is satisfied if the distribution of $1/R$ belongs to a maximum domain of attraction. See their paper for more details and the proof of this result. The result can be used to show that the Archimedean Joe copula belongs to the maximum domain of attraction of the Gumbel copula.

Example 3.7. *The Joe copula is the Archimedean copula induced by the generator* $\psi(t) = 1-(1-\exp(-x))^{1/\theta}$ *(see Section 2.3). Since*

$$\lim_{t\to\infty} \frac{1-\psi((tx)^{-1})}{1-\psi(x^{-1})} = \lim_{t\to\infty} \frac{(1-\exp(-(tx)^{-1})^{1/\theta}}{(1-\exp(-t^{-1})^{1/\theta}} = \lim_{t\to\infty} \left(\frac{(tx)^{-1}}{t^{-1}}\right)^{1/\theta} = x^{-1/\theta}$$

where we used that $\exp(-tx) \sim 1 - tx$ *for* $t \to 0$*, Theorem 3.5 implies that the Joe copula with parameter* θ *belongs to the maximum domain of attraction of the Gumbel copula with parameter* θ*.*

The result can be used in a similar fashion to show that both the Frank and the Clayton copula belong to the maximum domain of attraction of the independence copula. Several generators are shown in Figure 3.3.1. The figure shows that the generator function of the Frank copula approaches zero somewhat faster than the generator function of the Gumbel copula. This causes the Frank copula to converge to Π, whereas the Gumbel copula is tail dependent.

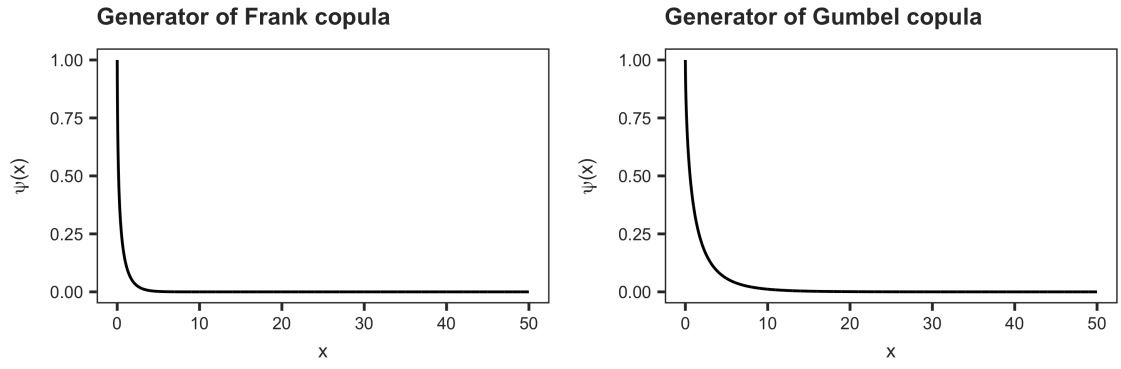

Figure 3.3.1: Generators for the Frank ($\theta = 3.45$) and Gumbel copula ($\theta = 1.54$). The model parameters lead to a bivariate rank correlation of $\rho = 0.5$ for both copulas.

For elliptical distributions, Schmidt (2002) and Hult and Lindskog (2002) show that the behavior of the tail of the generating variable determines the extremal dependence structure between the random variables. This presents a clear parallel to the theory for Archimedean copulas. The result is captured by the following theorem.

Theorem 3.6. *(Theorem 2.4 in Schmidt (2002), Theorem 4.3 in Hult and Lindskog (2002)). Let* $\mathbf{X} = (X_1, \ldots, X_d)$ *be a d-dimensional random vector following an elliptical distribution, with stochastic representation* $\mathbf{X} \stackrel{d}{=} R\mathbf{U}$ *(see Section 2.2) and denote with* F_R *the distribution function of the generating variable R. If* F_R *has a regularly varying tail, then all margins of the distribution of* $\mathbf{X}$ *are tail dependent.*

Proof. (Hult and Lindskog, 2002). The proof is based on the characterization of tail dependence based on multivariate regular variation (see Appendix A). We give an outline of the proof here. Note that since the d-dimensional random vector $\mathbf{U}$ is uniformly distributed on the unit sphere in $\mathbb{R}^d$ and since $\mathbf{U}$ is distributed independently of R (Section 2.2), the following equivalences hold,

$$\frac{\mathbb{P}\left(|\mathbf{X}| > tx, \mathbf{X}/|\mathbf{X}| \in \cdot\right)}{\mathbb{P}\left(|\mathbf{X}| > t\right)} = \frac{\mathbb{P}\left(R > tx, \mathbf{AU} \in \cdot\right)}{\mathbb{P}\left(R > t\right)} = \frac{\mathbb{P}\left(R > tx\right)\mathbb{P}\left(\mathbf{AU} \in \cdot\right)}{\mathbb{P}\left(R > t\right)}.$$

Furthermore, recall that if R is regularly varying, then by definition,

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(R > tx\right)}{\mathbb{P}\left(R > t\right)} = x^{-\alpha},$$

with $\alpha > 0$ (see Appendix A). Hence,

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(|\mathbf{X}| > tx, \mathbf{X}/|\mathbf{X}| \in \cdot\right)}{\mathbb{P}\left(|\mathbf{X}| > t\right)} = \lim_{t\to\infty} \frac{\mathbb{P}(R > tx)\mathbb{P}(\mathbf{AU} \in \cdot)}{\mathbb{P}(R > t)} = x^{-\alpha}\mathbb{P}\left(\mathbf{AU} \in \cdot\right),$$

which implies that $\mathbf{X}$ is multivariate regularly varying with index $\alpha > 0$. For more detailed computations see Hult and Lindskog (2002). ■

This result can be used to explore the maximum domain of attraction of both the t-copula and the Gaussian copula. Recall from Section 2.2 that for the Normal copula, the distribution of the generating variable is defined by $R^2 \sim \chi^2_d$, whereas the distribution of the generating variable for the t-copula is defined by $R^2/d \sim F(d, \nu)$. The densities of both resulting distributions for R are shown in Figure 3.3.2. For the t-copula, the density can be derived as follows: if $Y := R^2/d \sim F(d,\nu) =: F_Y$ then $R = \sqrt{dY}$. Hence, $F_R(x) = \mathbb{P}(R \leq x) = \mathbb{P}(\sqrt{dY} \leq x) = \mathbb{P}(Y \leq \frac{1}{d}x^2) = F_Y(\frac{1}{d}x^2)$. Therefore, we find that $f_R(x) = \frac{\mathrm{d}}{\mathrm{d}x}F_R(x) = \frac{\mathrm{d}}{\mathrm{d}x}F_Y(\frac{1}{d}x^2) = \frac{2}{d}xf_Y(\frac{1}{d}x^2)$, where f_Y is the density function of the $F(d,\nu)$ distribution. The density for the generating variable of the Normal copula can be determined similarly: if f_Y is the density function of the χ^2_d distribution, then the generating variable of the Normal copula follows the following density, $f_R(x) = 2xf_Y(x^2)$.

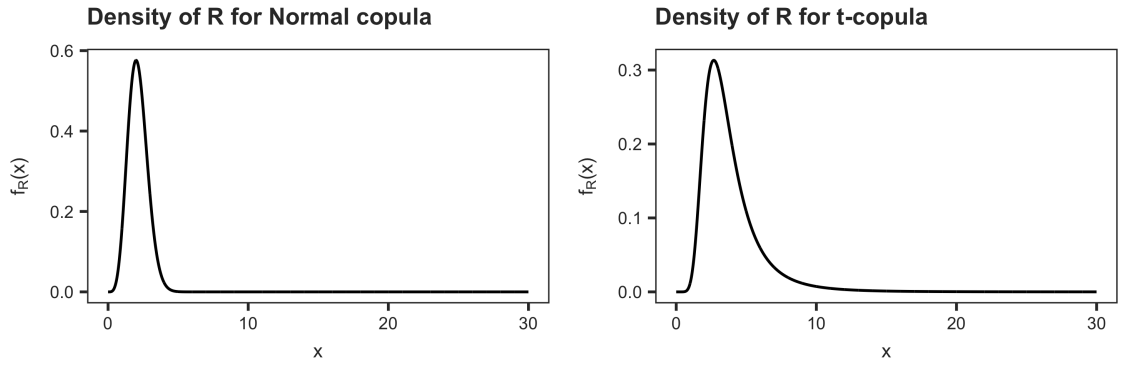

Figure 3.3.2: Densities of the generating variable for the Normal and t-copula, with the degrees of freedom equal to $d = 10$ and $\nu = 4$.

By Karamata's theorem (see Appendix A), the distribution of the generating variable R has a regularly varying tail if the density has a regularly varying tail. For the t-copula, we find that

$$\lim_{t\to\infty} \frac{f_R(tx)}{f_R(t)} = \lim_{t\to\infty} \frac{\frac{2}{d}txf_{F(d,\nu)}(\frac{1}{d}(tx)^2)}{\frac{2}{d}tf_{F(d,\nu)}(\frac{1}{d}t^2)} = \lim_{t\to\infty} x\frac{f_{F(d,\nu)}(\frac{1}{d}(tx)^2)}{f_{F(d,\nu)}(\frac{1}{d}t^2)},$$

and since the F-distribution belongs to the domain of attraction of the Fréchet distribution, it can be concluded that the t-copula has a generating variable with a regularly varying tail, and, therefore, has bivariate margins that are tail dependent. In contrast, for the Normal copula, the χ^2 distribution belongs to the domain of attraction of the Gumbel distribution, hence the generating variable does not have a regularly varying tail. This leads to the conclusion that the Gaussian copula belongs to the MDA of the independence copula Π, a result that has actually already been known since Sibuya (1960). Although the Normal distribution converges to Π, a

better approximation to the extreme value copula can be given in the bivariate case when the correlation parameter ρ is close to 1. This result is due to Hüsler and Reiss (1989) who show that under the assumption that the correlation coefficient ρ can vary with the sample size n, $\rho = \rho_n$, and does so in such a way that $\rho_n \to 1$ as $n \to \infty$, the bivariate Gaussian copula converges to the Hüsler-Reiss extreme value copula. This copula will be further assessed based on the dependence functions introduced in the Section 3.4.

3.3.2 Copula tail densities

Analogue to the tail density for the multivariate extreme value distribution introduced in Section 3.1.2, it is possible to consider the tail density of an extreme value copula. The copula tail density can be an useful tool to analyse the extremal dependence structure of distributions that are (most easily) defined by their density, such as the t-copula and vine copulas. The concept of copula tail densities has been pioneered by Li (2013) and further developed in Li and Hua (2015). Defining the partial differentiation operator as $D_w = \frac{\partial^d}{\partial w_1 \dots \partial w_d}$, and recalling from Section 2.1 that

$$\overline{C}(1 - uw_i, 1 \leq i \leq d) = 1 - \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} (-1)^{|S|-1} C_S(1 - uw_i, i \in S),$$

it follows that

$$D_w \overline{C}(1 - uw_i, 1 \leq i \leq d) = (-1)^d D_w C(1 - uw_i, 1 \leq i \leq d) = u^d c(1 - uw_i, 1 \leq i \leq d).$$

Li (2013) then defines the copula tail density as follows.

Definition 3.7. *Let* $\mathbf{X}$ *be a* d*-dimensional random vector with joint distribution function* F *and copula* C_F*. Denoting by* c_F *the copula density and by* $\overline{C}_F$ *the survival function of* C_F*, the copula tail density* $\Upsilon : \mathbb{R}_+^d \to \overline{\mathbb{R}}_+$ *is defined as*

$$\Upsilon(\mathbf{w}) = \lim_{u \to 0} u^{d-1} c_F(1 - uw_i, 1 \leq i \leq d) = \lim_{u \to 0} \frac{D_w \overline{C}_F(1 - uw_i, 1 \leq i \leq d)}{u}, \tag{3.3.6}$$

if the limits exist.

To ensure that the limit in the definition holds, uniform convergence on $\mathbb{R}_+^d \backslash \{0\}$ of the partial derivatives of the tail dependence functions as $u \to 0$ is required.

It is clear that if the copula density c_F is independent and therefore equal to 1, the tail density is zero on its entire domain. Unfortunately, the opposite is not true. That is, a tail density that is zero everywhere on its domain does not imply tail independence. This can be seen from the following example.

Example 3.8. *Consider a trivariate* D*-vine copula, where* c_{12} *is independent,* c_{23} *is the density of a bivariate t-copula with parameters* $\rho > 0$ *and* $\nu < \infty$*, and therefore tail dependent, and where the conditional copula density* $c_{13|2} = 1$*, corresponding to independence (note that this results in a so-called Markov tree). The density of the vine copula is given by*

$$c_F(u_1, u_2, u_3) = c_{12}(u_1, u_2) c_{23}(u_2, u_3) c_{13|2}\left(C_{1|2}(u_1|u_2), C_{3|2}(u_3|u_2)\right) = c_{23}(u_2, u_3).$$

Since the density of the vine model is equal to the density of the bivariate t-copula, the extreme value copula of the vine is given by the bivariate t-EV copula. Hence, the vine copula exhibits tail dependence. However, the tail density of the vine copula is equal to zero,

$$\Upsilon(w_1, w_2, w_3) = \lim_{u \to 0} u^2 c_{23}(1 - uw_2, 1 - uw_3) = \lim_{u \downarrow 0} u \Upsilon_{23}(w_2, w_3) = 0,$$

where Υ_{23} *is the tail density of the t-copula, defined in Proposition 3.2 below.*

Marginal tail densities can be retrieved from the full tail density by integrating out the appropriate variables. This is illustrated by the following example.

Example 3.9. *Suppose Υ_{123} is the trivariate copula tail density of the random vector $\mathbf{X} = (X_1, X_2, X_3)$ with copula density c_{123}. The marginal copula tail density with respect to the variables X_1 and X_2, denoted by Υ_{12}, can be retrieved from the full copula tail density Υ_{123} by integrating out the third variable as follows,*

$$\begin{aligned}
\Upsilon_{12} &= \lim_{u\to 0} uc_{12}(1-uw_1, 1-uw_2) \\
&= \lim_{u\to 0} u \int_0^1 c_{123}(1-uw_1, 1-uw_2, v_3)\mathrm{d}v_3 \\
&= \lim_{u\to 0} u \int_{1/u}^0 -uc_{123}(1-uw_1, 1-uw_2, 1-uw_3)\mathrm{d}w_3 \\
&= \int_0^\infty \lim_{u\to 0} u^2 c_{123}(1-uw_1, 1-uw_2, 1-uw_3)\mathrm{d}w_3 \\
&= \int_0^\infty \Upsilon_{123}(w_1, w_2, w_3)\mathrm{d}w_3,
\end{aligned}$$

where uniform convergence has to be assumed in order to be able to exchange limit and integral in the fourth step. Obviously, the same procedure can be employed to retrieve marginal tail densities for other subsets of variables and in higher dimensions.

The tail density for the t-copula is presented in Proposition 3.2.

Proposition 3.2. *(Li and Wu, 2013). The tail density function Υ for the t-copula is given by*

$$\Upsilon(\mathbf{w}) = |\Sigma|^{-\frac{1}{2}} \nu^{1-d} \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)\pi^{(d-1)/2}} \frac{\left[\left(\mathbf{w}^{-\frac{1}{\nu}}\right)^T \mathbf{\Sigma}^{-1}\mathbf{w}^{-\frac{1}{\nu}}\right]^{-\frac{\nu+d}{2}}}{\prod_{i=1}^d w_i^{\frac{\nu+1}{\nu}}}, \tag{3.3.7}$$

for $\mathbf{w} = (w_1, \ldots, w_d) \in \mathbb{R}^d \backslash \{0\}$. If $w_j = 0$ for $j \in \{1, \ldots, d\}$, then $\Upsilon(\mathbf{w}) = 0$.

Proof. (Li and Wu, 2013). Consider a d-dimensional symmetric t-distribution $t_d(\nu, \Sigma)$ with mean 0 and density function given by,

$$f_t(\mathbf{x}; \nu, \mathbf{\Sigma}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{d/2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu}\left(\mathbf{x}^T \mathbf{\Sigma}^{-1}\mathbf{x}\right)\right]^{-\frac{\nu+d}{2}},$$

where $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $\nu > 0$ is the degree of freedom, and $\mathbf{\Sigma} = (\rho_{ij})$ is a $d \times d$ symmetric dispersion matrix. The one dimensional marginal t-distribution has the density

$$f_i(x_i) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{1/2}} \left(1 + \frac{x_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x_i \in \mathbb{R}, 1 \le i \le d.$$

It can be shown that f_i has regularly varying tails with index $nu+1$. This implies that the margin F_i has a regulalry varying tail with $1 - F_i(x_i) \approx \nu^{-1} x_i^{-\nu} \mathscr{L}(x_i, \nu)$ as $x_i \to \infty$, where

$$\mathscr{L}(x_i, \nu) \approx \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \left(\frac{1}{x_i^2} + \frac{1}{\nu}\right)^{-\frac{\nu+1}{2}} \to \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \nu^{\frac{\nu+1}{2}} =: \ell, \quad \text{as } x_i \to \infty.$$

The limiting constant $\ell > 0$ is an explicit constant only dependending on ν. If we set $F_i(x_i) = 1 - uw_i$ then we have $\nu^{-1} x_i^{-\nu} \ell \approx uw_i$ as $u \to 0$. Thus, we obtain the following estimate for sufficiently small u,

$$F_i^{-1}(1 - uw_i) \approx \nu^{-\frac{1}{\nu}} \ell^{\frac{1}{\nu}} (uw_i)^{-\frac{1}{\nu}}, \quad 1 \le i \le d.$$

Plugging this estimate into the density of the t-copula $c(1 - uw_i, 1 \le i \le d)$ gives that as $u \to 0$,

$$\begin{aligned}
c(1 - uw_i, 1 \le i \le d) &= f_t\left(F_1^{-1}(1-uw_1), \ldots, F_d^{-1}(1-uw_d)\right) \prod_{i=1}^d \left[f_i\left(F_i^{-1}(1-uw_i)\right)\right]^{-1} \\
&\approx u^{1-d} |\mathbf{\Sigma}|^{-\frac{1}{2}} \nu^{(1-d)\left(\frac{\nu}{2}+1\right)} \ell^{d-1} \frac{\Gamma\left(\frac{\nu+d}{2}\right)\Gamma^{d-1}\left(\frac{\nu}{2}\right)}{\Gamma^d\left(\frac{\nu+1}{2}\right)} \frac{\left[\left(\mathbf{w}^{-\frac{1}{\nu}}\right)^T \mathbf{\Sigma}^{-1}\mathbf{w}^{-\frac{1}{\nu}}\right]^{-\frac{\nu+d}{2}}}{\prod_{i=1}^d w_i^{\frac{\nu+1}{\nu}}}.
\end{aligned}$$

The result now follows immediately from the definition of the copula tail density. ■

The tail density of a D-vine copula is presented in Theorem 3.7 below. This theorem will turn out to be very useful in analyzing the tail behavior of vines. For example, Proposition 3.3 illustrates that a D-vine always exhibits tail dependence if all baseline linking copulas are tail dependent.

Theorem 3.7. *(Li and Wu, 2013). Suppose C_F is a d-dimensional* D*-vine copula and assume that all bivariate linking copulas $C_{F;ij}$ have continuous densities and satisfy the uniform convergence properties. If the baseline linking copula $C_{F;i,i+1}$ are all upper tail dependent, then*

$$\frac{\Upsilon_{\{1,\ldots,d\}}(\mathbf{w})}{\Upsilon_{\{2,\ldots,d-1\}}(\mathbf{w}_{\{2,\ldots,d-1\}})} = \frac{\Upsilon_{\{1,\ldots,d-1\}}(\mathbf{w}_{\{1,\ldots,d-1\}})}{\Upsilon_{\{2,\ldots,d-1\}}(\mathbf{w}_{\{2,\ldots,d-1\}}} \frac{\Upsilon_{\{2,\ldots,d\}}(\mathbf{w}_{\{2,\ldots,d\}})}{\Upsilon_{\{2,\ldots,d-1\}}(\mathbf{w}_{\{2,\ldots,d-1\}})} \cdot c_{1,d|2,\ldots,d-1}\left(1-t_{1|2,\ldots,d-1}\left(w_1|\mathbf{w}_{\{2,\ldots,d-1\}}\right), 1-t_{d|2,\ldots,d-1}\left(w_d|\mathbf{w}_{\{2,\ldots,d-1\}}\right)\right),$$

where for $j \in \{1,\ldots,d\}$, and $S \subseteq \{1,\ldots,d\}\backslash\{j\}$,

$$\begin{aligned}\bar{t}_{j|S}\left(w_j|\mathbf{w}_S\right) &= \lim_{u\downarrow 0} \overline{C}_{j|S}\left(1-uw_j|1-uw_i, i\in S\right)\\ &= \int_0^{w_j} \frac{\Upsilon_{S\cup\{j\}}(\tilde{w}_j,\mathbf{w}_S)}{\Upsilon_S(w_S)}\mathrm{d}\tilde{w}_j.\end{aligned}$$

Proof. The proof is included in Appendix B.1. ■

Proposition 3.3. *Based on the representation of the copula tail density function for a* D*-vine copula, the following results can be deduced:*

1. *If some baseline linking copulas $C_{i,i+1}$ are tail independent (i.e., $\Upsilon_{i,i+1}=0$ for some $1\le i\le d$), then $\Upsilon(\mathbf{w})=0$ (Li and Wu, 2013).*
2. *If all baseline linking copulas $C_{i,i+1}$ are tail dependent (i.e., $\Upsilon_{i,i+1}>0$ for some $1\le i\le d$) and if all linking copulas $C_{\{i,j|i+1,\ldots,j-1\}}$ have a non-zero density everywhere on $[0,1]^2$, then the* D*-vine copula is tail dependent (i.e., $\Upsilon(\mathbf{w})>0$).*

Proof. The proof is included in Appendix B.1. ■

As an example, a 3-dimensional D-vine is considered. Theorem 3.7 implies that the tail density of a trivariate D-vine can be written in terms of the bivariate tail densities of the linking copulas. Specifically,

$$\begin{aligned}\Upsilon_{123}(x_1,x_2,x_3) &= \Upsilon_{12}(x_1,x_2)\Upsilon_{23}(x_2,x_3)c_{1,3|2}\left(1-\overline{C}_{F;1|2}(x_1|x_2), 1-\overline{C}_{F;3|2}(x_3|x_2)\right)\\ &= \Upsilon_{12}(x_1,x_2)\Upsilon_{23}(x_2,x_3)c_{13|2}\left(1-\int_0^{x_1}\Upsilon_{12}(w_1,x_2)\mathrm{d}w_1, 1-\int_0^{x_1}\Upsilon_{23}(x_2,w_3)\mathrm{d}w_3\right).\end{aligned}$$

3.4 Dependence functions

This section presents three dependence functions that originate from the extreme value copula. Specifically, the stable tail dependence function, the Pickands dependence function, and the tail copula are discussed. These functions are sometimes known by different names in literature, but these are most commonly used. Given certain conditions, each of these functions can convey the same amount of information on the dependence structure between multivariate extremes as the extreme value copula. However, subtle differences in the properties, evaluation methods, and interpretation of these dependence functions can provide advantages over the original extreme value copula.

3.4.1 Stable tail dependence function

Since the MDA-convergence-based definition of the extreme value copula (Equation 3.3.3) can be difficult to employ for some copulas, another general expression that shows the connection between the copula C_F and the corresponding extreme value copula C is derived. Taking the logarithm on both sides of Equation 3.3.3 and using that $\log(x) \sim x - 1$ for $x \to 1$, it can be derived that

$$\begin{aligned}\lim_{n\to\infty} \log C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)^n &= \lim_{n\to\infty} n \log\left(C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)\right) \\ &= -\lim_{n\to\infty} n\left(1 - C_F\left(u_1^{1/n}, \ldots, u_d^{1/n}\right)\right) \\ &= -\lim_{t\downarrow 0} \frac{1 - C_F(u_1^t, \ldots, u_d^t)}{t} \qquad = \log C(u_1 \ldots, u_d),\end{aligned}$$

for all $(u_1, \ldots, u_d) \in [0,1]^d$. By setting $u_j = \exp(-x_j)$, $j = 1, \ldots, d$, and using the approximation $\exp(-tx) \sim 1 - tx$ for $t \to 0$, the following useful reformulation of the copula convergence condition is given,

$$\lim_{t\downarrow 0} \frac{1 - C_F(1 - tx_1, \ldots, 1 - tx_d)}{t} = -\log C\left(e^{-x_1}, \ldots, e^{-x_d}\right). \tag{3.4.1}$$

Segers (2012b) notes that this way, the copula domain-of-attraction condition, originally involving the vector of componentwise sample maxima, is replaced by a condition on the upper tail of a single random vector. This is akin to switching from the block maxima approach to the peaks-over-threshold approach. This relationship between the copula C_F and its extreme value copula C is the basis for the definition of the so-called stable tail dependence function (STDF). This dependence function has its roots in the work of Huang (1992) and Drees and Huang (1998). Originally, the STDF was defined based on the simple extreme value distribution G_*, i.e., the multivariate extreme value distribution with margins transformed to unit Fréchet distributions. We define the STDF based on the copula convergence in Equation 3.4.1.

Definition 3.8. *Let $X_1, \ldots, X_d$ be random variables with marginal distributions $F_1, \ldots, F_d$, joint distribution function F, and copula C_F. The stable tail dependence function $\ell : \mathbb{R}_+^d \to \mathbb{R}_+$ is defined for $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}_+^d$ as*

$$\ell(x_1, \ldots, x_d) = \lim_{t\downarrow 0} \frac{1 - C_F(1 - tx_1, \ldots, 1 - tx_d)}{t}, \tag{3.4.2}$$

if the limit exists.

Intuitively, the stable tail dependence function (STDF) concerns the event that at least one among the d components $X_1, \ldots, X_d$ exceeds a high percentile of its own distribution. By letting this high threshold go to its limit, the extremal dependence structure is captured by the ℓ-function. The coefficients $(x_1, \ldots, x_d)$ can be interpreted as determining the direction from which the limit is approached. The existence of the limit is equivalent to the statement that the joint distribution function of the simple random vector $\mathbf{X}^*$, i.e., the random vector $\mathbf{X}$ with marginals transformed to unit Fréchets, is in the maximum domain of attraction of a d-variate extreme value distribution with unit Fréchet margins (Kiriliouk et al., 2014). See Einmahl et al. (2006) for methods to test this assumption. If the limit exists, the STDF of the extreme value copula is the same as the STDF of the copula C_F. This is captured by the following proposition.

Proposition 3.4. *If the limit in Equation 3.4.2 exists and if $C_F \in MDA(C)$, then the STDF of C_F is equal to the STDF of its extreme value copula C.*

Proof. This follows immediately by Equation 3.4.1 and the fact that an extreme value copula belongs to its own domain of attraction. ■

By this relationship, the extreme value copula can be characterized by the stable tail dependence function as follows.

Proposition 3.5. *The copula C is an extreme value copula if and only if it has the representation*

$$C(u_1, \ldots, u_d) = \exp\left(-\ell(-\log u_1, \ldots, -\log u_d)\right). \tag{3.4.3}$$

Proof. This relationship follows immediately from the definition of the stable tail dependence function and its relationship with the extreme value copula given in Equation 3.4.1. ■

Because of this representation, the STDF is sometimes also called the exponent function which is usually denoted by an a. Relationships with the extreme value distribution, the exponent measure, and the spectral measure are given by the following relationships,

$$\ell(x_1,\ldots,x_d) = -\log G_*(1/x_1,\ldots,1/x_d) = \nu_*\left([0,\infty]^d \backslash [\mathbf{0},\mathbf{1/x}]\right) = \int_\Xi \max_{j=1,\ldots,d}\left(\frac{\omega_i}{||\omega||_a}x_i\right) S(\mathrm{d}\omega). \tag{3.4.4}$$

See Molchanov (2008) and Ressel (2013) for more information on the properties of the STDF. Marginals can be easily retrieved by setting some components x_j to zero, making it easy to retrieve lower-dimensional margins of the extreme value copula. Each stable tail dependence function satisfies four necessary properties.

Theorem 3.8. *Each stable tail dependence function $\ell : \mathbb{R}_+^d \to \mathbb{R}_+$ satisfies the following properties:*

1. *Homogeneity: $\ell(cx_1,\ldots,cx_d) = c\ell(x_1,\ldots,x_d)$ for all $c > 0$;*
2. *$\ell(\mathbf{e}_j) = 1$ for $\mathbf{e}_j$ the j-th unit vector in $\mathbb{R}^d$, $j = 1,\ldots,d$;*
3. *Bounds: $\max(x_1,\ldots,x_d) \leq \ell(x_1,\ldots,x_d) \leq x_1 + \cdots + x_d$;*
4. *Convexity: ℓ is a convex function.*

Proof. (1) The homogeneity property follows from the definition of the stable tail dependence function as,

$$\begin{aligned}\ell(cx_1,\ldots,cx_d) &= \lim_{t\downarrow 0}\frac{1 - C_F(1-ctx_1,\ldots,1-ctx_d)}{t} = c\lim_{t\downarrow 0}\frac{1 - C_F(1-ctx_1,\ldots,1-ctx_d)}{ct} \\ &= c\lim_{\omega\downarrow 0}\frac{1 - C_F(1-\omega x_1,\ldots,1-\omega x_d)}{\omega} = c\ell(x_1,\ldots,x_d).\end{aligned}$$

(2) Suppose $j = 1$. The second property follows from the observation that $C(u,1,\ldots,1) = u$ for any $u \in [0,1]$ and from the definition:

$$\ell(e_1) = \ell(1,0,\ldots,0) = -\log C\left(e^{-1},1,\ldots,1\right) = -\log\left(e^{-1}\right) = 1.$$

The argument is similar for all $j \in \{1,\ldots,d\}$.
(3) The bounds from ℓ follow from the fact that the extreme value copula C is bounded between independence and comonotonicity:

$$\Pi_d = u_1\cdots u_d \leq C(u_1,\ldots,u_d) \leq \min(u_1,\ldots,u_d) = M_d.$$

Hence,

$$\exp\left(-(x_1+\cdots+x_d)\right) \leq C\left(e^{-x_1},\ldots,e^{-x_d}\right) \leq \min\left(e^{-x_1},\ldots,e^{-x_d}\right),$$

and

$$\max(x_1,\ldots,x_d) = -\log\min\left(e^{-x_1},\ldots,e^{-x_d}\right) \leq -\log C\left(e^{-x_1},\ldots,e^{-x_d}\right) \leq x_1+\cdots+x_d.$$

(4) The convexity of ℓ follows trivially from the fact that the logarithm is a concave function. ■

In the bivariate case, these four properties are sufficient for a function ℓ to be a tail dependence function. However, in higher dimensions, $d > 2$, these properties are necessary but not sufficient. This is illustrated by the following example taken from Beirlant et al. (2004).

Example 3.10. *Consider the trivariate function $\ell : [0,\infty)^3 \to [0,\infty)$ given by*

$$\ell(x_1,x_2,x_3) = \max\left\{(x_1+x_2),(x_2+x_3),(x_1+x_3)\right\}.$$

This function satisfies the four properties of the stable tail dependence function. However, it turns out that it is not a tail dependence function. This can be established by the fact that $\ell(1,1,0) = \ell(1,0,1) = \ell(0,1,1) = 2$, implying pairwise independence, implying full independence, $\ell(1,1,1) = 3$, by Theorem 3.3. This is in contradiction with the fact that $\ell(1,1,1) = 2$, so ℓ cannot be a tail dependence function.

The stable tail dependence function of the Gumbel copula can be easily derived from the relationship between the STDF and the extreme value copulas,

$$\begin{aligned}\ell(x_1,\ldots,x_d) &= -\log C\left(e^{-x_1},\ldots,e^{-x_d}\right)\\ &= -\log\left(\exp\left(-\left[(-\log(e^{-x_1}))^\theta+\cdots+(-\log(e^{-x_d}))^\theta\right]^{1/\theta}\right)\right)\\ &= \left(x_1^\theta+\cdots+x_d^\theta\right)^{1/\theta},\end{aligned} \tag{3.4.5}$$

with $\theta \geq 1$. Note that for any $d \geq 2$, the tail independent STDF ($\ell(\mathbf{x}) = x_1 + \cdots + x_d$) is retrieved for $\theta = 1$, while for $\theta \to \infty$ the STDF of the Gumbel copula converges to the comonotonic case ($\ell(\mathbf{x}) = \max(x_1,\ldots,x_d)$). Although Theorem 3.5 already offers the necessary tools to assess whether a certain Archimedean copula belongs to the maximum domain of attraction of the Gumbel copula, the following theorem of Charpentier and Segers (2009) can also be used to immediately find the stable tail dependence function of an Archimedean copula. Contrary to the previous results, Charpentier and Segers (2009) employ the inverse generator of the Archimedean copula and investigate its behavior near 1. The result is presented below.

Theorem 3.9. *(Based on Theorem 4.1 in Charpentier and Segers (2009)). Let $\mathbf{X} = (X_1,\ldots,X_d)$ be a random vector with joint distribution function F and Archimedean copula C_F with generator ψ and inverse generator ψ^{-1}. If the limit*

$$\lim_{t\downarrow 0}\frac{t\left(\psi^{-1}\right)'(1-t)}{\psi^{-1}(1-t)} = -\theta, \quad \theta \geq 1,$$

exists, then the stable tail dependence function ℓ of $\mathbf{X}$ is given by

$$\ell(x_1,\ldots,x_d) = \begin{cases} x_1+\cdots+x_d & \text{if } \theta = 1\\ \left(x_1^\theta+\cdots+x_d^\theta\right)^{1/\theta} & \text{if } 1<\theta<\infty\\ \max(x_1,\ldots,x_d) & \text{if } \theta=\infty\end{cases}, \quad (x_1,\ldots,x_d)\in\mathbb{R}_+^d.$$

Proof. (Based on the proof presented in Charpentier and Segers (2009)). First of all, note that the function $x \mapsto \psi^{-1}(1-x)$ is regularly varying at zero with index θ by Karamata's theorem (see Appendix A). Hence,

$$\lim_{t\downarrow 0}\frac{\psi^{-1}(1-tx)}{\psi^{-1}(1-t)} = x^\theta, \quad x\in\mathbb{R}_+.$$

This, in turn, implies that the function $x \mapsto 1/\psi^{-1}(1-1/x)$ is regularly varying at ∞ with index $1/\theta$ and thus that the function $1-\psi$ is regularly varying at zero with index $1/\theta$. First, we consider $1 \leq \theta < \infty$. The STDF ℓ can in this case be written in terms of the generator as follows,

$$\begin{aligned}\ell(x_1,\ldots,x_d) &= \lim_{t\downarrow 0}\frac{1-C_F(1-tx_1,\ldots,1-tx_d)}{t} = \lim_{t\downarrow 0}\frac{1-\psi\left(\psi^{-1}(1-tx_1)+\cdots+\psi^{-1}(1-tx_d)\right)}{t}\\ &= \lim_{t\downarrow 0}\frac{1}{1-\psi(\psi^{-1}(1-t))}\left(1-\psi\left(\psi^{-1}(1-t)\left(\frac{\psi^{-1}(1-tx_1)}{\psi^{-1}(1-t)}+\cdots+\frac{\psi^{-1}(1-tx_d)}{\psi^{-1}(1-t)}\right)\right)\right).\end{aligned}$$

By the uniform convergence theorem, the right-hand-side of the equation above converges to $\left(x_1^\theta+\cdots+x_d^\theta\right)^{1/\theta}$, as required. If $\theta = 1$, this simplifies to $x_1+\cdots+x_d$.

Now consider the case $\theta = \infty$. Let $1 < c < \infty$ be a constant. Since $x \mapsto \psi^{-1}(1-x)$ is regularly varying at zero with index ∞, we have

$$\lim_{t\downarrow 0}\frac{\psi^{-1}(1-ct)}{\psi^{-1}(1-t)} = \infty.$$

Therefore, it follows that

$$\begin{aligned}\psi^{-1}\left((1-t\left((\max(x_1,\ldots,x_d)\right)\right) &\leq \psi^{-1}(1-tx_1)+\cdots+\psi^{-1}(1-tx_d)\\ &\leq d\psi^{-1}\left(1-t\left((\max(x_1,\ldots,x_d)\right)\right)\\ &\leq \psi^{-1}\left((1-ct\left((\max(x_1,\ldots,x_d)\right)\right),\end{aligned}$$

for all t in a right neighbourhood of zero. By applying the function $1-\psi$ to the various parts of the inequalities and multiplying by t^{-1} we find

$$\max(x_1,\dots,x_d) \leq \frac{1-\psi\left(\psi^{-1}(1-tx_1)+\cdots+\psi^{-1}(1-tx_d)\right)}{t} \leq c\max(x_1,\dots,x_d).$$

By letting c decrease to one and by taking the limit $t\to 0$, this becomes

$$\max(x_1,\dots,x_d) \leq \ell(x_1,\dots,x_d) \leq \max(x_1,\dots,x_d),$$

finishing the proof in case $\theta=\infty$. ■

Example 3.11. *The inverse generator of the Joe copula is given by*

$$\psi^{-1}(t) = -\log\left(1-(1-t)^{\theta}\right), \quad \theta \geq 1$$

with first derivative

$$\left(\psi^{-1}\right)'(t) = -\frac{\theta(1-t)^{\theta-1}}{1-(1-t)^{\theta}}.$$

Since

$$\lim_{t\downarrow 0} -\frac{t\left(\psi^{-1}\right)'(1-t)}{\psi^{-1}(1-t)} = \lim_{t\downarrow 0} -\frac{\theta t^{\theta}}{(1-t^{\theta})\log(1-t^{\theta})} = \lim_{t\downarrow 0}\frac{\theta t^{\theta}}{(1-t^{\theta})t^{\theta}} = \theta,$$

where the relation $\log(x) = x-1$ *for* $x\to 1$ *is used again, it follows by Theorem 3.9 that the Joe copula belongs to the maximum domain of attraction of the Gumbel copula with parameter* θ.

For elliptical copulas and vine copulas it is less straightforward to derive the STDF since these copulas are defined by their density while the STDF is defined based on the copula distribution function. One approach to derive the STDF for these copulas employs the copula tail density (see Section 3.3.2). It can be observed that the tail density is related to the derivative of the ℓ-function. This is shown in the following proposition.

Proposition 3.6. *Let* $\mathbf{X}=(X_1,\dots,X_d)$ *be a random vector with joint distribution function* F *and copula* C_F *with density* c_F. *If* C_F *converges to an extreme value copula* C *and if its tail density* Υ *and stable tail dependence function* ℓ *exist, the tail density can be derived from the* ℓ*-function as follows,*

$$\Upsilon(x_1,\dots,x_d) = (-1)^{d-1}\frac{\partial^d}{\partial x_1\dots\partial x_d}\ell(x_1,\dots,x_d), \quad (x_1,\dots,x_d)\in\mathbb{R}_+^d.$$

Proof. Assuming uniform convergence of the ℓ-function such that the limit and partial derivatives can be exchanged, the result is retrieved as follows,

$$\begin{aligned}\frac{\partial^d}{\partial x_1\dots\partial x_d}\ell(x_1,\dots,x_d) &= \frac{\partial^d}{\partial x_1\dots\partial x_d}\lim_{t\downarrow 0}\frac{1-C_F(1-tx_1,\dots,1-tx_d)}{t}\\ &= \lim_{t\downarrow 0}(-t)^{d-1}c_F(1-tx_1,\dots,1-tx_d)\\ &= (-1)^{d-1}\Upsilon(x_1,\dots,x_d).\end{aligned}$$

■

This proposition can be used to find the tail density for Archimedean copulas. Note that it is required to know the tail density for Archimdean copulas to assess the tail density of vines that consist of one or more bivariate Archimedean linking copulas. The tail density of a d-dimensional Archimedean copula with a regularly varying generator is presented in the following proposition.

Proposition 3.7. *(Li and Wu, 2013). Let* $C_F(\mathbf{u}) = \psi\left(\sum_{i=1}^d \psi^{-1}(u_i)\right)$ *be an Archimedean copula with generator* ψ^{-1}, *where* ψ *is regularly varying at* 1 *with tail index* $\beta>1$. *The upper tail density of* C_F *is given by*

$$\Upsilon(\mathbf{w}) = \prod_{i=2}^{d}((i-1)\beta-1)\left(\prod_{i=1}^{d}w_i\right)^{\beta-1}\left(\sum_{i=1}^{d}w_i^{\beta}\right)^{-d+1/\beta}. \tag{3.4.6}$$

Proof. Since it is assumed that the generator is regularly varying at 1 with tail index $\beta > 1$, Theorem 3.9 implies that the STDF of the Archimedean copula is given by the STDF of the Gumbel copula with parameter β. By Proposition 3.6 it follows that the tail density of the Gumbel copula can be determined by taking partial derivatives of its stable tail dependence function. Hence,

$$\begin{aligned}
\frac{\partial^d}{\partial x_1 \dots \partial x_d} \ell(x_1, \dots, x_d) &= \frac{\partial^d}{\partial x_1 \dots \partial x_d} \left(x_1^\theta + \dots + x_d^\theta\right)^{1/\theta} \\
&= \theta x_1^{\theta-1} \frac{1}{\theta} \frac{\partial^{d-1}}{\partial x_2 \dots \partial x_d} \left(x_1^\theta + \dots + x_d^\theta\right)^{-1+1/\theta} \\
&= \theta^d \left(x_1 \cdots x_d\right)^{\theta-1} \left(\prod_{i=1}^{d} (-i+1+1/\theta)\right) \left(x_1^\theta + \dots + x_d^\theta\right)^{-d+1/\theta} \\
&= \prod_{i=1}^{d} (-(\theta(i-1)-1)) \left(\prod_{i=1}^{d} x_i\right)^{\theta-1} \left(\sum_{i=1}^{d} x_i^\theta\right)^{-d+1/\theta} \\
&= (-1)^{d-1} \prod_{i=2}^{d} ((i-1)\theta - 1) \left(\prod_{i=1}^{d} x_i\right)^{\theta-1} \left(\sum_{i=1}^{d} x_i^\theta\right)^{-d+1/\theta}.
\end{aligned}$$

■

The results above show that the copula tail density Υ can be retrieved from the stable tail dependence function. In the opposite direction, we show that the STDF can be retrieved from the copula tail density by integrating over an appropriate region in $\mathbb{R}_+^d$ (Proposition 3.8). In particular, this leads to expressions for the STDF of the t-copula and of D-vine copulas. To our knowledge, an expression for the STDF of a D-vine has not yet been given in the literature so far. Although Joe et al. (2010), Li (2013), and Joe (2015) have explored the extremal behavior of D-vines, their focus was on determining the bivariate tail dependence structure of variable pairs that are not directly specified in the vine stucture. Since this can be done based on the tail copula, another dependence function which will be introduced in Section 3.4.3, an explicit expression for the STDF of a D-vine was not developed in their work.

Proposition 3.8. *Let $\mathbf{X} = (X_1, \dots, X_d)$ be a random vector with joint distribution function F and copula C_F with density c_F. If C_F admits a copula tail density function Υ and a stable tail dependence function ℓ, the STDF can be determined by integrating marginal tail densities as follows,*

$$\ell(x_1, \dots, x_d) = \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} (-1)^{|S|-1} \int_0^{x_{i_1}} \cdots \int_0^{x_{i_{|S|}}} \Upsilon_S(v_{i_1}, \dots, v_{i_{|S|}}) \mathrm{d}v_{i_1} \dots \mathrm{d}v_{i_{|S|}}.$$

Here, Υ_S denotes the marginal copula tail density concerning the variables $X_i, i \in S$, and $|S|$ denotes the number of elements in the set S.

Proof. The proof follows by the inclusion-exclusion formula and an appropriate substitution:

$$\begin{aligned}
\ell(x_1, \dots, x_d) &= \lim_{t \downarrow 0} \frac{1 - C_F(1 - tx_1, \dots, 1 - tx_d)}{t} \\
&= \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(\bigcup_{j=1}^{d} \{U_j > 1 - tx_j\}\right) \\
&= \lim_{t \downarrow 0} t^{-1} \left(\sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} (-1)^{|S|-1} \mathbb{P}(U_i > 1 - tx_i, i \in S)\right) \\
&= \lim_{t \downarrow 0} \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} t^{-1} (-1)^{|S|-1} \int_{1-x_{i_1}}^{1} \cdots \int_{1-tx_{i_{|S|}}}^{1} c(w_{i_1}, \dots, w_{i_{|S|}}) \mathrm{d}w_{i_1} \dots \mathrm{d}w_{i_{|S|}} \\
&= \lim_{t \downarrow 0} \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} (-1)^{|S|-1} \int_0^{x_{i_1}} \cdots \int_0^{x_{i_{|S|}}} c(1 - tv_{i_1}, \dots, 1 - tv_{i_{|S|}}) t^{|S|-1} \mathrm{d}v_{i_1} \dots \mathrm{d}v_{i_{|S|}} \\
&= \sum_{\emptyset \neq S \subseteq \{1,\dots,d\}} (-1)^{|S|-1} \int_0^{x_{i_1}} \cdots \int_0^{x_{i_{|S|}}} \Upsilon_S(v_{i_1}, \dots, v_{i_{|S|}}) \mathrm{d}v_{i_1} \dots \mathrm{d}v_{i_{|S|}}.
\end{aligned}$$

In order to exchange limit and integral in the last step, uniform convergence of the tail density has to be assumed. This is implied by the definition of the copula tail densiy. ■

Using the expression for the copula tail density of a D-vine given in Theorem 3.7, the following example illustrates how the STDF of a trivariate D-vine can be determined.

Example 3.12. *For a 3-dimensional* D*-vine,*

$$\begin{aligned}\ell(x_1,x_2,x_3) &= x_1+x_2+x_3-\int_0^{x_1}\int_0^{x_2}\Upsilon_{12}(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2-\int_0^{x_1}\int_0^{x_3}\Upsilon_{13}(w_1,w_3)\mathrm{d}w_1\mathrm{d}w_3\\ &\quad-\int_0^{x_2}\int_0^{x_3}\Upsilon_{23}(w_2,w_3)\mathrm{d}w_2\mathrm{d}w_3+\int_0^{x_1}\int_0^{x_2}\int_0^{x_3}\Upsilon_{123}(w_1,w_2,w_3)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3\\ &= x_1+x_2+x_3-\int_0^{x_1}\int_0^{x_2}\Upsilon_{12}(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2-\int_0^{x_1}\int_0^{x_3}\Upsilon_{13}(w_1,w_3)\mathrm{d}w_1\mathrm{d}w_3\\ &\quad-\int_0^{x_2}\int_0^{x_3}\Upsilon_{23}(w_2,w_3)\mathrm{d}w_2\mathrm{d}w_3+\int_0^{x_1}\int_0^{x_2}\int_0^{x_3}\Upsilon_{12}(x_1,x_2)\Upsilon_{23}(x_2,x_3)\\ &\qquad\cdot c_{13|2}\left(1-\int_0^{x_1}\Upsilon_{12}(w_1,x_2)\mathrm{d}w_1,1-\int_0^{x_1}\Upsilon_{23}(x_2,w_3)\mathrm{d}w_3\right)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3.\end{aligned}$$

Recall from Section 3.3.2 that the marginal tail densities can be retrieved from the full tail density by integrating out the appropriate variables. For example,

$$\int_0^{x_1}\int_0^{x_2}\Upsilon_{12}(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2=\int_0^{x_1}\int_0^{x_2}\int_0^{\infty}\Upsilon_{123}(w_1,w_2,w_3)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3.$$

Since the numerical integration for (semi-)infinite bounds can be troublesome, especially in higher dimensions, the semi-infinite interval is rescaled to a bounded interval through application of the substitution $w_i=-\log(1-\tilde{w}_i)$. To explore this substitution, consider the simple integral $\int_0^\infty x\mathrm{d}x$. After the substitution $x=-\log(1-y)\Leftrightarrow y=1-e^{-x}$, the integral can be written with finite bounds as $\int_0^1-\log(1-y)/(1-y)\mathrm{d}y$. Using the `adaptIntegrate` function in `R` the multivariate integrals can be evaluated.

3.4.2 Pickands dependence function

The homogeneity of the STDF implies that to determine the behavior of ℓ on its entire range it is sufficient to know the behavior of ℓ on the unit simplex $\Delta_d=\{\mathbf{w}\in\mathbb{R}_+^d: w_1+\cdots+w_d=1\}$. The restriction of the STDF to the unit simplex is called the Pickands dependence function, denoted with A. When the ℓ-function is evaluated on the unit simplex, it is precisely the same as the Pickands A-function. Hence, it can be defined as follows.

Definition 3.9. *For* $\mathbf{x}=(x_1,\ldots,x_d)\in\Delta_d$ *Pickands dependence function is defined as*

$$A(x_1,\ldots,x_d)=\ell(x_1,\ldots,x_d)=\lim_{t\downarrow 0}\frac{1-C_F(1-tx_1,\ldots,1-tx_d)}{t}. \tag{3.4.7}$$

Due to the restriction to the unit simplex, the dependence function actually becomes $(d-1)$-dimensional, since the last coordinate can be characterized in terms of the sum of the previous coordinates. In reverse, the ℓ-function can be retrieved from the A-function through a simple scaling transformation based on its homogeneity: for $\mathbf{x}\in\mathbb{R}_+^d$,

$$\ell(x_1,\ldots,x_d)=(x_1+\cdots+x_d)A\left(\frac{x_1}{x_1+\cdots+x_d},\ldots,\frac{x_d}{x_1+\cdots+x_d}\right). \tag{3.4.8}$$

Originally, Pickands (1981) defined the Pickands dependence function for the bivariate case based on the spectral representation (Section 3.2.2) employing the L_1-norm for both norms $||\cdot||_p$ and $||\cdot||_q$, i.e., for $x\in[0,1]$,

$$A(x)=\int_0^1\max(\omega(1-x),(1-\omega)x)\mathrm{d}S(\omega)$$

with S the spectral measure defined on $[0, 1]$ satisfying

$$\int_0^1 \omega \mathrm{d}S(\omega) = \int_0^1 (1-\omega)\mathrm{d}S(\omega) = 1.$$

For the multivariate case, this can be extended to

$$A(x_1, \ldots, x_d) = \int_{\Delta_d} \max(\omega_1 x_1, \ldots, \omega_d x_d)\mathrm{d}S(\omega),$$

with S the spectral measure on the d-dimensional unit simplex. Because of the relationship between the STDF and the spectral measure (Equation 3.4.4), this definition of Pickands dependence function is equivalent to the STDF-based definition presented above.

Similarly to the STDF, the Pickands dependence function can represent the extreme value copula. This leads to the following Proposition 3.9. For the bivariate case, this representation is known as Pickands representation (Pickands, 1981). This representation is used quite often in literature and is sometimes employed to define the extreme value copula.

Proposition 3.9. *The copula C is an extreme value copula if and only if it has the representation*

$$C(u_1, \ldots, u_d) = \exp\left\{ A\left(\frac{\log u_1}{\sum_{j=1}^d \log u_j}, \ldots, \frac{\log u_d}{\sum_{j=1}^d \log u_j} \right) \sum_{j=1}^d \log u_j \right\}.$$

Proof. The result follows immediately from Proposition 3.5 and the relationship between the STDF and Pickands dependence function given in Equation 3.4.8. ■

Because of its direct relationship with the stable tail dependence function (Equation 3.4.7), the Pickands dependence function inherits the properties of the STDF. Hence, Theorem 3.8 also holds for the restriction of the STDF to the unit simplex. That is, the A-function is homogeneous and convex, is equal to one for unit vectors ($A(\mathbf{e}_j) = 1$ for $\mathbf{e}_j \in \mathbb{R}^d$ the j-th unit vector, $j = 1, \ldots, d$) and is bounded between comonotonicity and independence,

$$\max(\omega_1, \ldots, \omega_d) \leq A(\omega_1, \ldots, \omega_d) \leq 1, \quad \omega \in \Delta_d. \tag{3.4.9}$$

The properties of the Pickands dependence function allow for an intuitively attractive method to visualize the tail dependence structure for bivariate and even trivariate dependence structures. For example, Figure 3.4.1 shows the general range and bounds of the bivariate Pickands dependence function. Due to the restriction to the unit simplex, the parametrization of the A-function is one-dimensional for a bivariate dependence structure.

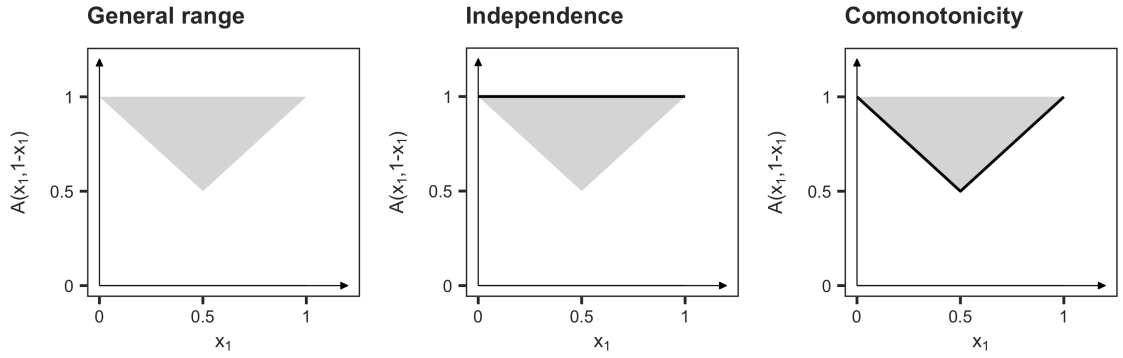

Figure 3.4.1: General range and bounds on the bivariate Pickands dependence function.

For a trivariate dependence structure, the parametrization of the A-function is two-dimensional, which can be visualized in a 3D-plot. Figure 3.4.2 shows the bounds of the A-function for a trivariate tail dependence structure.

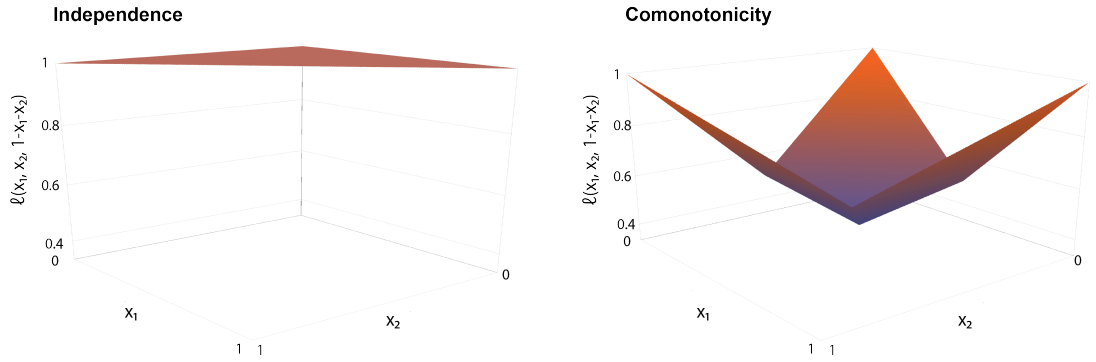

Figure 3.4.2: 3D visualizations of the bounds on the trivariate Pickands dependence function. Left-hand panel: independence bound. Right-hand panel: comonotonicity bound.

Using this visualization method, the tail dependence structure of the Gumbel copula can be further assessed. The STDF of the Gumbel copula was given in Equation 3.4.5. The 3D-plots of the Pickands dependence function for a trivariate Gumbel copula with either light dependence ($\theta = 2$) or strong dependence ($\theta = 5$) are shown in Figure 3.4.3. In addition, the pairwise marginal bivariate dependence functions are shown in Figure 3.4.4. Note that although the Gumbel copula can obtain a wide range of dependence strengths, it is not very flexible in dimensions $d > 2$ because equivalent tail dependence structures are imposed on all pairs. Especially for high dimensions this can become a large limitation of the model.

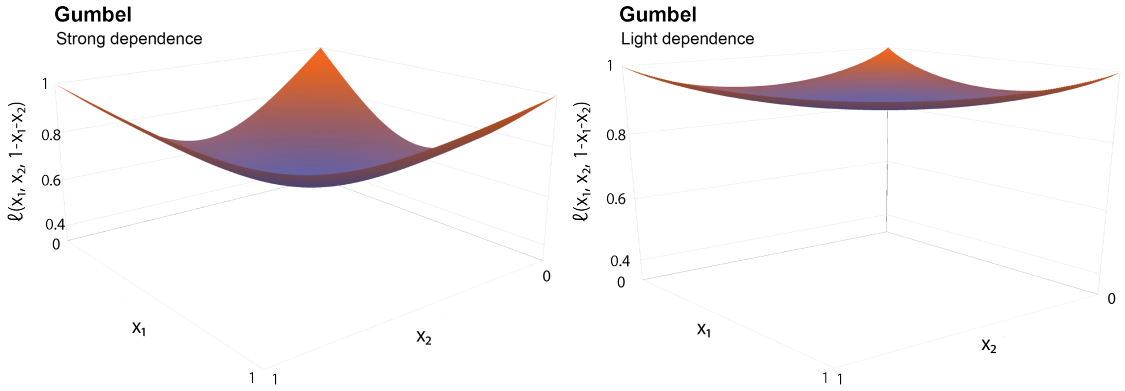

Figure 3.4.3: 3D visualizations of the Pickands dependence function for the trivariate Gumbel copula. The left-hand figure shows the Gumbel Pickands dependence function for strong tail dependence ($\theta = 5$), while the right-hand figure shows the Gumbel A-function for light tail dependence ($\theta = 2$).

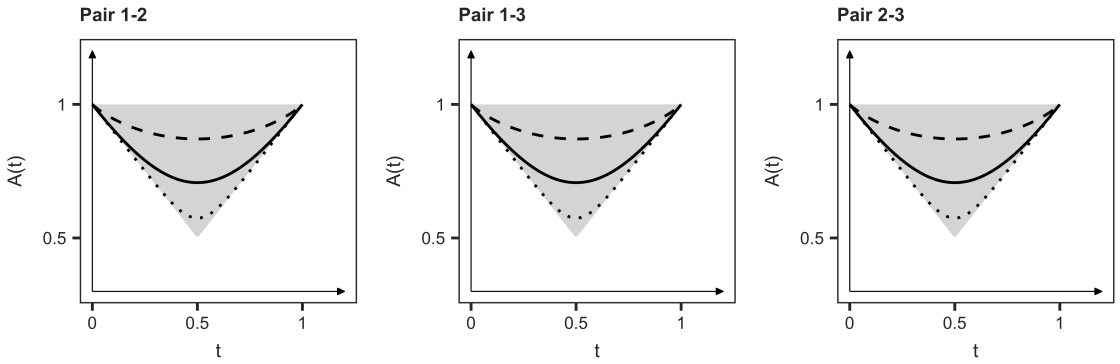

Figure 3.4.4: Bivariate marginal Pickands dependence functions for the trivariate Gumbel copula with parameters $\theta = 5$ (dashed line), $\theta = 2$ (solid line), and $\theta = 1.2$ (dotted line).

In order to visualize the dependence strucure of vines and elliptical copulas, the relationship between the STDF and the marginal copula tail densities can be employed (Proposition 3.8). To determine a trivariate ℓ-function, this entails the evaluation of the following expression,

$$\ell(x_1,x_2,x_3) = x_1+x_2+x_3 - \int_0^{x_1}\int_0^{x_2} \Upsilon_{12}(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2 - \int_0^{x_1}\int_0^{x_3} \Upsilon_{13}(w_1,w_3)\mathrm{d}w_1\mathrm{d}w_3 \\ - \int_0^{x_2}\int_0^{x_3} \Upsilon_{23}(w_2,w_3)\mathrm{d}w_2\mathrm{d}w_3 + \int_0^{x_1}\int_0^{x_2}\int_0^{x_3} \Upsilon_{123}(w_1,w_2,w_3)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3.$$

Although numerical integration of the marginal copula tail densities is possible, it is computationally intensive to evaluate the Pickands dependence function on its entire range. Alternatively, an expression for the Pickands dependence function of the t-EV copula was derived by Nikoloulopoulos et al. (2009). Denoting by $T_{d,\nu}$ the d-dimensional t-distribution function with ν degrees of freedom and partial correlation matrix $\mathbf{R}$, the t-EV copula is given by the following Pickands dependence function,

$$A(w_1,\ldots,w_d) = \sum_{j=1}^{d} w_j T_{d-1,\nu+1,\mathbf{R}_j}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho_{ij}^2}}\left[\left(\frac{w_i}{w_j}\right)^{-1/\nu} - \rho_{ij}\right], i\neq j; \mathbf{R}_j\right), \tag{3.4.10}$$

where $\mathbf{R}_j$ is equal to the partial correlation matrix $\mathbf{R}$ with the j-th column and j-th row removed. That is, $\mathbf{R}_j = (\rho_{k_1,k_2;j})_{k_1,k_2\neq j}$. For $d=2$, $\mathbf{R}_1$ and $\mathbf{R}_2$ are equal to 1. In the bivariate case, this simplifies to

$$A(w) = wT_{\nu+1}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\left[\left(\frac{w}{1-w}\right)^{-1/\nu} - \rho\right]\right) + (1-w)T_{\nu+1}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\left[\left(\frac{1-w}{w}\right)^{-1/\nu} - \rho\right]\right).$$

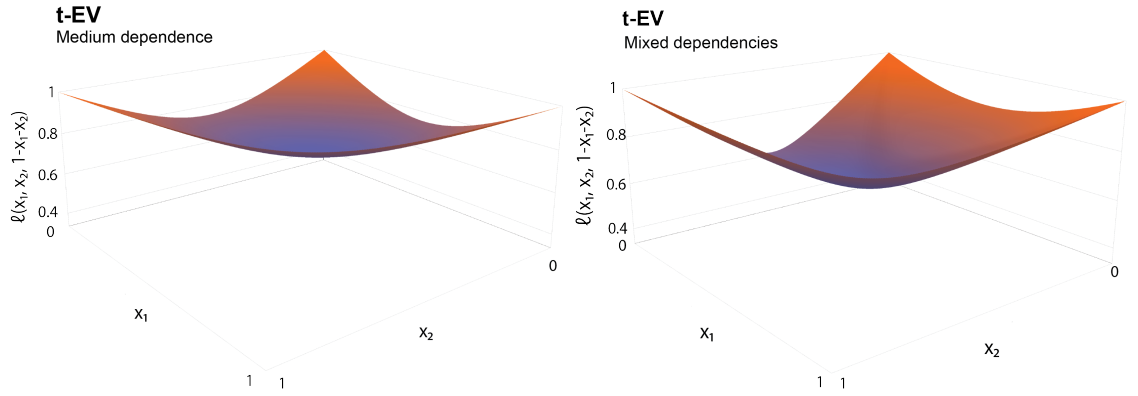

Figure 3.4.5: 3D visualizations of the Pickands dependence function for the 3-dimensional t-copula. The left-hand figure shows the t-EV Pickands dependence function for medium dependence ($\rho = 0.5$ for all bivariate margins). The right-hand figure shows the t-EV A-function for mixed bivariate dependencies ($\rho = 0.2, 0.5, 0.8$).

The maximum tail dependence strength that can be attained with the t-EV model is retrieved for $\nu \downarrow 0$ and $\rho_{ij} \to 1$. In contrast, the minimum tail dependence strength that can be attained with the t-EV model is retrieved for $\nu \to \infty$ and $\rho_{ij} \to 0$. Nikoloulopoulos et al. (2009) show that as $\nu \to \infty$ the t-EV copula converges to the Hüsler-Reiss model. As discussed in Section 3.3, this is the model the Normal copula converges to if correlations become stronger as the number of observations increases. The spatial analogue of the Hüsler-Reiss model is the well-known Brown-Resnick model. In the bivariate case, Hüsler and Reiss (1989) showed that on condition that $(1-\rho_n)\log n \to \delta > 0$ as $n \to \infty$, the Normal copula belongs to the MDA of the extreme value copula defined by the following Pickands dependence function,

$$A(w) = (1-w)\Phi\left(\delta^{-1} + \frac{\delta}{2}\log\frac{1-w}{w}\right) + w\Phi\left(\delta^{-1} + \frac{\delta}{2}\log\frac{w}{1-w}\right).$$

Properties of the bivariate Hüsler-Reiss dependence function are given in Falk et al. (2011) for example. The multivariate case was addressed by Hüsler and Reiss (1989) and also by (Hashorva,

2006), but the most appealing expression for the multivariate Hüsler-Reiss dependence function is presented by Nikoloulopoulos et al. (2009). On condition that

$$\lim_{n\to\infty} (1-\rho_{ij}(n)) \log n =: \delta_{ij},$$

for positive parameters δ_{ij}, $i \neq j$, with $\delta_{ij} = \delta_{ji}$, they show that the multivariate Normal copula belongs to the MDA of the extreme value copula defined by the following Pickands dependence function,

$$A(w_1,\ldots,w_d) = \sum_{j=1}^{d} w_j \Phi_{d-1,\mathbf{R}_j}\left(\delta_{ij}^{-1} + \frac{\delta_{ij}}{2}\log\frac{w_j}{w_i}, i \in I_j\right), \tag{3.4.11}$$

where $\mathbf{R}_j$ is a dispersion matrix with entries

$$\rho_{i,k;j} = \frac{\delta_{ij}^{-2} + \delta_{kj}^{-2} - \delta_{ik}^{-2}}{2\delta_{ij}^{-1}\delta_{kj}^{-1}},$$

for $i \in I$ and $k \in I_j$ and with $\delta_{ii}^{-1} = 0$. Since the previously mentioned expressions involve multivariate integration and other computational complexities, this one is to be preferred. The maximal tail dependence strength that can be attained with the Hüsler-Reiss model is achieved for $\delta_{ij} \uparrow 1$. In contrast, the minimum tail dependence strength is attained for $\delta_{ij} \downarrow 0$ and corresponds to tail independence.

For the bivariate case, several additional examples of parametric forms of A are presented in Hutchinson and Lai (1990) and in Smith (1994). However, since the aim of this research is to characterize *multivariate* tail dependence structures, we focus on dependence function that are also defined for higher dimensions.

3.4.3 Tail copula

A final dependence function that is considered is the so-called tail copula. The concept of a tail copula was introduced by Schmidt and Stadtmüller (2006) and again under a different name in Joe et al. (2010). It can be defined as follows.

Definition 3.10. *Let $X_1,\ldots,X_d$ be random variables with marginal distributions $F_1,\ldots,F_d$, joint distribution F, and copula C_F. The tail copula function $b : \mathbb{R}_+^d \to \mathbb{R}_+$ is defined for $\mathbf{x} = (x_1,\ldots,x_d) \in \mathbb{R}_+^d$ as*

$$b(x_1,\ldots,x_d) = \lim_{t\downarrow 0} \frac{\overline{C}_F(1-tx_1,\ldots,1-tx_d)}{t}, \tag{3.4.12}$$

if the limit exists. Here, $\overline{C}_F$ denotes the survival function of the copula C_F.

The tail copula can be seen as an adaption from the STDF: the tail copula measures the limiting probability that all variables are large given that one of them is large. In contrast, the stable tail dependence function was concerned with the dependence in the region where at least one variable is large, but not all variables have to be. Figure 3.4.6 illustrates the difference for the 2-dimensional case. One may note that this difference is actually trivial for the bivariate case because the marginals are known standard uniforms. This, however, is not the case for higher dimensions. Just as for the stable tail dependence function, the coefficients $(x_1,\ldots,x_d)$ determine the direction in which the limit goes to the upper right corner.

Although the STDF and the tail copula are non-trivially different for higher dimensions, the inclusion-exclusion formula can be used to find the following relationship between the two dependence functions,

$$\ell(x_1,\ldots,x_d) = \sum_{S\subseteq I, S\neq\emptyset} (-1)^{|S|-1} b(x_i, i \in S). \tag{3.4.13}$$

This shows that the STDF contains the same information that is contained in the tail copulas of all possible subsets of a group of variables under consideration. Hence, the stable tail dependence

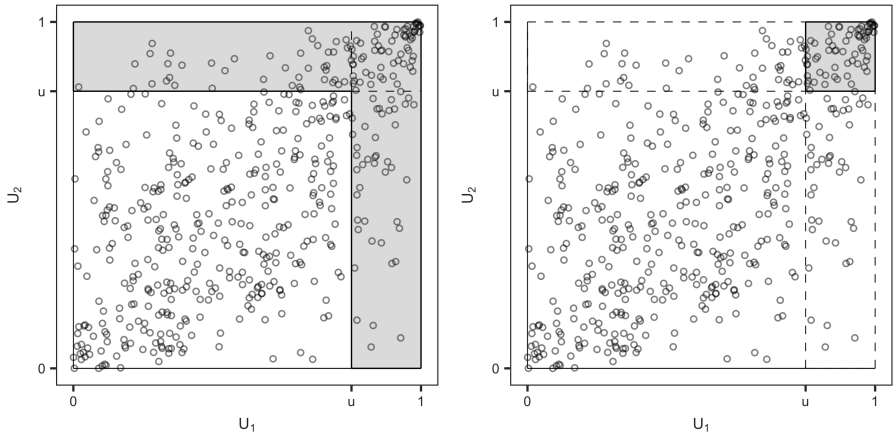

Figure 3.4.6: Left: region where at least one variable is large. The stable tail dependence function is concerned with this region. Right: region where all variables are large. The tail copula is concerned with this region.

function provides far more information than the tail copula evaluated in one point. On the other hand, both functions contain exactly the same information when the functions are known entirely.

It can be shown that the upper tail copula exists on $\mathbb{R}_+^d$ and $b \neq 0$ if the associated distribution function F lies in the maximum domain of attraction of an extreme value distribution with dependent margins (Resnick, 1987). On the other hand, if $b = 0$ everywhere on $\mathbb{R}_+^d$, this does not necessarily imply that the corresponding copula C_F belongs to the MDA of the independence copula Π. Only if the tail copula of all marginal subsets $\emptyset \neq S \subseteq \{1, \ldots, d\}$ with $|S| \geq 2$ is equal to zero everywhere, can it be concluded that $C_F \in MDA(\Pi)$. However, note that because of Theorem 3.3, it is sufficient to check all bivariate subsets only.

Since the b-function can be used to describe the dependence structure of the extreme-value distribution, the function is called a tail copula, even though it does not possess all copula properties. To avoid confusion, Joe et al. (2010) call the b-function "the tail dependence function" but since this label was already used in the previous section, we do not adopt their terminology. Although the tail copula is not a copula, Schmidt and Stadtmüller (2006) show that the bivariate tail copula satisfies most copula properties. The following theorem asserts some essential properties of the tail copula for higher dimensions. See Schmidt and Stadtmüller (2006) for more properties of the bivariate tail copula.

Theorem 3.10. *Each tail copula function $b : \mathbb{R}_+^d \to \mathbb{R}_+$ satisfies the following properties:*

1. *Homogeneity: $b(cx_1, \ldots, cx_d) = cb(x_1, \ldots, x_d)$ for all $c > 0$;*
2. *Groundedness: if there exists an $x_j, 1 \leq j \leq d$, such that $x_j = 0$ then $b(x_1, \ldots, x_d) = 0$;*
3. *Bounds: $0 \leq b(x_1, \ldots, x_d) \leq \min(x_1, \ldots, x_d)$;*

Proof. (1) The homogeneity property follows from the definition of the tail copula as,

$$b(cx_1, \ldots, cx_d) = \lim_{t \downarrow 0} \frac{\overline{C}_F(1 - ctx_1, \ldots, 1 - ctx_d)}{t} = c \lim_{t \downarrow 0} \frac{\overline{C}_F(1 - ctx_1, \ldots, 1 - ctx_d)}{ct}$$
$$= c \lim_{\omega \downarrow 0} \lim_{t \downarrow 0} \frac{\overline{C}_F(1 - \omega x_1, \ldots, 1 - \omega x_d)}{\omega} = cb(x_1, \ldots, x_d).$$

(2) W.l.o.g., suppose $x_1 = 0$.

$$b(0, x_2, \ldots, x_d) = \lim_{t \downarrow 0} \frac{\overline{C}_F(1, 1 - tx_2, \ldots, 1 - tx_d)}{t}$$
$$= \lim_{t \downarrow 0} \frac{\mathbb{P}(U_1 > 1, U_2 > 1 - tx_2, \ldots, U_d > 1 - tx_d)}{t} = 0,$$

since $U_1 \in [0, 1]$ and therefore $\mathbb{P}(U_1 > 1) = 0$.

(3) Defining the uniform random vector $\mathbf{V} = (V_1, \dots, V_d)$ as $\mathbf{V} = (1 - U_1, \dots, 1 - U_d)$ it is clear that

$$\begin{aligned}\overline{C}_F(1 - tx_1, \dots, 1 - tx_d) &= \mathbb{P}(U_1 > 1 - tx_1, \dots, U_d > 1 - tx_d) \\ &= \mathbb{P}(V_1 \leq tx_1, \dots, V_d \leq tx_d) = C_{F^*}(tx_1, \dots, tx_d)\end{aligned}$$

for some copula C_{F^*}. Using the Frechet-Hoeffding bounds for copulas, it follows that

$$\max(tx_1 + \cdots + tx_d - (d-1), 0) \leq C_{F^*}(tx_1, \dots, tx_d) \leq \min(tx_1, \dots, tx_d).$$

With $t \downarrow 0$ it follows that $0 \leq C_{F^*}(tx_1, \dots, tx_d) \leq t \min(x_1, \dots, x_d)$. Hence,

$$0 \leq \lim_{t \downarrow 0} \frac{\overline{C}_F(1 - tx_1, \dots, 1 - tx_d)}{t} \leq \min(x_1, \dots, x_d).$$

■

Similarly to the STDF, the tail copula is related to the tail density. This is shown in the following proposition.

Proposition 3.10. *(Li and Wu, 2013). Let Υ be the d-dimensional tail density of variables $X_1, \dots, X_d$ joint by a copula $C_F \in MDA(C)$. The tail copula of $(X_1, \dots, X_d)$ is given by*

$$b(\mathbf{x}) = \int_0^{x_1} \cdots \int_0^{x_d} \Upsilon(\mathbf{w}) \mathrm{d}w_1 \dots \mathrm{d}w_d.$$

Proof. The proof follows straightforwardly by employing an appropriate substitution:

$$\begin{aligned}b(\mathbf{x}) &= \lim_{t \downarrow 0} t^{-1} \overline{C}_F(1 - tx_1, \dots, 1 - tx_d) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(U_1 > 1 - tx_1, \dots, U_d > 1 - tx_d\right) \\ &= \lim_{t \downarrow 0} t^{-1} \int_{1-tx_1}^{1} \cdots \int_{1-tx_d}^{1} c_F(w_1, \dots, w_d) \mathrm{d}w_1 \dots \mathrm{d}w_d \\ &= \lim_{t \downarrow 0} t^{-1} \int_{x_1}^{0} \cdots \int_{x_d}^{0} c_F(1 - tv_1, \dots, 1 - tv_d)(-t)^d \mathrm{d}v_1 \dots \mathrm{d}v_d \\ &= \lim_{t \downarrow 0} \int_0^{x_1} \cdots \int_0^{x_d} t^{d-1} c_F(1 - tv_1, \dots, 1 - tv_d) \mathrm{d}v_1 \dots \mathrm{d}v_d \\ &= \int_0^{x_1} \cdots \int_0^{x_d} \Upsilon(\mathbf{v}) \mathrm{d}v_1 \dots \mathrm{d}v_d.\end{aligned}$$

In order to exchange limit and integral in the last step, uniform convergence of the copula tail density has to be assumed. This is implied by the definition of the copula tail density. ■

Employing this relationship, the tail copula function b of a D-vine copula can be determined based on the copula tail density derived in Theorem 3.7. The following example shows an expression for the tail copula of a trivariate D-vine.

Example 3.13. *For a 3-dimensional* D*-vine,*

$$\begin{aligned}b(x_1, x_2, x_3) = \int_0^{x_1} \int_0^{x_2} \int_0^{x_3} &\Upsilon_{12}(w_1, w_2) \Upsilon_{23}(w_2, w_3) \\ &\cdot c_{13|2}\left(1 - \int_0^{w_1} \Upsilon_{12}(\tilde{w}_1, w_2) \mathrm{d}\tilde{w}_1, 1 - \int_0^{w_3} \Upsilon_{23}(w_2, \tilde{w}_3) \mathrm{d}\tilde{w}_3\right) \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3.\end{aligned}$$

Using numerical integration procedures, it is possible to evaluate the multivariate tail copula for a given D-vine. Alternatively, Joe et al. (2010) developed expressions for the lower tail copula of a vine copula based on the Euler representation for homogeneous functions. We fashioned similar results for the upper tail copula, since this is the dependence function of interest (see Appendix B.2). However, when taking the limit, the final formulas presented in Joe et al. (2010) are equivalent to the tail density expressions shown above and do not offer computational advantages.

3.5 Summary

This chapter presented an extensive overview of full characterizations of the multivariate tail dependence structure. Starting with the joint distribution function of a d-dimensional vector of componentwise maxima, the multivariate extreme value distribution (MEVD) was identified as asymptotic distribution. The marginal distributions of the MEVD are univariate extreme value distributions which only depend on the tail index γ. There are infinitely many ways to couple these univariate margins into a proper MEVD. The dependence structure between multivariate extremes is, therefore, more complicated to assess than the marginal distributions of the extremes. The convergence-based maximum domain of attraction condition that underpins the MEVD definition can be challenging to assess, especially for higher dimensions. The exponent measure and the spectral measure are two characterizations of the MEVD that can alternatively be employed to assess the joint behavior of extremes. The exponent measure is a rescaled version of the MEVD that is homogeneous and can, therefore, be used to rescale regions. Moreover, based on the exponent measure, the critical result that pairwise tail independence is equivalent to multivariate tail independence (Theorem 3.3) was shown. The spectral measure decomposes the exponent measure into a radial part and a spectral part. An attractive property of the spectral measure is that it represents tail dependence as the distribution of the relative size of the contributions of the marginal components to their sum, given that the vector is large (Kiriliouk et al., 2014).

Both statistically and intuitively, these first few characterizations of multivariate tail dependence can be troublesome to work with. The main problem with the representations of the MEVD is that they also contain information on the marginal behavior of the maxima of $X_1, \ldots, X_d$. A copula-based approach that disentangles the marginal behavior from the dependence behavior mitigates this problem and leads to the definition of the extreme value copula which can be seen as the copula inherent to the MEVD. After considering the extreme value copula of several commonly used copula models, three dependence functions based on the extreme value copula are considered: the stable tail dependence function (STDF), Pickands dependence function, and the tail copula. The properties of these dependence functions make them more attractive to use some situations than the original extreme value copula. Furthermore, Segers (2012b) note that max-stable models such as the MEVD and the extreme value copula have the drawback that they are too coarse to provide an accurate characterization of the tail dependence structure if variables are asymptotic independent. Both the STDF and the tail copula can be extended easily to the scenario of tail independence. These extensions are however beyond the scope of this thesis.

Intuitively, the STDF captures the conditional probability that one or more variables exceed a high threshold, given that one of them does. Amongst several attractive properties, the STDF is shown to be convex and homogeneous of order. It is therefore sufficient to evaluate the STDF on the unit simplex to know its behavior everywhere. This restriction of the STDF yields the Pickands dependence function, which can be employed to visualize the entire tail dependence structure for bivariate and trivariate problems. Alternatively, the tail copula captures the conditional probability that *all* variables exceed a high threshold, given that one of them is extreme. The tail copula, therefore, captures only the most extreme part of the multivariate tail dependence in dimensions $d > 2$, while the STDF completely describes it. Segers (2012b) summarizes the interpretation of ℓ as "trouble in the air," whereas b only considers events extreme when "the sky is falling". Especially for large dimensions d, the probability that *all* variables are extreme at the same time will be negligible and therefore less relevant to consider. Moreover, a practical issue for large d is that joint d-dimensional exceedances are rarely observed in finite samples. Unless a sample contains an observation with all marginal being extreme, the tail copula indicates tail independence. On the other hand, the STDF incorporates events in which a single component becomes extreme, and hence finite samples provide more relevant observations. The STDF is furthermore convenient because it immediately yields the joint tail of the distribution function and can be retrieved for almost all commonly used copula models. In particular, an expression for the STDF of a d-dimensional D-vine was derived that can be evaluated numerically.

Chapter 4

Tail Dependence Measures

Whereas the previous chapter focused on describing the tail dependence structure fully, this chapter is concerned with statistics that can summarize to some extent the nature of the extremal dependence structure. Similar to the correlation measure for general dependence, a tail dependence summary measure can characterize the strength of the extremal dependence structure. In most cases, these summary measures only provide a partial description of the extremal dependence structure between random variables. Nevertheless, a tail dependence summary statistic can be employed to get an overview of the tail dependence structure and is much more accessible to interpret than an entire dependence function.

For the bivariate case, $d = 2$, the tail dependence coefficient (TDC) is the standard summary statistic to assess the tail dependence strength between two random variables. For higher dimensions, $d > 2$, however, multiple possible extensions of the bivariate TDC have been suggested in the literature. Section 4.2 is concerned with an overview of these alternatives and their properties. Most importantly, we prove that the multivariate extension of the TDC based on the stable tail dependence function can identify tail independence (Main Theorem 2). Besides these tail dependence coefficients, second order measures have been considered in the literature to further asses the tail dependence structure for variables that are asymptotically independent. The tail order is the topic of Section 4.3.

4.1 Desirable properties

Before introducing summary measures to describe bivariate and multivariate tail dependencies, it is instructive to consider the ideal properties a tail dependence measure should satisfy such that it can convey the maximum amount of information on the tail dependence structure. Initially, Rényi (1959) introduced a set of axioms to which a general dependence measure should adhere. However, although theoretically desirable, some of them are too strong and cannot be attained in practice. Schweizer and Wolff (1981) and Embrechts (2002) have therefore contemplated the properties a dependence measure for two continuous random variables should reasonably display. We extend their work to identify desirable properties of a dependence measure for higher dimensions and to *tail* dependence, specifically. The following properties are selected to be vital for a multivariate tail dependence measure $\delta(X_1, \ldots, X_d)$.

(P1) Existence
$\delta(X_1, \ldots, X_d)$ is defined for any $X_1, \ldots, X_d$.

(P2) Exchangeability
$\delta(X_1, \ldots, X_d) = \delta(X_{\pi(1)}, \ldots, X_{\pi(d)})$, where π is an arbitrary bijection $\pi : \{1, \ldots, d\} \to \{1, \ldots, d\}$ and represents a finite permutation of the indices $\{1, \ldots, d\}$. Hence, the dependence measure should yield the same result regardless of the order of the variables. For a dependence measure for bivariate data this property is known as symmetry, $\delta(X_1, X_2) = \delta(X_2, X_1)$.

(P3) Normalization
$0 \leq \delta(X_1, \ldots, X_d) \leq 1$, with the lower bound corresponding to weak tail dependence and the upper bound to strong tail dependence. Note that this is an appropriate normalization since the

dependence structure of an extreme value distribution cannot contain negative dependence (see Section 3.1.2). However, the exact bounds of the normalization are unimportant since any scaling operation can be used to obtain the required bounds, as long as the range of δ is bounded.

(P4) Invariance under monotone transformations
If the variable X_1 is transformed with a strictly monotone function f, then the dependence structure between $f(X_1)$ and $X_2, \ldots, X_d$ is the same as the dependence structure between X_1 and $X_2, \ldots, X_d$. Formally, if $f : \mathbb{R} \to \mathbb{R}$ is strictly monotone almost surely, then

$$\delta(f(X_1), X_2, \ldots, X_d) = \delta(X_1, \ldots, X_d).$$

Schweizer and Wolff (1981) show that if $(X_1, \ldots, X_d)$ follow a joint distribution function that is invariant under strictly increasing transformations, then any property of the joint distribution function is solely a function of the copula. Therefore, for variables $X_1, \ldots, X_d$ with marginal distributions $F_1, \ldots, F_d$, joint by a copula C_F, it is desirable for a tail dependence measure $\delta(X_1, \ldots, X_d)$ to be a function solely of the copula C_F:

$$\delta(X_1, \ldots, X_d) = \delta(C_F).$$

This implies that the tail dependence measure should not be influenced by the marginal behavior of the variables $X_1, \ldots, X_d$.

(P5) Identification of independence
In order to be able to identify tail independence, the following must hold for random variables $X_1, \ldots, X_d$ with continuous marginal distributions $F_1, \ldots, F_d$, joint by a copula C_F:

$$C_F \in MDA(\Pi) \Leftrightarrow \delta(X_1, \ldots, X_d) = 0.$$

Since it can be difficult to find a dependence measure that satisfies this property, Embrechts (2002) proposes the weaker property that dependence measures should be equal to zero in case of independence. For tail dependence measures this translates to the following:

$$C_F \in MDA(\Pi) \Rightarrow \delta(X_1, \ldots, X_d) = 0.$$

(P6) Identification of comonotonicity
In order to be able to identify asymptotic comonotonicity, the following must hold for random variables $X_1, \ldots, X_d$ with continuous marginal distributions $F_1, \ldots, F_d$, joint by a copula C_F:

$$C_F \in MDA(M_d) \Leftrightarrow \delta(X_1, \ldots, X_d) = 1,$$

with M_d the d-dimensional comonotonicity copula. Again, the weaker property can be defined as

$$C_F \in MDA(M_d) \Rightarrow \delta(X_1, \ldots, X_d) = 1.$$

One might argue that the identification of comonotonicity is less important than the identification of independence, since asymptotic comonotonic variables are not very common in practice.

(P7) Relationship between $\delta(C_F)$ and $\delta(C)$
For *asymptotic* tail dependence, it is desirable that the tail dependence measure yields the same result for the extreme value copula C as for the copula C_F. That is,

$$C_F \in MDA(C) \Rightarrow \delta(C_F) = \delta(C).$$

Another desirable property would be the reverse relationship,

$$\delta(C_F) = \delta(C) \Rightarrow C_F \in MDA(C),$$

since this would imply that the extreme value copula C of C_F could be identified based on the tail dependence measure $\delta(C_F)$. This is a more general case of identification properties (P5) and (P6) mentioned above. However, since infinitely many tail dependent extreme value copulas exist, this is in general infeasible for tail dependent cases.

4.2 Tail dependence coefficients

4.2.1 The TDC for bivariate data

The concept of bivariate dependence between extremes was developed mainly by Geffroy (1958) and Sibuya (1960), although not through the language of copulas. The tail dependence coefficient was first introduced in Sibuya (1960) for two random variables X_1 and X_2 sharing the same marginal distribution with right endpoint $z^* \in \overline{\mathbb{R}}$ as

$$\chi = \lim_{z \to z^*} \mathbb{P}\left(X_2 > z | X_1 > z\right).$$

It represents the conditional probability that X_2 is extreme given that X_1 is extreme. To evaluate the tail dependence strength between two variables that do not share the same marginal distribution, the marginal behavior of the variables has to be removed. Joe (1993) reformulate the definition of the tail dependence coefficient (TDC) in terms of copulas as follows.

Definition 4.1. *The tail dependence coefficient (TDC) for two random variables X_1 and X_2 with marginal distributions F_1 and F_2, joint by a copula C_F, is defined as*

$$\lambda(C_F) = \lim_{u \uparrow 1} \mathbb{P}\left(F_2(X_2) > u | F_1(X_1) > u\right) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u)}{1-u}, \tag{4.2.1}$$

where $\overline{C}_F$ denotes the survival function of C_F.

Intuitively, the copula-based TDC represents the conditional probability that X_2 is extreme relative to its own distribution F_2, given that X_1 is extreme relative to its own distribution F_1. Since the definition of the TDC can be formulated in terms of the copula only, it can be concluded that the TDC does not depend on the marginal distributions. Furthermore, it can be shown that the TDC of a copula C_F is the same as the TDC of its extreme value copula C. This is captured by Theorem 4.1.

Theorem 4.1. *Let C_F be a bivariate copula with tail dependence coefficient $\lambda(C_F)$. If $C_F \in MDA(C)$ for an extreme value copula C with tail dependence coefficient $\lambda(C)$, then $\lambda(C_F) = \lambda(C)$. That is, the tail dependence coefficients of the copula C_F and its extreme value copula C are the same.*

Proof. Using the inclusion-exclusion principle (Equation 2.1.6) and that $\log(x) \sim x - 1$ as $x \to 1$, it can be derived that

$$\lambda(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u)}{1-u} = \lim_{u \uparrow 1} \frac{1 - 2u + C_F(u,u)}{1-u} = 2 - \lim_{u \uparrow 1} \frac{1 - C_F(u,u)}{1-u} = 2 - \lim_{u \uparrow 1} \frac{\log C_F(u,u)}{\log u}.$$

Using this expression for the tail dependence coefficient and the definition of the extreme value copula (Equation 3.3.3), the statement can be deduced as follows:

$$\begin{aligned}
\lambda(C) &= 2 - \lim_{u \uparrow 1} \frac{\log C(u,u)}{\log u} = 2 - \lim_{u \uparrow 1} \lim_{n \to \infty} \frac{\log C_F\left(u^{1/n}, u^{1/n}\right)^n}{\log u} \\
&= 2 - \lim_{u \uparrow 1} \lim_{n \to \infty} \frac{\log C_F\left(u^{1/n}, u^{1/n}\right)}{1/n \log u} = 2 - \lim_{u \uparrow 1} \lim_{n \to \infty} \frac{\log C_F\left(u^{1/n}, u^{1/n}\right)}{\log u^{1/n}} \\
&= 2 - \lim_{u \uparrow 1} \lim_{t \downarrow 0} \frac{\log C_F\left(u^t, u^t\right)}{\log u^t} = 2 - \lim_{w \uparrow 1} \frac{\log C_F(w,w)}{\log w} = \lambda(C_F).
\end{aligned}$$

■

The result implies that in order to determine the TDC of a certain copula C_F, it is sufficient to know by which extreme value copula C the copula C_F is attracted and what the TDC of the extreme value copula is. Note that this does not imply that the extreme value copula of C_F can be identified based on the TDC of C_F since the TDC of an extreme value copula is not necessarily unique for that specific extreme value copula.

Since the TDC summarizes the extremal dependence structure, it is not surprising that it is related to the characterizations of the multivariate tail dependence structure considered in Chapter 3. For example, the TDC is equal to the bivariate tail copula b evaluated in the point $(1, 1)$:

$$\lambda(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u, u)}{1 - u} = \lim_{t \downarrow 0} \frac{\overline{C}_F(1 - t, 1 - t)}{t} = b(1, 1). \tag{4.2.2}$$

The TDC is also related to the bivariate stable tail dependence function and, consequently, the bivariate Pickands dependence function as follows,

$$\lambda(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u, u)}{1 - u} = 2 - \lim_{u \uparrow 1} \frac{1 - C_F(u, u)}{1 - u} = 2 - \ell(1, 1) = 2\left(1 - A(1/2)\right). \tag{4.2.3}$$

That is, the TDC is a rescaled version of the STDF evaluated in the point $(1, 1)$ and a rescaled version of the Pickands dependence function evaluated in the point $1/2$. Due to this relationship with the Pickands dependence function, the TDC can be seen as a point on the range of the Pickands dependence function evaluated in the point $1/2$ (see the left panel in Figure 4.2.1).

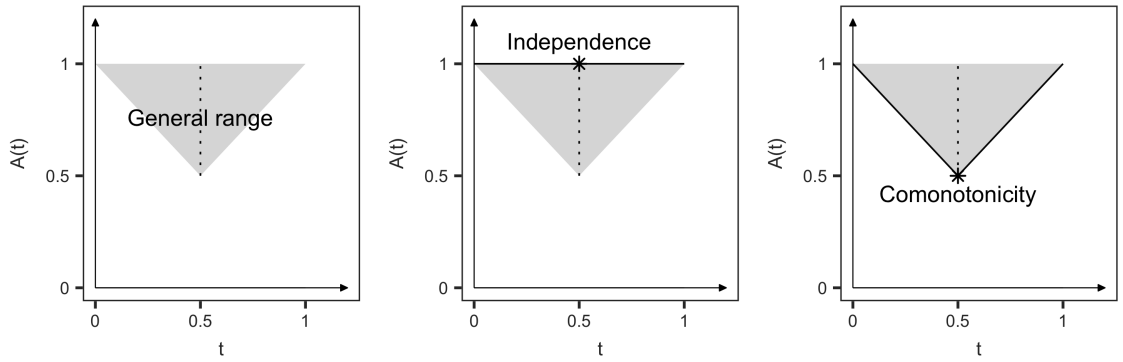

Figure 4.2.1: Evaluation of the Pickands dependence function in $1/2$.

By the relationship of the TDC with the dependence functions, it follows that in the tail independent case, the TDC is given by $\lambda(\Pi) = 2(1 - 1) = 0$. Similarly, for the comonotonic case, the TDC is equal to $\lambda(M) = 2(1 - 0.5) = 1$. Both of these are desirable properties. The scenarios are depicted in the middle and right-hand-side panels of Figure 4.2.1. Even better, it can be shown that the bivariate TDC has the desirable property that it is equal to zero if *and only if* X_1 and X_2 are tail independent (Main Theorem 1). Intuitively, this can be seen from Figure 4.2.1 as follows: since the Pickands dependence function is convex, it is only possible that $A(1/2) = 1$ (and hence $\lambda = 2(1 - A(1/2)) = 0$) if $A(t) = 1$ for all $t \in [0, 1]$, representing exactly the case of asymptotic independence. The result is captured in the following main theorem.

Main Theorem 1. *Let X_1 and X_2 be two continuous random variables joint with a copula C_F that belongs to the maximum domain of attraction of an extreme value copula. The variables X_1 and X_2 are asymptotically independent if and only if their tail dependence coefficient $\lambda(C_F)$ is equal to zero. That is,*

$$C_F \in MDA(\Pi) \Leftrightarrow \lambda(C_F) = 0.$$

Proof. Recall that the independence bound of the bivariate STDF equals $\ell(x_1, x_2) = x_1 + x_2$. Hence, by the relationship between the TDC and the stable tail dependence function (Equation 4.2.3) it follows immediately that if $C_F \in MDA(\Pi)$, then $\lambda(C_F) = 2 - \ell(1, 1) = 2 - (1 + 1) = 0$. Conversely, we have to show that if $\lambda(C_F) = 0$, then $C_F \in MDA(\Pi)$. By the relationship between the extreme value copula and the STDF (Proposition 3.5), it is sufficient to show that if $\lambda(C_F) = 0$, then $\ell(x_1, x_2) = x_1 + x_2$ for any $x_1, x_2 \in \mathbb{R}_+$. Alternatively, by using the restriction of the ℓ-function to the unit simplex Δ_d, i.e., the Pickands dependence function A, it is sufficient to show that if $\lambda(C_F) = 0$, then $A(t) = A(t, 1 - t) = 1$ for any $t \in [0, 1]$. Recall from Section 3.4.2 that the A-function is convex and is equal to 1 if it is evaluated in one of the unit vectors. Hence, $A(1, 0) = A(0, 1) = 1$. Note that the unit vectors correspond to the bounds of the domain of A. Since $A(t) \leq 1$, $t \in [0, 1]$, this implies that the Pickands dependence function attains maxima

on the boundaries of its domain, which is a known property of convex functions. Moreover, the maximum principle of convex functions states that if a convex function attains a global maximum at an interior point of its domain, the function must be constant (see, e.g., Theorem 3.4.6 in Niculescu and Persson (2004)). Since $\lambda(C_F) = 0$ implies that $A(1/2) = 1$, the maximum principle yields that A is constant. We find that if $\lambda(C_F) = 0$, then $A(t) = A(t, 1-t) = 1$ for any $t \in [0, 1]$, which concludes the proof. ■

It is also possible to prove the result without making use of the maximum principle for convex functions. This (more tedious) approach is shown in Appendix B.3.1. Similarly, it can be shown that the bivariate TDC is equal to 1 if and only if X_1 and X_2 are asymptotically comonotonic. Intuitively, this can be seen from Figure 4.2.1 as follows: since the Pickands dependence function is convex, the function values have to be smaller than or equal to the straight line joining the points $t = 0$ and $t = 0.5$. However, because of the bounds on the A-function, the function values cannot be below this straight line segment. Hence, the A-function is equal to the straight line segment. A similar argument holds for the line segment joining the point $t = 0.5$ and $t = 1$. The result is given by the theorem below.

Theorem 4.2. *Let X_1 and X_2 be two continuous random variables joint with a copula C_F that belongs to the maximum domain of attraction of an extreme value copula. The variables X_1 and X_2 are asymptotically comonotonic if and only if their tail dependence coefficient $\lambda(C_F)$ is equal to 1. That is,*

$$C_F \in MDA(M) \Leftrightarrow \lambda(C_F) = 1.$$

Proof. If $C_F \in MDA(M)$, the STDF is equal to $\ell(x_1, x_2) = \max(x_1, x_2)$. By the relationship between the TDC and the STDF (Equation 4.2.3), it follows immediately that in this case $\lambda(C_F) = 2 - \ell(1, 1) = 2 - 1 = 1$. Conversely, we have to show that if $\lambda(C_F) = 1$, then $C_F \in MDA(M)$. By the relationship between the extreme value copula and the STDF, it is sufficient to show that if $\lambda(C_F) = 1$, then $\ell(x_1, x_2) = \max(x_1, x_2)$ for any $x_1, x_2 \in \mathbb{R}_+$. Alternatively, by using the restriction of the ℓ-function to the unit simplex Δ_d, i.e., the Pickands dependence function A, it is sufficient to show that if $\lambda(C_F) = 1$, then $A(t) = A(t, 1-t) = \max(t, 1-t)$ for any $t \in [0, 1]$. Using the properties of the A-function, this result is shown with the following steps.

- Because of the bounds $\max(t, 1-t) \leq A(t) \leq 1$, $t \in [0, 1]$, we establish that $A(1) = A(0) = 1$. Since these bounds imply that $A(t) \geq \max(t, 1-t)$ for all $t \in [0, 1]$, it is sufficient to show that $A(t) \leq \max(t, 1-t)$ for all $t \in [0, 1]$ in order to prove that $A(t) = \max(t, 1-t)$ for all $t \in [0, 1]$.
- We assume that $\lambda(C_F) = 1$, which corresponds to $A(1/2) = 1/2$.
- The convexity of A implies that for all $t_1, t_2 \in [0, 1]$ and $\alpha \in [0, 1]$, the following defining relationship holds,

$$A\left(\alpha t_1 + (1-\alpha)t_2\right) \leq \alpha A(t_1) + (1-\alpha)A(t_2).$$

- As a special case, consider specifying α such that $\alpha t_1 + (1-\alpha)t_2 = t$ for any $t \in [0, 1]$, with $t_1 \in \{0, 1\}$ and $t_2 = 1/2$.
  - If $t \in (0, 1/2)$, we find by putting $t_1 = 0$ and $\alpha = 1 - 2t \in (0, 1)$ that

$$A(t) \leq \alpha A(0) + (1-\alpha)A(1/2) = (1-2t) + \frac{2t}{2} = 1 - t = \max(t, 1-t).$$

  - Similarly, if $t \in (1/2, 1)$, we find by putting $t_1 = 1$ and $\alpha = 2t - 1 \in (0, 1)$ that

$$A(t) \leq \alpha A(1) + (1-\alpha)A(1/2) = (2t-1) + \frac{2-2t}{2} = t = \max(t, 1-t).$$

- Hence, for any $t \in [0, 1]$ we conclude that $\max(t, 1-t) \leq A(t) \leq \max(t, 1-t)$, indicating that $A(t) = \max(t, 1-t)$. This is what needed to be shown to finish the proof.

■

In summary, the TDC for bivariate data is copula-based, can identify both independence and comonotonicity and is bounded between zero and one. It satisfies all desirable properties introduced in Section 4.1. Moreover, the TDC is directly connected to the characterizations of the tail dependence structure that were discussed in Chapter 3 and therefore measures *asymptotic* tail dependence. An overview of TDC values for a wide range of copulas is given by Heffernan (2000) and Joe (2015) for example. For an overview of the upper and lower tail coefficients of more Archimedean copulas, see Charpentier and Segers (2009). A sample of these values is shown in Table 4.2.1.

Table 4.2.1: Selection of extreme value copulas, several copulas that belong to their maximum domain of attraction and the corresponding tail dependence coefficients.

Extreme value copula C	Copula C_F	TDC λ
Comonotonicity M	Comonotonicity	1
Independence Π	Independence	0
	Gaussian	0
	Frank	0
	Clayton	0
Gumbel	Gumbel	$2-2^{1/\theta}$
	Joe	$2-2^{1/\theta}$
	BB1	$2-2^{1/\theta}$
Galambos	Galambos	$2^{-1/\theta}$
	BB4	$2^{-1/\delta}$
t-EV	Student-t	$2T_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}}\right)$
Hüsler-Reiss	Gaussian ($\rho_n \to 1$)	$2-2\Phi(\delta^{-1})$
BB5	BB5	$2-\left(2-2^{-1/\delta}\right)^{1/\theta}$

Based on the connection with the STDF, it is straightforward to assess the TDC for several bivariate copulas by evaluating the STDF in the point $(1,1)$ or the Pickands dependence function in the point $(1/2,1/2)$. Alternatively, the TDC for the Gumbel copula can be easily derived analytically as follows.

Example 4.1. *For the bivariate Gumbel copula with parameter $\theta \geq 1$,*

$$C_\theta^{Gu}(u_1,u_2) = \exp\left[-\left((-\log(u_1))^\theta + (-\log(u_2))^\theta\right)^{1/\theta}\right],$$

it can be inferred that

$$\begin{aligned}\lambda(C_\theta^{Gu}) &= 2-\lim_{u\uparrow 1}\frac{\log C_\theta^{Gu}(u,u)}{\log u} = 2-\lim_{u\uparrow 1}\frac{-\left((-\log(u))^\theta+(-\log(u))^\theta\right)^{1/\theta}}{\log u}\\ &= 2-\lim_{u\uparrow 1}\frac{-\left(2(-\log(u))^\theta\right)^{1/\theta}}{\log u} = 2-2^{1/\theta}.\end{aligned}$$

Based on Theorem 4.2.1 it can be concluded that all copulas that belong to the maximum domain of attraction of the Gumbel copula have a TDC equal to $2-2^{1/\theta}$. Theorem 3.5 or Theorem 3.9 can be used to verify whether a given Archimedean copula belongs to the maximum domain of attraction of the Gumbel copula.

For elliptical copulas, Schmidt (2002) shows that the tail behavior of the copula is determined by the tail behavior of the distribution function of the generating variable R. Specifically, let $\mathbf{X} = (X_1,\ldots,X_d)$ be spherically distributed with stochastic representation $\mathbf{X} \overset{d}{=} R\mathbf{U}$ (see Section 2.2). Suppose that F_R, the distribution function of R, has a regularly varying tail. Then all

bivariate margins have the tail dependence property and the tail dependence coefficient λ for an elliptically contoured random vector (X_1, X_2) with 2×2-dimensional positive definite covariance matrix $\mathbf{\Sigma}$, and with regular varying generating variable with index $-\alpha < 0$, is given by

$$\lambda = \frac{\int_0^{h(\rho)} \frac{u^\alpha}{\sqrt{1-u^2}} \mathrm{d}u}{\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}} \mathrm{d}u},$$

where $h(\rho) := \left(1 + \frac{(1-\rho)^2}{1-\rho^2}\right)^{-1/2}$ and where ρ denotes the correlation coefficient between the variables X_1 and X_2. Note that the TDC only depends on the correlation coefficient ρ and the regular variation index α. Alternatively, the TDC for the t-EV extreme value copula can be found by evaluating its Pickands dependence in the point $(1/2, 1/2)$ and rescaling it (Nikoloulopoulos et al., 2009). That is,

$$\begin{aligned}\lambda &= 2(1 - A(1/2, 1/2)) = 2\left(1 - T_{\nu+1}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}}(1-\rho)\right)\right) \\ &= 2T_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}}\right),\end{aligned}$$

with T_ν the univariate cumulative distribution function of the t-distribution with ν degrees of freedom and ρ the correlation between the variables X_1 and X_2. The TDC for the Hüsler-Reiss model can be determined similarly.

For vine copulas it can be of interest to assess the bivariate TDC for pairs of variables that are not directly linked through a baseline example, consider a 3-dimensional D-vine with linking copula (see Section 2.4). For two tail dependent Archimedean baseline linking copulas c_{12} and c_{23} with model parameters that are not necessarily the same. The conditional copula on the second level $c_{13|2}$ is chosen to be either a Gumbel copula or the independence copula in order to investigate the possible dependence strengths for between the variables X_1 and X_3. For the two Archimedean linking copulas, the bivariate TDCs are known. However, the bivariate TDC describing the tail dependence strength between X_1 and X_3 is not directly observable. Based on Proposition 3.10 and the relationship between the TDC and the tail copula (Equation 4.2.2), the TDC λ_{13} can be determined as follows,

$$\lambda_{13} = b_{13}(1,1) = \int_0^1 \int_0^1 \Upsilon_{13}(w_1, w_3)\mathrm{d}w_1\mathrm{d}w_3 = \int_0^1 \int_0^\infty \int_0^1 \Upsilon_{123}(w_1, w_2, w_3)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3.$$

Hence, by plugging in the tail density for a trivariate D-vine copula (Theorem 3.7),

$$\begin{aligned}\lambda_{13} = \int_0^1 \int_0^\infty \int_0^1 &\Upsilon_{\{1,2\}}(x_1, x_2)\Upsilon_{\{2,3\}}(x_2, x_3) \\ &\cdot c_{13|2}\left(1 - \int_0^1 \Upsilon_{\{1,2\}}(w_1, x_2)\mathrm{d}w_1, 1 - \int_0^1 \Upsilon_{\{2,3\}}(x_2, w_3)\mathrm{d}w_3\right)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3.\end{aligned}$$

As explained in Section 3.4.1 these expressions can be numerically evaluated in `R`. Results are shown in Table 4.2.2. Similar to findings of Joe et al. (2010), we find that vine copulas can have a flexible range of tail dependence with different tail dependence strengths for each bivariate margin. Moreover, the results show that in order for a vine copula to have tail dependence for all bivariate margins, it is only necessary for the bivariate copulas in the baseline linking level to have tail dependence and it is not necessary for the conditional bivariate copulas in level $2, \ldots, d-1$ to have tail dependence. For example, the D-vines with two tail dependent Archimedean copulas as baseline linking copulas c_{12} and c_{23} also exhibits tail dependence for the bivariate margin of c_{13} even when the conditional copula $c_{13|2}$ is specified to be tail independent.

Table 4.2.2: Numerical evaluation of the unknown bivariate TDC in a 3-dimensional D-vine.

c_{12}	λ_{12}	c_{23}	λ_{23}	$c_{13\|2}$	λ_{13}
Gumbel, $\theta = 2.56$ (Strong dep.)	0.690	Gumbel, $\theta = 1.54$ (Medium dep.)	0.431	Gumbel, $\theta = 2.56$ (Strong dep.)	0.641
Gumbel, $\theta = 2.56$ (Strong dep.)	0.690	Gumbel, $\theta = 1.54$ (Medium dep.)	0.431	Independence	0.369
Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 2.56$ (Strong dep.)	0.377
Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Independence	0.052

4.2.2 Tail copula extensions

A first multivariate extension of the bivariate tail dependence coefficient is based on the tail copula. It is defined as follows.[1]

Definition 4.2. *For a vector of random variables $\mathbf{X} = (X_1, \dots, X_d)$ with marginal distributions $F_1, \dots, F_d$, joint distribution function F, and copula C_F, the tail copula-based multivariate tail dependence coefficient τ is given by*

$$\tau(C_F) = \lim_{u \uparrow 1} \mathbb{P}\left(F_1(X_1) > u, \dots, F_d(X_d) > u | F_d(X_d) > u\right) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u, \dots, u)}{1-u}. \tag{4.2.4}$$

The tail copula-based multivariate TDC τ measures the probability that all variables $X_1, \dots, X_d$ are extreme given that one of them is extreme. This extension of the TDC is therefore sometimes called the upper orthant tail dependence coefficient. Similar as for the bivariate TDC, the multivariate TDC is equal to the tail copula evaluated in $(1, \dots, 1)$:

$$\tau(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u, \dots, u)}{1-u} = \lim_{t \downarrow 0} \frac{\overline{C}_F(1-t, 1-t)}{t} = b(1, \dots, 1).$$

In the definition above, the conditioning variable is taken to be X_d, but it does not matter on which variable the probability is conditioned since the transformed marginals are all uniforms. It is questionable whether this irrelevance of the specific conditioning variable is reasonable for a multivariate tail dependence measure. The following example illustrates the problem.

Example 4.2. *Consider the theoretical example for three variables (U_1, U_2, U_3) with copula C_F where U_1 is independent of U_2 and where U_2 and U_3 are comonotonic,*

$$U_1 \perp U_2, \quad U_1 \perp U_3, \quad U_2 = U_3.$$

In this case, the variables are joint by the following copula,

$$C_F(u_1, u_2, u_3) = u_1 \min(u_2, u_3).$$

Using the inclusion-exclusion formula, we find that the multivariate tail dependence coefficient is

[1] The tail copula-based multivariate TDC τ should not be confused with Kendalls tau which is an often used dependence measure for copulas.

equal to zero,

$$\begin{aligned}\tau(C_F) &= \lim_{u\uparrow 1}\frac{\overline{C}_F(u,u,u)}{1-u} = \lim_{u\uparrow 1}\frac{1-\sum_{\emptyset\neq S\subseteq\{1,2,3\}}(-1)^{|S|+1}\mathbb{P}\left(\cap_{i\in S}\{U_i\leq u\}\right)}{1-u}\\ &= \lim_{u\uparrow 1}\frac{1-\Big(C_F(u,1,1)+C_F(1,u,1)+C_F(1,1,u)-C_F(u,u,1)\Big)}{1-u} + \\ &\qquad\qquad \lim_{u\uparrow 1}\frac{1-\Big(-C_F(u,1,u)-C_F(1,u,u)+C_F(u,u,u)\Big)}{1-u}\\ &= \lim_{u\uparrow 1}\frac{1-\left(3u-2u^2-u+u^2\right)}{1-u} = \lim_{u\uparrow 1}\frac{1-2u+u^2}{1-u} = \lim_{u\uparrow 1}\frac{(1-u)^2}{1-u} = \lim_{u\uparrow 1}-(1-u) = 0.\end{aligned}$$

The example shows that the multivariate TDC τ is equal to zero for the variables U_1, U_2 and U_3 regardless of the conditioning variable. However, the comonotonic variables U_2 and U_3 contain more information than the independent variable U_1. That is, conditional on knowing U_2 the value of U_3 is also known (and vice versa) but conditional on knowing U_1 we are not informed on the other variables. It would be desirable for this additional information to be reflected in the multivariate tail dependence coefficient.

Another problem with the multivariate TDC τ can be identified in Example 4.2. It turns out that the copula $C_F(u_1,u_2,u_3) = u_1\min(u_2,u_3)$ from the example is in fact an extreme value copula and lies in its own domain of attraction because it satisfies the max-stability condition (Equation 3.3.4):

$$\lim_{n\to\infty} C_F\left(u_1^{1/n},u_2^{1/n},u_3^{1/n}\right)^n = \lim_{n\to\infty}\left(u_1^{1/n}\min\left(u_2^{1/n},u_3^{1/n}\right)\right)^n = u_1\min(u_2,u_3) = C_F(u_1,u_2,u_3).$$

This illustrates that the multivariate TDC of a copula C_F can be zero when this copula does not belong to the domain of attraction of the independence copula, implying that

$$\tau(C_F) = 0 \not\Rightarrow C_F \in MDA(\Pi).$$

That is, the identifying relationship no longer holds. In fact, it can be shown that if one pair of the random variables $(X_1,\ldots,X_d)$ is tail independent, then the TDC will be equal to zero, no matter how strong the asymptotic dependencies between the other variables may be: assuming without loss of generality that X_1 and X_2 are tail independent,

$$\begin{aligned}0\leq\tau(C_F) &= \lim_{u\uparrow 1}\frac{\overline{C}_F(u,\ldots,u)}{1-u} = \lim_{u\uparrow 1}\frac{\mathbb{P}(F_1(X_1)>u,F_2(X_2)>u,\ldots,F_d(X_d)>u)}{1-u}\\ &\leq \lim_{u\uparrow 1}\frac{\mathbb{P}(F_1(X_1)>u,F_2(X_2)>u)}{1-u} = \lim_{u\uparrow 1}\frac{(1-u)^2}{1-u} = 0.\end{aligned}$$

Hence, if one bivariate pair is tail independent, information on the tail dependence structure between the other variables is lost. Especially for higher dimensions, this limits the utility of this tail dependence measure. More generally, it can be seen that the weakest pairwise dependencies have the highest impact on the tail copula. Consider for example again the 3-dimensional D-vine with two tail dependent Archimedean baseline linking copulas c_{12} and c_{23} with model parameters that are not necessarily the same. The conditional copula on the second level $c_{13|2}$ is chosen to be either a Gumbel copula or the independence copula. Using the integrating procedure described in Section 3.4.1, τ is evaluated as follows,

$$\begin{aligned}\tau(C_F) = b(1,1,1) &= \int_0^1\int_0^1\int_0^1 \Upsilon_{12}(w_1,w_2)\Upsilon_{23}(w_2,w_3)\\ &\quad\cdot c_{13|2}\left(1-\int_0^{w_1}\Upsilon_{12}(\tilde{w}_1,w_2)\mathrm{d}\tilde{w}_1, 1-\int_0^{w_3}\Upsilon_{23}(w_2,\tilde{w}_3)\mathrm{d}\tilde{w}_3\right)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3.\end{aligned}$$

The results in Table 4.2.3 show that the multivariate TDC τ reflects the weakest bivariate tail dependence of the multivariate vine copula. This is not very informative and might lead to underestimation of the tail dependence strength in dimensions $d > 2$.

A possible solution to the loss of information on the tail dependence structure if one or more pairwise dependencies are weakly tail dependent or tail independent, is to assess the upper orthant

Table 4.2.3: Numerical evaluation of the tail copula-based multivariate TDC τ in a 3-dimensional D-vine.

c_{12}	λ_{12}	c_{23}	λ_{23}	$c_{13\|2}$	λ_{13}	τ_{123}
Gumbel, $\theta = 2.56$ (Strong dep.)	0.690	Gumbel, $\theta = 1.54$ (Medium dep.)	0.431	Gumbel, $\theta = 2.56$ (Strong dep.)	0.641	0.418
Gumbel, $\theta = 2.56$ (Strong dep.)	0.690	Gumbel, $\theta = 1.54$ (Medium dep.)	0.431	Independence	0.369	0.333
Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 2.56$ (Strong dep.)	0.377	0.159
Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Independence	0.052	0.044

TDCs of several subsets $S \subseteq \{1, \ldots, d\}$ of the variables $X_1, \ldots, X_d$. For example, by first assessing the upper orthant TDC τ for all bivariate pairs, which is effectively the same as determining the bivariate TDC λ, it can be determined whether the variables $(X_1, \ldots, X_d)$ are tail independent or tail independent. Note that this follows from Theorem 3.3 in combination with Main Theorem 1. Higher order tail dependencies can next be assessed by determining the multivariate TDC τ for other subsets of the variables. Another idea is to vary the set of conditioning variables (Li, 2009). This leads to a different orthant TDC, which we denote by τ_S, defined as

$$\tau_S(C_F) = \lim_{u \uparrow 1} \mathbb{P}\left(F_j(X_j) > u, \forall j \notin S | F_i(X_i) > u, \forall i \in S\right), \tag{4.2.5}$$

for a subset $\emptyset \neq S \subseteq \{1, \ldots, d\}$. In terms of copulas, the definition of the upper orthant tail dependence coefficient is given as

$$\tau_S(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u, \ldots, u)}{\overline{C}_{F;S}(u, \ldots, u)}, \tag{4.2.6}$$

where $\overline{C}_F$ denotes the survival function of C_F and where $C_{F;S}$ is the $|S|$-dimensional copula joining the variables $X_i : i \in S$. In words, this orthant TDC τ_S is the conditional probability that all variables in a certain subset of the variables under consideration are extreme given that all variables not belonging to that set are in fact extreme. Consider for example a portfolio of 3 stocks, each generating random losses captured by the random variables X_1, X_2 and X_3. The orthant TDC of subset $S = \{2, 3\}$ describes the probability that the loss coming from the first stock X_1 is extreme, given that the losses generated by the second and third stock are extreme. We revisit Example 4.2 to illustrate how the subset-based orthant TDC τ_S provides more information of the tail dependence structure than the regular orthant TDC τ.

Example 4.3. *Consider again the theoretical example for three variables (U_1, U_2, U_3) with copula C_F, where U_1 is independent of U_2 and where U_2 and U_3 are comonotonic,*

$$U_1 \perp U_2, \quad U_1 \perp U_3, \quad U_2 = U_3.$$

In this case, the variables are joint by the following copula,

$$C_F(u_1, u_2, u_3) = u_1 \min(u_2, u_3).$$

Since conditioning on one variable leads to the multivariate TDC measure of the previous section, we already know that $\tau_{\{1\}}(C_F) = \tau_{\{2\}}(C_F) = \tau_{\{3\}}(C_F) = 0$. Using the inclusion-exclusion formula and calculation results from Example 4.2, we find the values for the upper orthant TDC for the remaining subsets as follows,

$$\tau_{\{1,2\}}(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u,u)}{\overline{C}_{F;\{1,2\}}(u,u)} = \lim_{u \uparrow 1} \frac{(1-u)^2}{1 - 2u + C_F(u,u,1)} = \lim_{u \uparrow 1} \frac{(1-u)^2}{1 - 2u + u^2} = \lim_{u \uparrow 1} \frac{(u-1)^2}{(u-1)^2} = 1,$$

$$\tau_{\{1,3\}}(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u,u)}{\overline{C}_{F;\{1,3\}}(u,u)} = \lim_{u \uparrow 1} \frac{(1-u)^2}{1 - 2u + C_F(u,1,u)} = \lim_{u \uparrow 1} \frac{(1-u)^2}{1 - 2u + u^2} = \lim_{u \uparrow 1} \frac{(u-1)^2}{(u-1)^2} = 1,$$

$$\tau_{\{2,3\}}(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u,u)}{\overline{C}_{F;\{2,3\}}(u,u)} = \lim_{u \uparrow 1} \frac{(1-u)^2}{1 - 2u + C_F(1,u,u)} = \lim_{u \uparrow 1} \frac{(1-u)^2}{1 - 2u + u} = \lim_{u \uparrow 1} \frac{(u-1)^2}{1-u} = 0.$$

Since there exist subsets $S \subset \{1,2,3\}$ for which $\tau_S(C_F) \neq 0$, the variables U_1, U_2, and U_3 are said to be upper-orthant tail dependent (Li, 2009).

Although the subset-based orthant TDC τ_S can convey more information than τ, a large drawback is presented by the fact that multiple measures have to be computed to capture the nature of the extremal dependence structure. Especially for higher dimensions it becomes a tedious task to assess the orthant TDC for all possible subsets and to draw conclusions from all of these individual measures on the general structure of the tail dependence structure. For example, for dimension $d = 3$ there are 6 TDC measures to evaluate; for dimension $d = 5$, this approach leads to 30 upper orthant TDC measures to evaluate and interpret.

4.2.3 Stable tail dependence function extension

The previous section extended the bivariate TDC λ based on the relationship with the tail copula. In this section, the extension of the TDC λ based on its relationship with the stable tail dependence function is researched. Since the bivariate TDC λ is a rescaled version of the STDF evaluated in $(1,1)$, a multivariate TDC Λ^* could be based on the d-dimensional STDF evaluated in $(1,\ldots,1)$,

$$\Lambda^* = \ell(1,\ldots,1) = \lim_{u\uparrow 1} \frac{1 - C_F(u,\ldots,u)}{1-u} = -\log C(e^{-1},\ldots,e^{-1}) = dA\left(1/d,\ldots,1/d\right).$$

It turns out that this extension leads to the definition of the so-called extremal coefficients, discussed by Schlather and Tawn (2003), for example. The STDF evaluated in the point $(1,\ldots,1)$ can be interpreted as the probability that one or more components are extreme, given that one component is extreme. Kiriliouk (2016) shows that the resulting number can be interpreted as the effective number of tail independent variables among $X_1,\ldots,X_d$.

Because of the bounds of the STDF, this quantity will be bounded as $1 \leq \Lambda^* \leq d$ for the comonotonic case and the independent case, respectively. By scaling the measure as follows,

$$\Lambda(C_F) = \frac{\Lambda^* - d}{1-d} = \frac{d - \Lambda^*}{d-1},$$

the measure will now be bounded between $0 \leq \Lambda \leq 1$ for the independent case and the comonotonic case, respectively. We therefore define the STDF-based multivariate TDC as follows.

Definition 4.3. *Let $\mathbf{X} = (X_1,\ldots,X_d)$ be a d-dimensional random vector with copula C_F that belongs to the domain of attraction of an extreme value copula and with stable tail dependence function ℓ. The multivariate STDF-based TDC Λ is then defined as*

$$\Lambda(C_F) = \frac{d - \ell(1,\ldots,1)}{d-1}. \tag{4.2.7}$$

Note that for $d = 2$, we have $\Lambda(C_F) = 2 - \ell(1,1)$, which is exactly the definition for the bivariate TDC. In order to explore how this multivariate tail dependence measure behaves, we consider Example 4.2 one last time.

Example 4.4. *Consider the theoretical example for three variables (U_1, U_2, U_3) with copula C_F, where U_1 is independent of U_2 and where U_2 and U_3 are comonotonic,*

$$U_1 \perp U_2, \quad U_1 \perp U_3, \quad U_2 = U_3.$$

In this case, the variables are joint by the following copula,

$$C_F(u_1,u_2,u_3) = u_1 \min(u_2,u_3).$$

Since

$$\Lambda^* = \ell(1,1,1) = \lim_{u\uparrow 1} \frac{1 - C_F(u,\ldots,u)}{1-u} = \lim_{u\uparrow 1} \frac{1-u^2}{1-u} = \lim_{u\uparrow 1} \frac{(1-u)(1+u)}{1-u} = 2,$$

the STDF-based TDC gives that $\Lambda = (2-3)/(1-3) = 1/2$, indicating that the dependence structure is between independence ($\Lambda = 0$) and comonotonicity ($\Lambda = 1$).

Since the STDF of a copula C_F is the same as its extreme value copula C, if ℓ exists and if $C_F \in MDA(C)$ (see Proposition 3.4), it is straightforward to establish that $\Lambda(C_F) = \Lambda(C)$. Formally, this is captured by the following proposition.

Proposition 4.1. *Let C_F be a d-dimensional copula with multivariate tail dependence coefficient $\Lambda(C_F)$. If $C_F \in MDA(C)$ for an extreme value copula C with multivariate tail dependence coefficient $\Lambda(C)$, then $\Lambda(C_F) = \Lambda(C)$. That is, the multivariate tail dependence coefficients of the copula C_F and its extreme value copula C are the same.*

Proof. The result follows immediately from the STDF-based definition of the multivariate tail dependence coefficient and the notion that the stable tail dependence function ℓ is the same for C_F and C if $C_F \in MDA(C)$ (Proposition 3.4). ■

Furthermore, it turns out that this multivariate TDC Λ has the ability to identify tail independence, just as the bivariate TDC (Main Theorem 1). This is a crucial observation which has not yet been made in literature to our knowledge. It is captured by Main Theorem 2.

Main Theorem 2. *Let $X_1, \ldots, X_d$ be continuous random variables joint with a copula C_F that belongs to the maximum domain of attraction of an extreme value copula. The variables $X_1, \ldots, X_d$ are asymptotically independent if and only if their STDF-based multivariate tail dependence coefficient $\Lambda(C_F)$ is equal to zero. That is,*

$$C_F \in MDA(\Pi_d) \Leftrightarrow \Lambda(C_F) = 0.$$

Proof. If $C_F \in MDA(\Pi_d)$, the STDF is equal to $\ell(x_1, \ldots, x_d) = x_1 + \cdots + x_d$. Hence, in that case it follows immediately that

$$\Lambda(C_F) = (d - \ell(1, \ldots, 1)/(d-1) = (d-d)/(d-1) = 0.$$

Conversely, we have to show that if $\Lambda(C_F) = 0$, then $C_F \in MDA(\Pi_d)$. By the relationship between the extreme value copula and the STDF (Proposition 3.5), it is sufficient to show that if $\Lambda(C_F) = 0$, then $\ell(\mathbf{x}) = x_1 + \cdots + x_d$ for any $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}_+^d$. Alternatively, by using the Pickands representation, it is sufficient to show that if $\Lambda(C_F) = 0$, then $A(\mathbf{x}) = 1$ for any $\mathbf{x} \in \Delta_d$. Recall from Section 3.4.2 that the A-function is convex and is equal to 1 if it is evaluated in one of the unit vectors: $A(\mathbf{e}_j) = 1$ for $\mathbf{e}_j$ the j-th unit vector, $1 \leq j \leq d$. Note that the unit vectors correspond to the extreme points of the unit simplex.[2] Since $A(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \Delta_d$, this implies that the Pickands dependence function attains maxima on the extreme points of its domain. Similar as for one-dimensional convex functions, this is a known property: a multivariate convex function attains its supremum at an extreme point of its convex domain (see, e.g., Theorem A.4.3 in Niculescu and Persson (2004)). Since $\Lambda(C_F) = 0$ implies that $A(1/d, \ldots, 1/d) = 1$, a maximum is also attained in the point $(1/d, \ldots, 1/d)$. However, unlike the maximum principle for one-dimensional convex functions, this does not immediately imply that the function is constant if it also attains its maximum at non-extreme points. Therefore, to finish the proof, we employ a strategy that does not make use of the maximum principle for convex functions. This approach is shown in Appendix B.3.2. ■

Since this multivariate TDC Λ is equivalent to a rescaled version of the STDF evaluated in the point $(1, \ldots, 1)$, it is straightforward to find Λ for several known copula models based on the expressions of the STDF that were derived in Section 3.4. For example, the Pickands dependence function for the t-EV copula evaluated in the point $(1/d, \ldots, 1/d)$ is given by,

$$A(1/d, \ldots, 1/d) = \sum_{j=1}^{d} \frac{1}{d} T_{d-1,\nu+1,\mathbf{R}_j} \left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho_{ij}^2}} \left(1 - \rho_{ij}\right), i \neq j \right),$$

[2] Extreme points of a set U are those points that are not an interior point of any linear segment in U. That is, if there do not exist $x, y \in U$ such that $z = \lambda x + (1-\lambda)y$ for $\lambda \in (0,1)$, then z is an extreme point of U (Niculescu and Persson, 2004).

where $\mathbf{R}_j$ is equal to the partial correlation matrix $\mathbf{R}$ with the j-th row and j-th column removed. That is, $\mathbf{R}_j = (\rho_{k_1,k_2;j})_{k_1,k_2\neq j}$. The multivariate TDC for the t-EV copula is therefore given by

$$\begin{aligned}\Lambda &= \frac{d-\ell(1,\ldots,1)}{d-1} = \frac{d-dA(1/d,\ldots,1/d)}{d-1} \\ &= \frac{1}{d-1}\left(d - \sum_{j=1}^{d} T_{d-1,\nu+1,\mathbf{R}_j}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho_{ij}^2}}(1-\rho_{ij}), i\neq j\right)\right).\end{aligned}$$

Similarly, Λ can be determined for the Hüsler-Reiss model. For Archimedean copulas, the ℓ-function immediately yields that $\ell(1,\ldots,1) = d^{1/\theta}$, with $\theta \geq 1$ the model parameter of the asymptotic Gumbel copula. Alternatively, a straightforward expression for Λ can be found for Archimedean copulas based on the behavior of their generator. That is, consider the ℓ-function evaluated in the point $(1,\ldots,1)$ for an Archimedean copula C_F with generator function ψ,

$$\ell(1,\ldots,1) = \lim_{u\uparrow 1}\frac{1-C_F(u,\ldots,u)}{1-u} = \lim_{u\uparrow 1}\frac{1-\psi\left(\psi^{-1}(u)+\cdots+\psi^{-1}(u)\right)}{1-u} = \lim_{x\downarrow 0}\frac{1-\psi(dx)}{1-\psi(x)},$$

where it is used that $\psi(0)=1$ and $\psi(1)=0$. It is implied by this expression that the multivariate extremal behavior of Archimedean copulas is determined by the regular variation of the function $1-\psi$ at 0. That is, if $1-\psi$ is regularly varying at 0 with index α, then

$$\ell(1,\ldots,1) = \lim_{x\downarrow 0}\frac{1-\psi(dx)}{1-\psi(x)} = d^{\alpha} \quad \Rightarrow \quad \Lambda = \frac{d-d^{\alpha}}{d-1}.$$

This, in turn, implies that for $\alpha = 1$, the Archimedean copula is asymptotically independent, while for $\alpha = 0$, the Archimedean copula is asymptotically comonotonic. Theorem 3.9 also leads to this conclusion (the results are comparable by taking $\theta = 1/\alpha$).

For vine copulas, the multivariate TDC Λ can be assessed based on the copula tail density of the D-vine (Theorem 3.7). Similar to finding the STDF function for vines, Λ can be determined as follows,

$$\ell(1,\ldots,1) = \sum_{\emptyset\neq S\subseteq\{1,\ldots,d\}}(-1)^{|S|-1}\int_0^1\cdots\int_0^1 \Upsilon_S(v_{i_1},\ldots,v_{i_{|S|}})\mathrm{d}v_{i_1}\ldots\mathrm{d}v_{i_{|S|}}.$$

(see Proposition 3.8). Table 4.2.2 and Table 4.2.3 already showed the bivariate TDC λ for the unspecified bivariate margin c_{13} and the multivariate TDC τ. To further illustrate the tail dependence behavior of a trivariate D-vine with Archimedean baseline linking copulas, the multivariate TDC Λ is also assessed. For a 3-dimension D-vine with baseline linking copulas c_{12} and c_{23}, with corresponding copula tail densities Υ_{12} and Υ_{23}, and first level conditional copula density $c_{13|2}$, the multivariate TDC Λ can be numerically evaluated based on the following expression,

$$\begin{aligned}\ell(1,1,1) &= 3 - \int_0^1\int_0^1 \Upsilon_{12}(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2 - \int_0^1\int_0^1 \Upsilon_{23}(w_2,w_3)\mathrm{d}w_2\mathrm{d}w_3 \\ &\quad - \int_0^1\int_0^1 \Upsilon_{13}(w_1,w_3)\mathrm{d}w_1\mathrm{d}w_3 + \int_0^1\int_0^1\int_0^1 \Upsilon_{123}(w_1,w_2,w_3)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3 \\ &= 3 - \int_0^1\int_0^1 \Upsilon_{12}(w_1,w_2)\mathrm{d}w_1\mathrm{d}w_2 - \int_0^1\int_0^1 \Upsilon_{23}(w_2,w_3)\mathrm{d}w_2\mathrm{d}w_3 \\ &\quad - \int_0^1\int_0^{\infty}\int_0^1 \Upsilon_{123}(w_1,w_2w_3)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3 \\ &\quad + \int_0^1\int_0^1\int_0^1 \Upsilon_{12}(w_1,w_2)\Upsilon_{23}(w_2,w_3) \\ &\qquad \cdot c_{13|2}\left(1-\int_0^{w_1}\Upsilon_{12}(\tilde{w}_1,w_2)\mathrm{d}\tilde{w}_1, 1-\int_0^{w_3}\Upsilon_{23}(w_2,\tilde{w}_3)\mathrm{d}\tilde{w}_3\right)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3.\end{aligned}$$

As discussed in Section 3.4.1, the semi-infinite bounds are transformed and the resulting integrals can be evaluated in `R` with the `adaptIntegrate` function. The results in Table 4.2.4 indicate that the multivariate TDC Λ provides an informative summary of the tail dependence strength in the vine.

Table 4.2.4: Numerical evaluation of the STDF-based multivariate TDC Λ in a 3-dimensional D-vine.

c_{12}	λ_{12}	c_{23}	λ_{23}	$c_{13\|2}$	λ_{13}	Λ_{123}
Gumbel, $\theta = 2.56$ (Strong dep.)	0.690	Gumbel, $\theta = 1.54$ (Medium dep.)	0.431	Gumbel, $\theta = 2.56$ (Strong dep.)	0.641	0.672
Gumbel, $\theta = 2.56$ (Strong dep.)	0.690	Gumbel, $\theta = 1.54$ (Medium dep.)	0.431	Independence	0.369	0.579
Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 2.56$ (Strong dep.)	0.377	0.293
Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Gumbel, $\theta = 1.16$ (Weak dep.)	0.185	Independence	0.052	0.190

4.3 Second order measures

The tail dependence coefficients are asymptotic dependence measures, and hence capture the limiting dependence structure of the extreme value copula to which a certain copula converges. However, depending on the copula C_F, the speed of convergence can vary significantly. Especially in the tail independent case, the TDC might not fully capture the strength of the extremal dependence for finite samples. For example, if $(X_1, X_2) \sim C_F$ are independent, then

$$\lambda(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u)}{1-u} = \lim_{u \uparrow 1} \frac{(1-u)^2}{1-u} = 0.$$

The tail dependence coefficient is also zero for any $1 < k \leq 2$

$$\lambda(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u)}{1-u} = \lim_{u \uparrow 1} \frac{(1-u)^k}{1-u} = 0,$$

but in these scenarios, the variables X_1 and X_2 are not independent. If the variables have not yet converged to Π, extreme values stemming from finite samples of C_F might still exhibit dependence. In order to account for this phenomenon, we need a second measure quantifying the amount of dependence that is left above high but finite thresholds. The second order measures introduced in this section assess the speed of convergence of the copula C_F to the independence copula Π for variables that are tail independent.

4.3.1 Tail order

The tail order is a first way to measure the residual dependence for asymptotically tail independent variables. This second order measure is based on the concept of regular variation (see Appendix A). Originally, the concept behind the tail order was presented for the bivariate case by Tawn and Ledford (1996) who introduced the so-called tail coefficient η for two random variables X_1 and X_2 with no right endpoint and joint distribution function F as follows,

$$\mathbb{P}(X_1 > z, X_2 > z) \sim \mathscr{L}(z) z^{-1/\eta} \text{ as } z \to \infty.$$

The power term $-1/\eta$ controls the speed of decay of the joint tail probability. Large values of η lead to a slower decay rate of the joint probability than small values of η. Hence, η measures the strength of dependence in the tails. After the initial work of Tawn and Ledford (1996), many others have researched this measure under different names. For example, De Haan and Zhou (2011) and Hashorva (2010) use the term residual dependence index. The tail order is defined by Hua and Joe (2011) as the reciprocal of the tail coefficient η. Using copulas and the tail order notation, $\kappa = 1/\eta$, the following multivariate definition of the tail order will be employed (Joe, 2015).

Definition 4.4. *Let $X_1, \ldots, X_d$ be random variables with marginal distributions $F_1, \ldots, F_d$, and copula C_F. If there exists a coefficient $\kappa(C_F) > 0$ such that, with a function $\mathscr{L}(u) \in \mathrm{RV}_0(0^+)$,*

$$\lim_{u \uparrow 1} \mathbb{P}(F_1(X_1) > u, \ldots, F_d(X_d) > u) = \lim_{u \uparrow 1} \overline{C}_F(u, \ldots, u) \sim (1-u)^{\kappa(C_F)} \mathscr{L}(1-u), \tag{4.3.1}$$

then we refer to $\kappa(C_F)$ as the tail order of C_F.

It is customary to assume that the tail order is bounded between perfect dependence and independence, i.e., $1 \leq \kappa \leq d$, since the multivariate extreme value distribution cannot contain negative dependence (see Section 3.1.2). However, since the tail order is concerned with the residual dependence structure in finite samples that converge to Π, negative dependence cannot be ruled out. If there is a negative association between variables at extreme but finite levels, the tail order might be larger than d.

For any bivariate copula C_F that can be written as Equation 4.3.1 with tail order $\kappa > 1$ and a function $\mathscr{L}(u)$ slowly varying at zero, the following holds:

$$\lambda(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,u)}{1-u} = \lim_{u \uparrow 1} \frac{(1-u)^\kappa \mathscr{L}(1-u)}{1-u} = 0.$$

Hence, a bivariate copula C_F with tail order $\kappa > 1$ is tail independent and therefore belongs to the maximum domain of attraction of the independence copula Π. For dimensions $d > 2$, the tail copula-based TDC τ is equal to zero,

$$\tau(C_F) = \lim_{u \uparrow 1} \frac{\overline{C}_F(u,\ldots,u)}{1-u} = \lim_{u \uparrow 1} \frac{(1-u)^\kappa \mathscr{L}(1-u)}{1-u} = 0,$$

but this does not necessarily imply that the copula C_F is tail independent (see Section 3.4.3). Hence, in contrast to the bivariate case, a d-dimensional copula C_F with $d > 2$ with tail order $\kappa > 1$ is not necessarily tail independent. On the other hand, for any dimension $d \geq 2$ it can be shown that a copula C_F with tail order $\kappa = 1$ is tail dependent. The tail copula-based multivariate TDC τ is in this case equal to $\tau(C_F) = \lim_{u \uparrow 1} \mathscr{L}(1-u)$.

The tail order and the slowly varying function can be determined analytically for several known copulas. For the d-dimensional independence copula it is easy to see that $\kappa = d$ with $\mathscr{L}(t) = 1$ since $\overline{C}_F(u,\ldots,u) = (1-u)^d$. Similarly, for the comonotonicity copula it holds that $\kappa = 1$ with $\mathscr{L}(t) = 1$, since $\overline{C}_F(u,\ldots,u) = 1-u$. To determine the tail order for other copulas, the computations are a bit more tedious. The following Example 4.5 shows that the d-dimensional Frank copula has a tail order equal to $\kappa = d$ with slowly varying function equal to $\mathscr{L}(u) = -\frac{1}{\delta}\left(e^{-\delta}-1\right)^{1-\delta}(-\delta)^d$. Hence, the Frank copula has the same tail order as the independence copula and therefore does not exhibit residual dependence.

Example 4.5. *Recall that the d-dimensional Frank copula is given by*

$$C(u_1,\ldots,u_d) = -\frac{1}{\delta}\log\left(1 + \left(e^{-\delta}-1\right)^{1-d}\prod_{i=1}^{d}\left(e^{-\delta u_i}-1\right)\right),$$

for $\delta \geq 0$ *(see Section 2.3). Because it is reflection symmetric,*

$$\lim_{u \uparrow 1} \overline{C}_F(u,\ldots,u) = \lim_{u \downarrow 0} C_F(u,\ldots,u).$$

Now, using that $\exp(-ux) \sim 1 - ux$ *as* $u \to 0$ *and that* $\log(x) \sim x - 1$ *for* $x \to 1$*, it follows that*

$$\begin{aligned}
\lim_{u \downarrow 0} C_F(u,\ldots,u) &= \lim_{u \downarrow 0} -\frac{1}{\delta}\log\left(1 + \left(e^{-\delta}-1\right)^{1-d}\left(e^{-\delta u}-1\right)^d\right) \\
&= \lim_{u \downarrow 0} -\frac{1}{\delta}\log\left(1 + \left(e^{-\delta}-1\right)^{1-d}(-\delta u)^d\right) \\
&= \lim_{u \downarrow 0} -\frac{1}{\delta}\left(e^{-\delta}-1\right)^{1-d}(-\delta u)^d,
\end{aligned}$$

hence,

$$C_F(u,\ldots,u) \sim -\frac{1}{\delta}\left(e^{-\delta}-1\right)^{1-d}(-\delta)^d \cdot u^d =: \mathscr{L}(u)u^\kappa,$$

with $\kappa = d$ *and* $\mathscr{L}(u) = -\frac{1}{\delta}\left(e^{-\delta}-1\right)^{1-d}(-\delta)^d$*. In the bivariate case* $(d = 2)$*, this simplifies to*

$$C_F(u,u) \sim \frac{\delta}{1-e^{-\delta}}u^2.$$

In contrast, the Normal copula does exhibit residual dependence. The following example shows that the tail order of the d-dimensional Normal copula with correlation matrix $\mathbf{\Sigma}$ is equal to $\kappa = \mathbf{1}_d \mathbf{\Sigma}^{-1} \mathbf{1}_d^T$. In the bivariate case, this simplifies to $\kappa = \rho$.

Example 4.6. *(Joe, 2015). Consider a d-dimensional Gaussian copula C_F with positive definite correlation matrix $\mathbf{\Sigma}$. Suppose that C_F satisfies $C_F(u, \ldots, u) \sim u^\kappa \mathscr{L}(u) = h^* u^\kappa (-\log u)^\zeta$, as $u \downarrow 0$, where h^* is a constant. Then, by the monotone density theorem, this is equivalent to $c(u, \ldots, u) \sim h u^{\kappa - d} (-\log u)^\zeta$, as $u \downarrow 0$, where h is another constant. Denoting with ϕ and Φ the Normal density and distribution function, we find that*

$$\begin{aligned} 1 &= \lim_{u \downarrow 0} \frac{c(u, \ldots, u)}{h u^{\kappa - d} (-\log u)^\zeta} = \lim_{u \downarrow 0} \frac{\phi_{\mathbf{\Sigma}} \left(\Phi^{-1}(u), \ldots, \Phi^{-1}(u) \right)}{h \phi^d \left(\Phi^{-1}(u) \right) u^{\kappa - d} (-\log u)^\zeta} \\ &= \lim_{z \to -\infty} \frac{\phi_{\mathbf{\Sigma}} (z, \ldots, z)}{h \phi^d(z) \left(\Phi(z) \right)^{\kappa - d} \left(-\log(\Phi(z)) \right)^\zeta} = \lim_{z \to -\infty} \frac{\phi_{\mathbf{\Sigma}}(z, \ldots, z)}{h \phi^\kappa(z) |z|^{d - \kappa} \left(-\log(\phi(z)/|z|) \right)^\zeta}. \end{aligned}$$

In the above calculations it is used that $\Phi(z) \sim \phi(z)/|z|$ as $z \to -\infty$. Since the exponent terms dominate the numerator and denominator, to cancel the exponent terms, a necessary condition is that $\kappa = \mathbf{1}_d \mathbf{\Sigma}^{-1} \mathbf{1}_d^T$, which turns out to be the tail order of the copula C. Furthermore, to cancel the term of $|z|$, we need that $d - \kappa + 2\zeta = 0$, so $\zeta = (\kappa - d)/2$.

An overview of the tail order of several other copulas can be found in Joe (2015) and Hua and Joe (2011). The issue of second order tail dependence structures in vines is still an unsolved issue in literature.

4.3.2 Asymptotic independence measure

Linked to the tail order is the asymptotic independence measure introduced by Coles et al. (1999). Just as the tail order, the asymptotic independence measure aims to capture the relative strength of dependence between variables that are asymptotic independent. Originally defined for two variables X_1 and X_2 sharing the same marginal distribution with right endpoint $z_* \in \overline{\mathbb{R}}$ as

$$\overline{\chi} = \lim_{z \to z^*} \frac{2 \log \mathbb{P}(X_1 > z)}{\log \mathbb{P}(X_1 > z, X_2 > z)} - 1,$$

the asymptotic independence measure assesses the rate at which $\mathbb{P}(X_2 > z | X_1 > z)$ approaches zero as $z \to z_*$. This can be established as follows. Suppose that as $z \to \infty$, the probabilities behave as $\mathbb{P}(X_1 > z) \sim z^{-k_1} \mathscr{L}(z)$ and $\mathbb{P}(X_1 > z, X_2 > z) \sim z^{-k_2} \mathscr{L}(z)$, with $k_1, k_2 > 0$. The parameters k_1 and k_2 control the rate at which the probabilities decay to zero and share a close resemblance to the tail order coefficient. Under these assumptions, it is implied that the conditional probability behaves as $\mathbb{P}(X_2 > z | X_1 > z) \sim z^{k_1 - k_2} \mathscr{L}(z)$ and that the asymptotic independence measure equals

$$\overline{\chi} = \lim_{z \to z^*} \frac{2 \log \left(z^{-k_1} \mathscr{L}(z) \right)}{\log \left(z^{-k_2} \mathscr{L}(z) \right)} - 1 \sim \frac{2k_1}{k_2} - 1.$$

Note that since $\mathbb{P}(X_1 > z) \geq \mathbb{P}(X_1 > z, X_2 > z)$ we must have that $k_1 \leq k_2$. The degree to which k_2 is larger than k_1 determines the tail dependence. The exact definition is based on a scaling argument. Assuming that $(X_1, X_2) \sim C_F$, the definition can be reformulated in terms of copulas as follows,

$$\overline{\chi} = \lim_{u \uparrow 1} \frac{2 \log(1 - u)}{\log \overline{C}_F(u, u)} - 1.$$

The connection with the tail order is given by Coles et al. (1999) and Heffernan (2000). Recalling that

$$\overline{C}_F(u, u) \sim (1 - u)^\kappa \mathscr{L} \left(1 - u \right),$$

as $u \to 1$, it follows that

$$\overline{\chi}(u) \sim \frac{2 \log(1 - u)}{\kappa \log(1 - u) + \log \mathscr{L} \left(1 - u \right)} - 1 \to \frac{2}{\kappa} - 1.$$

Because of the one-on-one relationship between the tail order and the asymptotic independence measure, results for the asymptotic independence measure follow trivially from the tail order results.

4.4 Summary

This chapter was concerned with measures to summarize the multivariate tail dependence structure. In the bivariate case, the tail dependence coefficient (TDC) λ, is commonly used to capture the strength of tail dependence between two random variables. Two extensions of the bivariate tail dependence coefficient to higher dimensions $d > 2$ have been considered: one based on the tail copula and another based on the stable tail dependence function. The tail copula-based multivariate TDC, τ, captures the conditional probability that all variables are extreme, given that one of them is. The measure is equal to the tail copula b evaluated in the point $(1, \ldots, 1)$. In contrast, the STDF-based multivariate TDC, Λ, captures the conditional probability that one or more variables are extreme given that one of them is. This measure is equal to a rescaled version of the stable tail dependence function ℓ evaluated in the point $(1, \ldots, 1)$.

The value of the tail copula-based TDC τ reduces significantly when one marginal pair of variables exhibits weak tail dependence, regardless of the tail dependence strength between the other variables. In particular, the measure τ is equal to zero when at least one marginal pair of variables is tail independent. Hence, in case of varying marginal tail dependencies, the tail copula-based multivariate TDC τ loses a significant amount of information. To mitigate this loss of information, a subset-based measure, τ_S, can be employed. However, for high dimensions, this approach requires a large number of subset-based TDCs to be evaluated in order to get a good picture of the tail dependence strength. In contrast, the STDF-based multivariate TDC Λ conveys most important information of the multivariate tail dependence structure. The ability of the multivariate TDC Λ to identify tail independence in all dimensions $d \geq 2$ (Main Theorem 2) is noteworthy and will be used in the next chapter to develop a test statistic to test the null hypothesis of tail independence.

In case of asymptotic independence, the tail order κ provides further information on the residual dependence that might be present in samples above high but finite thresholds. An overview of the considered tail dependence summary measures is given in Table 4.4.1 below. It can be seen that the considered tail dependence measures behave quite well and satisfy almost all desirable properties. Note that this is not naturally true for tail dependence measures. For example, another tail dependence that is sometimes considered in literature is the extreme correlation, which is defined as the correlation above high threshold values. The extreme correlation, similar to the general correlation measure, only exists if the variables to which it is applied have finite first and second moments. Moreover, it is not copula-based and is neither able to identify tail independence nor tail comonotonicity. We conclude that the multivariate TDC Λ is the best suited tail dependence measure to employ. In case of asymptotic independence, the tail order can provide additional information.

Table 4.4.1: Overview of tail dependence measures and their properties.

	TDC ($d = 2$)	TDC ($d > 2$) (STDF-based)	TDC ($d > 2$) (Tail copula-based)	Tail order
	λ	Λ	τ	κ
(P1) Existence*	✓	✓	✓	✓
(P2) Exchangeability	✓	✓	✓	✓
(P3) Normalisation	✓	✓	✓	✓
(P4) Copula-based	✓	✓	✓	✓
(P5) ID independence	✓	✓	✗	Only for $d = 2$
(P6) ID comonotonicity	✓	**	**	✗

* All tail dependence measures exist on condition that the limit in their definition exists. For all considered parametric copulas this is true.

** We conjecture that the STDF-based TDC Λ cannot identify asymptotic comonotonicity, whereas the tail copula-based TDC can identify asymptotic comonotonicity. It is, however, not straightforward to prove this and since asymptotic comonotonicity is not very relevant because it hardly ever occurs this problem is left unaddressed.

Chapter 5

Inference for Multivariate Extremal Dependence

After identifying appropriate characterizations for the multivariate tail dependence structure that either fully capture the dependence structure or summarize the strength of the dependence structure, this chapter is concerned with statistical procedures required for inference on these characterizations. The field of science concerned with statistical inference on multivariate extremal behavior is still in rapid development, and a large body of research on this topic already exists, although mostly for the bivariate case. There are many different approaches to tackle this problem: inference methods can be non-parametric or parametric, and, in the latter case, they can be likelihood-based, frequentist as well as Bayesian, or based on other techniques such as the method of moments or minimum distance estimation (Segers, 2012b). Moreover, inferences can be made based on any of the characterizations discussed in Chapter 3, presenting a wide range of possibilities to estimate the dependence structure between multivariate extremes.

Historically, the focus has been on estimating the full bivariate extreme value distribution through the exponent or spectral measure. See for example De Haan and Resnick (1993) for an estimator of the exponent measure and Einmahl et al. (1997), Einmahl et al. (2001), and Einmahl and Segers (2009), for a sequence of improving estimators of the spectral measure. Boldi and Davison (2007) and Guillotte et al. (2011) introduced a Bayesian approach to estimating the spectral measure, and Kiriliouk et al. (2014) presented an overview of nonparametric estimation procedures for the spectral measure. The bivariate Pickands dependence function also received quite some attention by researchers. See for example Pickands (1981) for a nonparametric approach and Tawn (1988) for a parametric approach. Abdous and Ghoudi (2005) and Vettori et al. (2017) both presented a comprehensive overview of nonparametric estimators of the Pickands dependence function. Unfortunately, many bivariate methods cannot be readily extended to the multivariate version of the Pickands dependence function. Nonparametric estimation of the tail copula was considered by Schmidt and Stadtmüller (2006) and Bücher and Dette (2013). Moreover, the relationships between the characterizations of the dependence structure imply that estimators for one characterization can also be employed to estimate another characterization. For example, an estimator for the Pickands dependence function can be defined based on an estimator for the spectral measure (Capéraà and Fougères, 2000).

In this chapter, however, the primary focus is on estimating the STDF and, inherently, the STDF-based multivariate TDC Λ. After discussing and evaluating several estimators, we elaborate on how the STDF estimators can be used to fashion test statistics and test the hypothesis of tail independence and, consequently, to classify variables as either tail dependent or tail independent. The testing of multivariate tail (in)dependence is a statistical problem that has received little attention in literature. We focus on these inference problems because they provide a comprehensive set of tools that can be used to (1) fully assess the multivariate tail dependence structure, (2) summarize the strength of multivariate extremal dependencies, and (3) classify data as either tail dependent or tail independent.

5.1 STDF estimators

This section introduces several estimators for the stable tail dependence function. Firstly, the empirical STDF is introduced along with several adjusted versions that smooth this nonparametric estimator or correct its bias. Thereafter, we briefly touch upon parametric methods to estimate the STDF and finally evaluate several estimators in a simulation study. Throughout this section, it is assumed that we have n independently and identically distributed (iid) data observations from a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with joint distribution F, marginal distribution functions $F_1, \ldots, F_d$, and a copula C_F that belongs to the maximum domain of attraction of an extreme value copula C and is therefore characterized by a stable tail dependence function ℓ. Data are denoted by $(X_{11}, \ldots, X_{1d}), \ldots, (X_{n1}, \ldots, X_{nd})$. The aim is to develop an estimator for the ℓ-function for $\mathbf{x} \in [0,1]^d$. Due to the homogeneity of the STDF, this estimator can be employed to retrieve estimates for all $\mathbf{x} \in \mathbb{R}_+^d$. It would also be sufficient to restrict the domain of the estimator to the unit simplex Δ_{d-1}, where we would actually be considering the problem of estimating the Pickands dependence function.

5.1.1 Empirical STDF

The empirical stable tail dependence function approximates the STDF nonparametrically in a straightforward manner. It was first introduced for the bivariate case by Huang (1992) and later by Drees and Huang (1998). Consider again the definition of the d-dimensional STDF,

$$\begin{aligned}\ell(x_1, \ldots, x_d) &= \lim_{t \downarrow 0} \frac{1 - C_F(1 - tx_1, \ldots, 1 - tx_d)}{t} \\ &= \lim_{t \downarrow 0} \frac{1}{t} \mathbb{P}\left(F_1(X_1) > 1 - tx_1 \text{ or } \ldots \text{ or } F_d(X_d) > 1 - tx_d\right).\end{aligned}$$

Looking closely at the definition of the STDF, it can be noticed there are three components of the ℓ-function that have to be approximated to estimate the function nonparametrically. Firstly, the limiting value $t \downarrow 0$ has to be replaced by a finite number. It is customary in extreme value theory to introduce an intermediate sequence $k := k_n \in \{1, \ldots, n\}$ such that $k \to \infty$ and $k/n \to 0$ as $n \to 0$. The limiting value $t \downarrow 0$ is approximated by the finite value k/n (step 1). Note that n is implied by the dataset, whereas k has to be chosen by the statistician. Secondly, the multivariate probability $\mathbb{P}$ has to be approximated. The most straightforward way to do this is using the empirical copula (step 2). Note that this is, in fact, the same problem as having to estimate copula as discussed in Section 2.5. Thirdly, the marginal distributions of the variables, $F_1, \ldots, F_d$, have to be estimated (step 3). This can be done employing the empirical distribution function, which is based on the ranks of data. By plugging these approximations into the definition of the STDF step by step, the empirical STDF is retrieved as follows,

$$\begin{aligned}\ell(x_1, \ldots, x_d) &= \lim_{t \downarrow 0} \frac{1}{t} \mathbb{P}\left(F_1(X_1) > 1 - tx_1 \text{ or } \ldots \text{ or } F_d(X_d) > 1 - tx_d\right) \\ &\approx \frac{n}{k} \mathbb{P}\left(F_1(X_1) > 1 - \frac{kx_1}{n} \text{ or } \ldots \text{ or } F_d(X_d) > 1 - \frac{kx_d}{n}\right) && \text{(Step 1)} \\ &\approx \frac{n}{k} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{F_1(X_{i1}) > 1 - \frac{kx_1}{n} \text{ or } \ldots \text{ or } F_d(X_{id}) > 1 - \frac{kx_d}{n}\right\} && \text{(Step 2)} \\ &\approx \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{\hat{F}_{n1}(X_{i1}) > 1 - \frac{kx_1}{n} \text{ or } \ldots \text{ or } \hat{F}_{nd}(X_{id}) > 1 - \frac{kx_d}{n}\right\} && \text{(Step 3)} \\ &= \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{X_{i1} > \hat{F}_{n1}^{-1}\left(1 - \frac{kx_1}{n}\right) \text{ or } \ldots \text{ or } X_{id} > \hat{F}_{nd}^{-1}\left(1 - \frac{kx_d}{n}\right)\right\} \\ &= \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{X_{i1} > X_{n-kx_1,n} \text{ or } \ldots \text{ or } X_{id} > X_{n-kx_d,n}\right\} \\ &= \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{R_{i1,n} > n - kx_1 \text{ or } \ldots \text{ or } R_{id,n} > n - kx_d\right\},\end{aligned}$$

where $R_{ij,n}$, $i \in \{1,\dots,n\}, j \in \{1,\dots,d\}$, denotes the rank of X_{ij} out of $1,\dots,n$. Intuitively, the number of vector observations where one or more of its components exceeds a high threshold is counted and properly rescaled. The high thresholds depend on the marginal distributions $F_1,\dots,F_d$. If the marginal distributions are estimated with the empirical distribution function, an observation exceeds a high quantile exactly when the rank of the observation is higher than the rank of the threshold value. Einmahl et al. (2012) introduced an alternative estimator with slightly better finite-sample properties:

$$\hat{\ell}_{n,k}(\mathbf{x}) := \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left\{R_{i1,n} > n + 1/2 - kx_1 \text{ or } \dots \text{ or } R_{id,n} > n + 1/2 - kx_d\right\}.$$

The variations of the estimator retrieved by specifying $n - kx_j$, $n + 1/2 - kx_j$, or $n + 1 - kx_j$, are all asymptotically equivalent but the latter two often yield better finite sample performances. Hence, we use the definition of Einmahl et al. (2012).

Definition 5.1. *Let $(X_{11},\dots,X_{1d}),\dots,(X_{n1},\dots,X_{nd})$ be d-dimensional vector observations from a copula C_F. Let $k := k_n \in \{1,\dots,n\}$ be such that $k \to \infty$, and $k/n \to 0$ if $n \to \infty$. The empirical stable tail dependence function is given by*

$$\hat{\ell}_{n,k}(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left\{R_{i1,n} > n + 1/2 - kx_1 \text{ or } \dots \text{ or } R_{id,n} > n + 1/2 - kx_d\right\}, \tag{5.1.1}$$

where $R_{ij,n}$, $i \in \{1,\dots,n\}, j \in \{1,\dots,d\}$, denotes the rank of X_{ij} out of $1,\dots,n$.

To get a feeling of how the empirical STDF behaves in practice, data are simulated from the bivariate Gumbel copula. The estimated $\hat{\ell}_{n,k}$-functions are shown below for different values of the sample size (Figure 5.1.1). The multivariate empirical STDF can be estimated in `R` with the `stdfEmp` function from the `tailDepFun` package. Alternatively, it is straightforward to implement the estimator (see Appendix 5.1). The figures show that the empirical STDF approximates the real ℓ-function quite well, even for a relatively small sample size. The performance of the estimator is further assessed in the simulation study presented in Section 5.1.5. A drawback of the empirical STDF is that it usually does not satisfy the shape constraints of the stable tail dependence function (see Theorem 3.8). Most notably, the empirical STDF is not continuous and not convex.

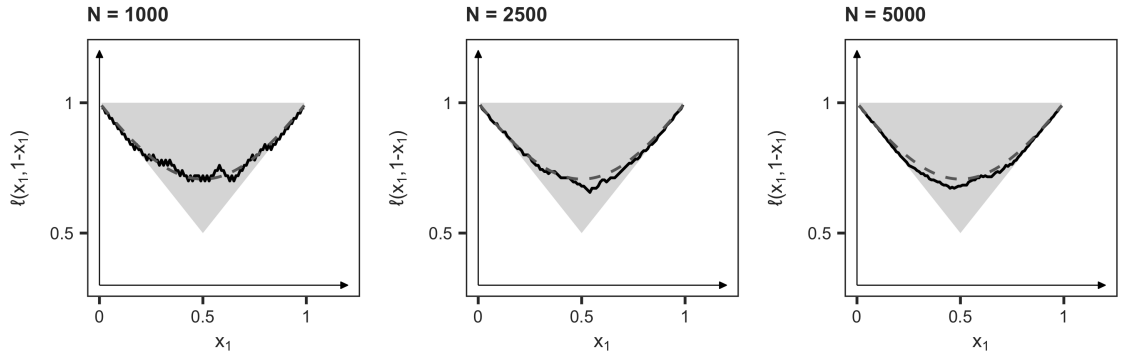

Figure 5.1.1: The empirical STDF for data simulated from the bivariate Gumbel copula with model parameter $\theta = 2$ using $k = 5\%$ of $N \in \{1000, 2500, 5000\}$ observations.

Asymptotic behavior

It is essential to know the asymptotic behavior of the empirical STDF to construct a test statistic that can be used for testing tail independence (Section 5.2). Furthermore, the asymptotic behavior of the empirical STDF is a building stone for the derivation of the asymptotic behavior of most other nonparametric and semi-parametric STDF estimators. Hence, in the following we address in detail the asymptotic properties of the empirical STDF estimator. It is known that, under certain regularity conditions, the empirical STDF converges to a zero-mean Wiener process with a variance that is dependent on the true and usually unknown stable tail dependence function ℓ. Recall that the Wiener process is a Gaussian stochastic process that is defined as follows.

Definition 5.2. *A Wiener process (or Brownian motion) W_t is a stochastic process characterized by the following properties: (1) $W_0 = 0$ a.s.; (2) W has independent increments; (3) W has Gaussian increments: $W_{t+s} - W_t \sim N(0, t-s)$; and (4) W has continuous paths.*

In order to state and prove the asymptotic result, some notations and definitions are introduced according to the exposition of Einmahl et al. (2012). Firstly, let W_ν denote a zero-mean Gaussian process indexed by a measure ν that works on Borel sets of $[0,\infty]^d \backslash \{(\infty,\ldots,\infty)\}$ and is related to the STDF as

$$\nu\left(\left\{\mathbf{u} \in [0,\infty]^d : u_1 \leq x_1 \text{ or } \ldots \text{ or } u_d \leq x_d\right\}\right) = \ell(x_1,\ldots,x_d).$$

The covariance structure of W_ν is given by

$$\mathbb{E}\left(W_\nu(A_1)W_\nu(A_2)\right) = \nu(A_1 \cap A_2) \tag{5.1.2}$$

for any two Borel sets A_1 and A_2 in $[0,\infty]^d \backslash \{(\infty,\ldots,\infty)\}$. Define

$$W_\ell(\mathbf{x}) = W_\nu\left(\{u \in [0,\infty]^d \backslash \{(\infty,\ldots,\infty)\} : u_1 \leq x_2 \text{ or } \ldots \text{ or } u_d \leq x_d\}\right), \tag{5.1.3}$$

and let $W_{\ell,j}$, $j = 1,\ldots,d$ be the marginal processes,

$$W_{\ell,j}(x_j) = W_\ell(0,\ldots,0,x_j,0,\ldots,0), \qquad x_j \geq 0. \tag{5.1.4}$$

Define ℓ_j to be the right-hand partial derivative of ℓ with respect to x_j, $j = 1,\ldots,d$. Lastly, define the zero-mean Gaussian process

$$B_\ell(\mathbf{x}) = W_\ell(\mathbf{x}) - \sum_{i=1}^{d} \ell_j(\mathbf{x}) W_{\ell,j}(x_j) \tag{5.1.5}$$

with variance defined by

$$\begin{aligned}\mathbb{E}\left(B_\ell(\mathbf{x})^2\right) &= \mathbb{E}\left(W_\ell(\mathbf{x})^2\right) - 2\sum_{j=1}^{d} \ell_j(\mathbf{x})\mathbb{E}\left(W_\ell(\mathbf{x})W_{\ell,j}(x_j)\right) + \mathbb{E}\left(\left(\sum_{j=1}^{d} \ell_j(\mathbf{x})W_{\ell,j}(x_j)\right)^2\right) \\ &= \ell(\mathbf{x}) - 2\sum_{j=1}^{d} \ell_j(\mathbf{x})\ell(0,\ldots,0,x_j,0,\ldots,0) + \sum_{j=1}^{d} \ell_j^2(\mathbf{x})\ell(0,\ldots,0,x_j,0,\ldots,0).\end{aligned}$$

Now we are ready to state the result.

Theorem 5.1. *(Einmahl et al., 2012). Assuming that*

1. *$\lim_{t\downarrow 0} t^{-1}\left(1 - C_F(1-tx_1,\ldots,1-tx_d)\right)$ exists and converges uniformly to $\ell(x_1,\ldots,x_d)$ on $[0,T]^d$ for $T > 0$ (first order condition);*
2. *$t^{-1}\left(1 - C_F(1-tx_1,\ldots,1-tx_d)\right) - \ell(\mathbf{x}) = \mathcal{O}\left(t^\alpha\right)$, uniformly in $\mathbf{x} \in [0,1]^d$ as $t \downarrow 0$, for some $\alpha > 0$ (second order condition);*
3. *$k = \mathcal{O}\left(n^{2\alpha/(1+2\alpha)}\right)$ for the positive number α used in the assumption above, and $k \to \infty$ as $n \to \infty$;*
4. *for all $j = 1,...,d$, the first-order partial derivative of ℓ with respect to x_j exists and is continuous on the set of points $\mathbf{x}$ such that $x_j > 0$;*

we have

$$\sup_{x\in[0,T]^d} \left|\sqrt{k}\left(\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right) - B_\ell(\mathbf{x})\right| \to 0,$$

for $T > 0$ as $n \to \infty$. Here $B_\ell(\mathbf{x})$ is a zero-mean Gaussian process defined in Equation 5.1.5.

Proof. An outline of the proof is given in Appendix B.4.1. ■

5.1.2 Smoothed versions of the empirical STDF

As mentioned above, a drawback of the empirical STDF is that it usually does not satisfy the shape constraints of the stable tail dependence function. Several corrective steps can be taken to solve this problem. In this section, two smoothed versions of the empirical STDF are considered that mitigate the discontinuities of the standard empirical STDF. The smoothed versions are more intuitive to interpret visually since they are more similar to true dependence functions. Moreover, the smoothing decreases the variance of the estimates, thereby potentially improving the finite sample performance compared to the standard empirical STDF (on the condition that the bias does not increase too much). Although both smoothed versions of the empirical STDF yield a continuous estimator, they are generally not convex. Other estimators that do meet the shape constraints of the STDF have been considered in literature, but mainly for the bivariate case. For example, Pickands (1981) employef the greatest convex minorant for the bivariate Pickands dependence function, Hall and Tajvidi (2000) suggested the use of constrained smoothed splines, and Fils-villetard et al. (2008) and Gudendorf and Segers (2012) developed methods to project initial estimates onto a set of proper dependence functions. Additionally, Marcon et al. (2017) explored the projection of initial multivariate estimates on a set of Bernstein polynomials. However, this method does not necessarily yield proper dependence functions either, and extensive computations are required, especially for higher dimensions.

Beta copula smoothing

Kiriliouk et al. (2018) introduce a smoothed version of the empirical stable tail dependence function based on the beta copula (see Section 2.5). By using this smoothed copula, the estimated STDF becomes continuous and always satisfies the theoretical bounds of the STDF. The beta-smoothed empirical STDF is defined as follows.

Definition 5.3. *Let $(X_{11}, \ldots, X_{1d}), \ldots, (X_{n1}, \ldots, X_{nd})$ be d-dimensional vector observations from a copula C_F. Let $k : k_n \in \{1, \ldots, n\}$ be such that $k \to \infty$, and $k/n \to 0$ if $n \to \infty$. The beta-smoothed empirical stable tail dependence function is given by*

$$\hat{\ell}^{\beta}_{n,k}(\mathbf{x}) = \frac{n}{k}\left(1 - \hat{C}^{\beta}_{n}\left(1 - \frac{k}{n}\mathbf{x}\right)\right),$$

where $\hat{C}^{\beta}_{n}$ is the empirical beta copula.

See Section 2.5 for more information on the beta copula. Kiriliouk et al. (2018) show that the asymptotic distribution of the beta-smoothed empirical STDF estimator is the same as the asymptotic distribution of the standard empirical STDF. Although these two estimators are asymptotically equivalent, the beta-smoothed estimator might yield better finite sample performance.

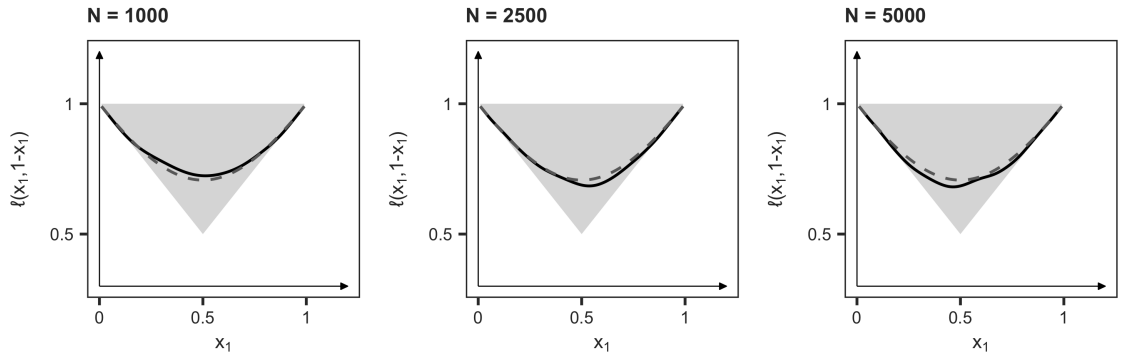

Figure 5.1.2: The beta-smoothed empirical STDF for data simulated from the bivariate Gumbel copula with model parameter $\theta = 2$ using $k = 5\%$ of $N \in \{1000, 2500, 5000\}$ observations.

To get a feeling of how the beta-smoothed estimator behaves, data are simulated from the bivariate Gumbel copula. The estimated $\hat{\ell}^{\beta}_{n,k}$-functions are shown for different values of the sample size (Figure 5.1.2). To our knowledge, there currently is no standard implementation of this estimator available in `R`. However, it is straightforward to estimate the beta-smoothed empirical

STDF by using the `C.n` function from the `copula` package that provides the option to estimate a beta-smoothed copula. The STDF estimator can be easily derived from this copula (see Appendix E.1 for our implementation). The figures show that the beta-smoothed empirical STDF approximates the real ℓ-function quite well for all considered sample sizes. Moreover, due to the continuity of the estimated functions, they resemble the real STDF more closely than the standard empirical STDF and can, therefore, be interpreted more readily.

Kernel smoothing

Another approach to retrieve a smoothed version of the STDF estimator is to smooth the empirical STDF instead of the empirical copula. Since the STDF is not bounded on $[0,1]^d$, a kernel smoothing approach can be taken without encountering boundary issues. Recall from Section 2.5 that a kernel function is a positive integrable function that is often used in nonparametric statistics to smooth estimators. We define the following kernel-smoothed STDF estimator.

Definition 5.4. *Let $(X_{11},\dots,X_{1d}),\dots,(X_{n1},\dots,X_{nd})$ be d-dimensional vector observations from a copula C_F. Let $k : k_n \in \{1,\dots,n\}$ be such that $k \to \infty$, and $k/n \to 0$ if $n \to \infty$. The kernel-smoothed empirical stable tail dependence function is given by*

$$\hat{\ell}^K_{n,k}(\mathbf{x}) = \frac{\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-1} \hat{\ell}_{n,k}(a_j \mathbf{x})}{\frac{1}{k}\sum_{j=1}^{k} K(a_j)} \tag{5.1.6}$$

where $\hat{\ell}_{n,k}$ is the standard empirical STDF, $a_j = \frac{j}{k+1}$ and K is a kernel function on $(0,1)$ such that $\int_0^1 K(u)\mathrm{d}u = 1$.

This specific kernel-smoothed empirical STDF is inspired by the bias-corrected estimator introduced by Beirlant et al. (2016), which will be discussed in the next section. In order to get a feeling of the behavior of the kernel-smoothed estimator, the simulated data from the bivariate Gumbel copula are used again to estimate the $\hat{\ell}^K_{n,k}$-functions. The results are shown in Figure 5.1.3. The implementation of this kernel-smoothed estimator is our own (see Appendix E.1). The figures show that the kernel-smoothed estimator yields very similar results to the beta-smoothed estimator. The kernel-smoothed functions are slightly less smooth, but this can be adjusted by specifying different parameters for the power kernel function.

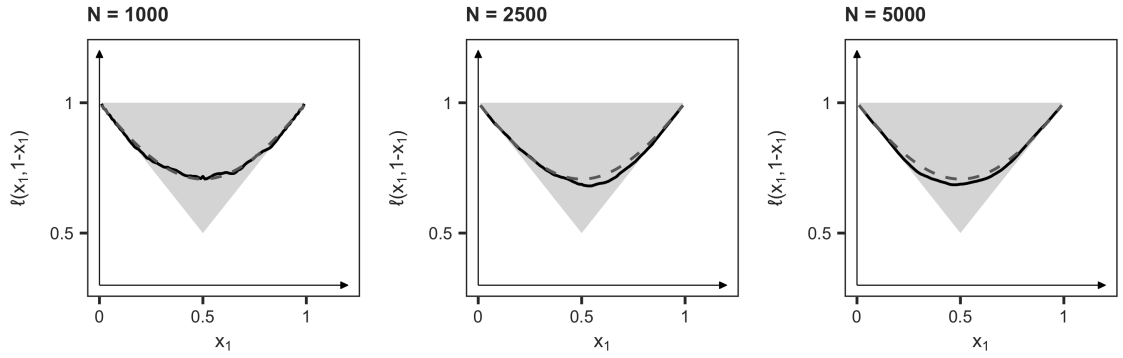

Figure 5.1.3: The kernel-smoothed empirical STDF for data simulated from the bivariate Gumbel copula with model parameter $\theta = 2$ using $k = 5\%$ of $N \in \{1000, 2500, 5000\}$ observations. The power kernel with model parameter $\tau = 5$ is used.

The asymptotic behavior of the kernel-smoothed estimator follows easily from Theorem 5.1 and is stated in the following proposition.

Proposition 5.1. *Under the conditions of Theorem 5.1, we have*

$$\sqrt{k}\left\{\hat{\ell}^K_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right\} \to \left(\int_0^1 K(u) u^{-1/2}\mathrm{d}u\right) B_\ell(\mathbf{x}), \tag{5.1.7}$$

as $n \to \infty$. Here $B_\ell(\mathbf{x})$ is a zero-mean Gaussian process defined in Equation 5.1.5, and K is a kernel function on $(0,1)$ such that $\int_0^1 K(u)\mathrm{d}u = 1$.

Proof. We start from the asymptotic behavior of the empirical STDF presented in Theorem 5.1, i.e.,

$$\sqrt{k}\left\{\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right\} \to B_\ell(\mathbf{x}).$$

By the continuous mapping theorem and by the observation that $\sum_{j=1}^{k} K(a_j) \to \int_0^1 K(u)\mathrm{d}u = 1$ as $n \to \infty$,

$$\sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^{k} K(a_j)\hat{\ell}_{n,k}(\mathbf{x}) - \sum_{j=1}^{k} K(a_j)\ell(\mathbf{x})\right\} \to \left(\int_0^1 K(u)\mathrm{d}u\right) B_\ell(\mathbf{x}) = B_\ell(\mathbf{x}).$$

By the homogeneity of ℓ and the fact that $B_\ell(a\mathbf{x}) \sim \sqrt{a}B_\ell(\mathbf{x})$ it next follows that

$$\sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^{k} K(a_j)a_j^{-1}\hat{\ell}_{n,k}(a_j^{-1}\mathbf{x}) - \sum_{j=1}^{k} K(a_j)\ell(\mathbf{x})\right\} \to \left(\int_0^1 K(u)u^{-1/2}\mathrm{d}u\right) B_\ell(\mathbf{x}).$$

The result follows by dividing everything by $\sum_{j=1}^{k} K(a_j)$. ■

In a similar setting, Beirlant et al. (2016) propose to use the power kernel $K(t) = (\tau+1)t^\tau \mathbb{1}\{t \in (0,1)\}$, $\tau > -1/2$. By Theorem 5.1, the asymptotic behavior of the power kernel-smoothed empirical STDF is given by

$$\sqrt{k}\left\{\hat{\ell}^K_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right\} \to 2\frac{1+\tau}{1+2\tau}B_\ell(\mathbf{x}), \tag{5.1.8}$$

as $n \to \infty$. For small values of τ the variance of the asymptotic distribution of $\hat{\ell}^K_{n,k}$ is inflated relative to the asymptotic distribution of the standard empirical STDF but for large values of τ this inflation factor stemming from the power kernel converges to one. Although asymptotically the estimators are almost equivalent, the finite sample performance might be different.

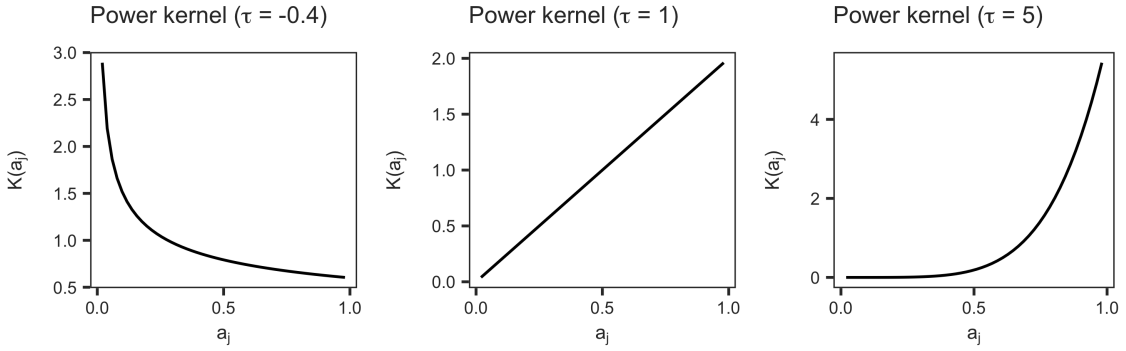

Figure 5.1.4: The power kernel for different model parameters.

5.1.3 Bias corrections

Besides the fact that the empirical STDF does not satisfy the shape constraints of the theoretical STDF, another possible drawback of the empirical STDF is that the bias increases quickly as the number of tail observations k increases. This could ruin the finite sample performance of the empirical STDF, especially if the trade-off between bias and variance is not optimally handled. To illustrate the problem, data are simulated from a bivariate Normal copula with model parameter $\rho = 0.5$. Recall that the Normal copula belongs to the maximum domain of attraction of Π. However, for the relatively small sample size of $N = 1000$, the empirical STDF indicates that the data are strongly tail dependent even though they are in fact tail independent. Also for the larger sample sizes, a bias remains, albeit somewhat smaller.

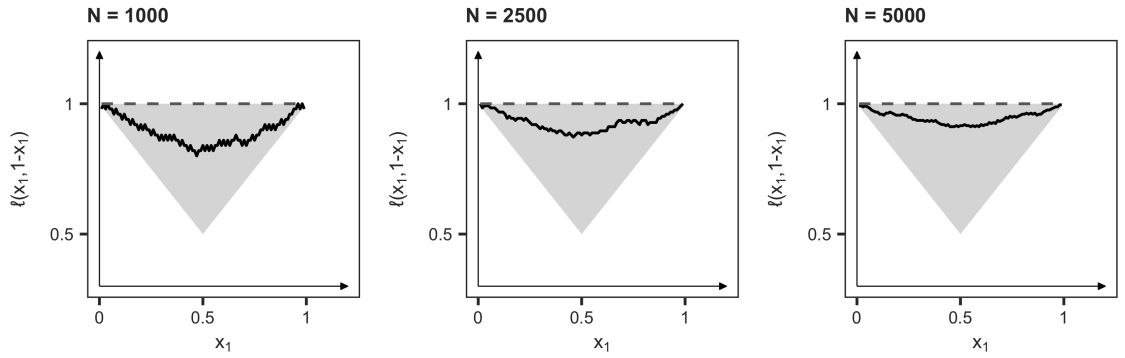

Figure 5.1.5: Estimated empirical ℓ-functions based on simulation results for the Normal copula in dimension $d = 2$ with model parameter $\rho = 0.5$ using $k = 5\%$ of the $N \in \{1000, 2500, 5000\}$ observations.

To mitigate this problem, bias-correction methods for the empirical STDF have been introduced in the literature by Fougères et al. (2015) and Beirlant et al. (2016). The bias-correction methodology is derived in the context of a slightly different asymptotic setup than before. Whereas it was first assumed that $k = \mathcal{O}\left(n^{2\alpha/(1+2\alpha)}\right)$ (assumption 3 in Theorem 5.1) we now allow k to go to ∞ at a slower rate and impose a third order condition in order to introduce a bias-term to the estimator than can be removed from an initial nonparametric estimator. Formally, this is specified in the following Theorem 5.2.

Theorem 5.2. *(Fougères et al., 2015). Assuming that*

1. $\lim_{t\downarrow 0} t^{-1}\left(1 - C_F(1-tx_1, \ldots, 1-tx_d)\right)$ *exists and converges uniformly to* $\ell(x_1, \ldots, x_d)$ *on* $[0,T]^d$ *for* $T > 0$ *(first order condition);*

2. *there exists a positive function* α*, such that* $\alpha(t) \to 0$ *as* $t \to \infty$*, and a nonzero function* M *such that for all* $\mathbf{x} \in \mathbb{R}_+^d$*,* $\lim_{t\downarrow 0} \frac{1}{\alpha(t)}\left(t^{-1}\left(1 - C_F(1-tx_1, \ldots, 1-tx_d)\right) - \ell(\mathbf{x})\right) = M(\mathbf{x})$*, uniformly on* $[0,T]^d$ *for* $T > 0$ *(second order condition);*

3. *there exists a positive function* β*, such that* $\beta(t) \to 0$ *as* $t \to \infty$*, and a nonzero function* N *such that for all* $\mathbf{x} \in \mathbb{R}_+^d$*,* $\lim_{t\downarrow 0} \frac{1}{\beta(t)}\left(\frac{1}{\alpha(t)}\left(t^{-1}\left(1 - C_F(1-tx_1, \ldots, 1-tx_d)\right) - \ell(\mathbf{x})\right) - M(\mathbf{x})\right)$ $= N(\mathbf{x})$*, uniformly on* $[0,T]^d$ *for* $T > 0$ *(third order condition);*

4. *the functions* M *and* N *are continuous and homogeneous of order* $1-\rho$ *and of order* $1-\rho-\rho'$ *respectively, with* $\rho, \rho' < 0$*, and the function* M *is differentiable;*

5. k *is such that* $\sqrt{k}\alpha(n/k) \to \infty$ *and* $\sqrt{k}\alpha(n/k)\beta(n/k) \to 0$ *as* $n \to \infty$*;*

6. *for all* $j = 1, ..., d$*, the first-order partial derivative of* ℓ *with respect to* x_j *exists and is continuous on the set of points* $\mathbf{x}$ *such that* $x_j > 0$*;*

we have

$$\sqrt{k}\left\{\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x}) - \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\right\} \to B_\ell(\mathbf{x}) \tag{5.1.9}$$

in $D([0,T]^d)$ *for every* $T > 0$ *as* $n \to \infty$*. Here* $B_\ell(\mathbf{x})$ *is a zero-mean Gaussian process dependent on the true tail dependence function defined in Equation 5.1.5.*

Proof. The proof follows the same line of logic as the proof of Theorem 5.1. However, because the intermediate sequence now satisfies $\sqrt{k}\alpha(n/k) \to \infty$, the D_2-term does not vanish but introduces a positive bias term equal to $\alpha(n/k)M(\mathbf{x})$. That is,

$$\sup_{\mathbf{x}\in[0,T]^d} \left|\sqrt{k}\left(D_2(\mathbf{x}) - \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\right)\right| \to 0 \quad \text{a.s.}$$

See Fougères et al. (2015) for the detailed proof. ■

Note that if $\sqrt{k}\alpha(n/k)$ tends to zero, which is implied by the assumptions of Theorem 5.1, the estimator is consistent and asymptotically Normal with convergence rate $\sqrt{k}$, conform Einmahl et al. (2012). However, an asymptotic bias exists if $\sqrt{k}\alpha(n/k)$ converges to a nonzero constant, as is implied by Theorem 5.2. The main idea is to estimate the bias αM that occurs and to remove it from the initial STDF estimator. In order to do so, consider a rescaled version, $\hat{\ell}_{n,k,a}(\mathbf{x}) = a^{-1}\hat{\ell}_{n,k}(a\mathbf{x})$ for $a > 0$, and define

$$\hat{\Delta}_{k,a}(\mathbf{x}) = \hat{\ell}_{n,k,a}(a\mathbf{x}) - \hat{\ell}_{n,k}(\mathbf{x}). \tag{5.1.10}$$

Recall that by the homogeneity of the true STDF, we have that $\ell(a\mathbf{x}) = a\ell(\mathbf{x})$ for $a > 0$ and that by the homogeneity of the M-function (assumption 4 in Theorem 5.2) we have that $M(a\mathbf{x}) = a^{1-\rho}M(\mathbf{x})$ for $a > 0$ and $\rho < 0$. Also note that for the Gaussian process B_ℓ (defined in Equation 5.1.5) it holds that $B_\ell(a\mathbf{x}) = \sqrt{a}B_\ell(\mathbf{x})$ for $a > 0$. Therefore, under the conditions of Theorem 5.2, we find that

$$\sqrt{k}\left\{\hat{\ell}_{n,k,a}(\mathbf{x}) - \ell(\mathbf{x}) - \alpha\left(\frac{n}{k}\right)a^{-\rho}M(\mathbf{x})\right\} \to a^{-1/2}B_\ell(\mathbf{x}).$$

Based on this expression, Fougères et al. (2015) propose the following bias-corrected estimator for the ℓ-function:

$$\hat{\ell}^{BC1}_{n,k,\overline{k}}(\mathbf{x}) = \hat{\ell}_{n,k,a}(\mathbf{x}) - \hat{\Delta}_{k,\left(a^{-\hat{\rho}_{\overline{k}}(\mathbf{x}^*)}+1\right)^{-1/\hat{\rho}_{\overline{k}}(\mathbf{x}^*)}}(\mathbf{x}), \tag{5.1.11}$$

with

$$\hat{\rho}_{\overline{k},a,r}(\mathbf{x}^*) = \min\left\{1 - \frac{1}{\log r}\log\left|\frac{\hat{\Delta}_{\overline{k},a}(r\mathbf{x}^*)}{\hat{\Delta}_{\overline{k},a}(\mathbf{x}^*)}\right|, 0\right\} \tag{5.1.12}$$

the estimator of ρ based on another intermediate sequence $\overline{k} = k_n$ such that $\overline{k}/k \to 0$ for a fixed vector $\mathbf{x}^* \in \mathbb{R}^d$ and with $r \in (0,1)$. Alternatively, to avoid the estimation of ρ, Fougères et al. (2015) propose a second bias-corrected estimator as follows,

$$\hat{\ell}^{BC2}_{n,k,\overline{k}}(\mathbf{x}) = \frac{\hat{\ell}_{n,k}(\mathbf{x})\hat{\Delta}_{\overline{k},a}(a\mathbf{x}) - \hat{\ell}_{n,k}(a\mathbf{x})\hat{\Delta}_{\overline{k},a}(\mathbf{x})}{\hat{\Delta}_{\overline{k},a}(a\mathbf{x}) - a\hat{\Delta}_{\overline{k},a}(\mathbf{x})}. \tag{5.1.13}$$

To get an idea of the behavior of these bias-corrected estimators, the simulated data from the bivariate Normal copula are now used to estimate the bias-corrected empirical ℓ-functions. The results are shown in Figure 5.1.6 and Figure 5.1.7 below. Since there are no ready-to-use implementations of the estimators from Fougères et al. (2015) in `R` to our knowledge, we created our own implementation (see Appendix E.1). Compared to the uncorrected empirical STDF shown in Figure 5.1.5, the bias-corrected estimators are in fact closer to the true independence STDF and therefore effectively reduce bias. However, the figures also illustrate that the bias-corrected estimators of Fougères et al. (2015) are quite irregular and, as a consequence, will have a relatively large variance.

Beirlant et al. (2016) propose a third bias-corrected estimator based on a kernel-smoothed version of an initial nonparametric STDF estimator. The bias correction is done based on a kernel-smoothed version of the rescaled STDF,

$$\tilde{\ell}_{n,k}(\mathbf{x}) = \frac{1}{k}\sum_{j=1}^{k} K(a_j)\hat{\ell}_{n,k,a_j}(\mathbf{x}), \qquad a_j = \frac{j}{k+1}, \quad j \in \{1,\ldots,k\}, \tag{5.1.14}$$

where K is a kernel function on $(0,1)$ such that $\int_0^1 K(u)\mathrm{d}u = 1$. Defining the difference between this kernel-smoothed rescaled STDF and the original kernel-smoothed STDF as

$$\Delta_{k,a}(\mathbf{x}) = a^{-1}\tilde{\ell}_{n,k}(a\mathbf{x}) - \tilde{\ell}_{n,k}(\mathbf{x}) \tag{5.1.15}$$

the bias-corrected estimator of Beirlant et al. (2016) is defined as follows.

$$\hat{\ell}^{BC3}_{n,k,\overline{k}}(\mathbf{x}) = \frac{\tilde{\ell}_{n,k}(\mathbf{x}) - \left(\frac{\overline{k}}{k}\right)^{\hat{\rho}_{\overline{k}}(\mathbf{x}^*)}\tilde{\alpha}_{\overline{k}}(\mathbf{x})\frac{1}{k}\sum_{j=1}^{k}K(a_j)a_j^{-\hat{\rho}_{\overline{k}}(\mathbf{x}^*)}}{\frac{1}{k}\sum_{j=1}^{k}K(a_j)} \tag{5.1.16}$$

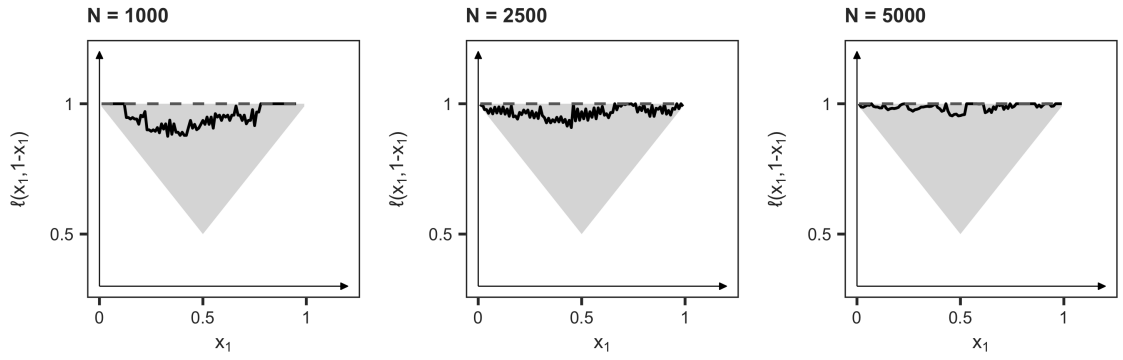

Figure 5.1.6: Estimated bias-corrected empirical $\hat{\ell}^{BC1}$-function of Fougères et al. (2015) based on simulation results for the bivariate Normal copula with model parameter $\rho = 0.5$ using $k = 5\%$ of the $N \in \{1000, 2500, 5000\}$ observations.

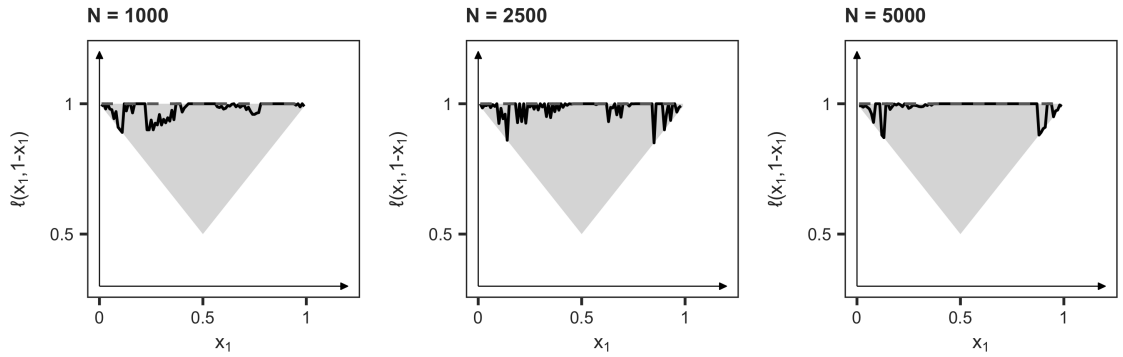

Figure 5.1.7: Estimated bias-corrected empirical $\hat{\ell}^{BC2}$-function of Fougères et al. (2015) based on simulation results for the bivariate Normal copula with model parameter $\rho = 0.5$ using $k = 5\%$ of the $N \in \{1000, 2500, 5000\}$ observations.

where $\hat{\rho}_{\overline{k}}$ is the estimator of ρ, defined in Equation 5.1.12, but now for the difference between the kernel-smoothed estimators and where

$$\tilde{\alpha}_k(\mathbf{x}) = \frac{\sum_{j=1}^{k}\sum_{l=1}^{k}\left(a_j^{-\hat{\rho}_k(\mathbf{x})} - a_l^{-\hat{\rho}_k(\mathbf{x})}\right)\hat{\ell}_{n,k,a_j}(\mathbf{x})}{\sum_{j=1}^{k}\sum_{l=1}^{k} a_j^{-\hat{\rho}_k(\mathbf{x})}\left(a_j^{-\hat{\rho}_k(\mathbf{x})} - a_l^{-\hat{\rho}_k(\mathbf{x})}\right)} \tag{5.1.17}$$

is the estimator of αM. Asymptotic normality holds with the limiting random vector equal in distribution to $\int_0^1 K(u)u^{-1/2}\mathrm{d}u$ times the limiting distribution of $\hat{\ell}_{n,k}$. This is formalized in the following proposition.

Proposition 5.2. *Under the conditions of Theorem 5.2, we have*

$$\sqrt{k}\left(\hat{\ell}^{BC3}_{n,k,\overline{k}}(\mathbf{x}) - \ell(\mathbf{x})\right) \to \left(\int_0^1 K(u)u^{-\frac{1}{2}}\mathrm{d}u\right) B_\ell(\mathbf{x}), \tag{5.1.18}$$

as $n \to \infty$, $k \to \infty$, $\overline{k} \to \infty$, and $\overline{k}/k \to 0$. Here $B_\ell(\mathbf{x})$ is a zero-mean Gaussian process defined in Equation 5.1.5, and K is a kernel on $(0,1)$ such that $\int_0^1 K(u)\mathrm{d}u = 1$.

Proof. See the proof in Beirlant et al. (2016). ■

The limiting process is independent of the value of ρ and has the same asymptotic variance as the uncorrected estimator. If the power kernel is taken, $K(t) = (\tau+1)t^\tau \mathbb{1}\{t \in (0,1)\}$, $\tau > -1/2$, (see Figure 5.1.4), the following expression can be found for the asymptotic distribution:

$$\sqrt{k}\left(\hat{\ell}^{BC3}_{n,k,\overline{k}}(\mathbf{x}) - \ell(\mathbf{x})\right) \to 2\frac{1+\tau}{1+2\tau}B_\ell(\mathbf{x}). \tag{5.1.19}$$

The bias-corrected STDF estimator from Beirlant et al. (2016) can be estimated in `R` using the `stdfEmpCorr` function from the `tailDepFun` package. However, since an adaption of this estimator is employed in the context of testing tail independence (Section 5.2), we also fashioned our own implementation of this estimator (see Appendix E.1 for the code). To get an idea of the behavior of this third bias-corrected estimator, the $\hat{\ell}^{BC3}$-function is estimated for data simulated from the bivariate Normal copula. The results are shown in Figure 5.1.8. For larger sample sizes, the estimator effectively reduces the bias associated with the standard empirical STDF (see Figure 5.1.5). Compared to the first two bias-corrected estimators introduced by Fougères et al. (2015) this estimator is more regular which will result in a lower variance of the estimator. The finite sample performance of the bias-corrected estimators will be further assessed in the simulation study presented in Section 5.1.5.

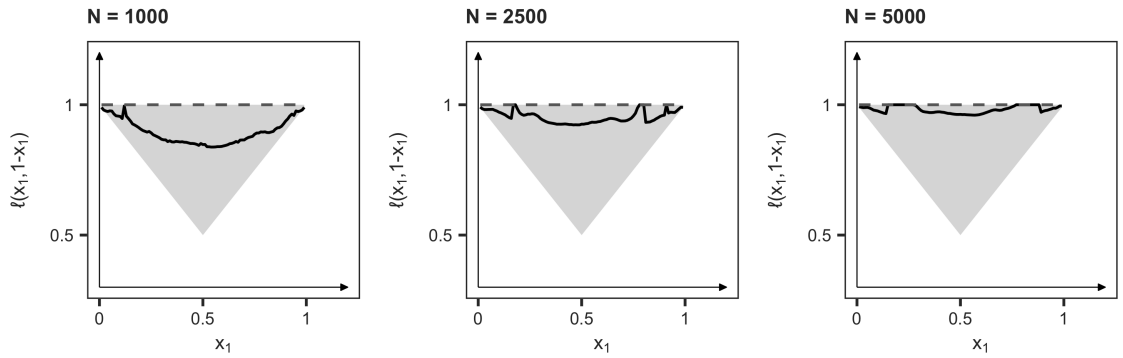

Figure 5.1.8: Estimated bias-corrected empirical $\hat{\ell}^{BC3}$-function of Beirlant et al. (2016) based on simulation results for the bivariate Normal copula with model parameter $\rho = 0.5$ using $k = 5\%$ of the $N \in \{1000, 2500, 5000\}$ observations.

5.1.4 Parametric estimation methods

Next to the nonparametric estimation of the STDF discussed in the previous sections, many parametric models have been introduced for modelling the stable tail dependence function. Although finite-dimensional parametric models can never cover the full class of Pickands dependence functions, parametric models can be attractive because of their analytic tractability and intuition. A summary on parametric estimation for the extremal dependence structure can be found in, e.g., Kotz and Nadarajah (2000). A panoramic overview will be presented here to get a sense of alternative estimation methods.

Likelihood-based methods can be applied based on the multivariate extreme value distribution (MEVD) or the extreme value copula. The maximum likelihood estimation (MLE) for bivariate extremes has been the topic of many research efforts. See, for example, Smith (1994), Tawn and Ledford (1996), Boldi and Davison (2007), Coles and Tawn (1991), and De Haan et al. (2008). However, likelihood-based inference remains challenging, mostly due to the lack of simple forms of the likelihood, especially in higher dimensions. The focus has therefore been on the relatively simple Gumbel copula, although a likelihood-based estimator for the Hüsler-Reiss model has been recently discussed in Engelke et al. (2015). Padoan et al. (2010) suggest maximizing a composite pairwise likelihood that only considers dependence information between bivariate pairs to simplify the likelihood. Also see Huser and Davison (2013) for this approach. However, by only considering the tail dependence between bivariate margins, information on the true joint extremal dependence structure might be lost. Another drawback of the MLE approach is that the STDF has to be differentiable for the method to work. If the spectral measure underlying the STDF is discrete, or if the STDF is based on a factor model, the STDF is not differentiable, causing the MLE method to fail. A final drawback is that the asymptotic behavior of the MLE estimator is only known for the bivariate case.

To overcome the disadvantages of the maximum likelihood methods, Einmahl et al. (2008) and Einmahl et al. (2012) propose, respectively, bivariate and multivariate moment estimators

(M-estimators) for the STDF. This estimator does not require the ℓ-function to be differentiable and asymptotic results for d dimensions are known. The approach is however ill-adapted to higher dimensions. Therefore, akin to the composite likelihood methods, Einmahl et al. (2016b) develop a pairwise approach. As an alternative to the M-estimator and the pairwise variant thereof, Einmahl et al. (2017) propose a weighted least squares (WLS) estimator. Just as the M-estimator, it can handle nondifferentiable tail dependence functions, but it does not involve the integration of functions of many variables. The WLS estimator can, therefore, provide an attractive alternative for high dimensional estimation problems. Einmahl et al. (2017) show that the WLS estimator outperforms the pairwise M-estimator of Einmahl et al. (2016b), especially for weak tail dependence. For strong dependence, the estimators perform similarly in terms of mean squared error. The WLS estimator also turns out to be a strong competitor to likelihood-based methods. Furthermore, they find that their WLS estimator has an almost constant MSE as the dimension increases. Huser and Genton (2016) observe a similar pattern for the pairwise composite likelihood method. Both the M-estimator and the WLS-estimator are in fact semiparametric methods because they depend on either matching moments or minimizing the distance between a nonparametric initial estimate of the STDF and a parametric function of the STDF. Both methods are currently only developed for the Gumbel copula, which does not provide any flexibility to model asymmetric dependencies in dimensions $d > 2$. The focus of this research on inference for multivariate extremes is therefore on nonparametric estimation procedures for the STDF.

5.1.5 Simulation study

A simulation study is conducted to evaluate and compare estimation methods for the STDF. The aim is to retrieve a comprehensive overview of the behavior of the different estimators by considering a range of different dependence types, dependence strengths, and dimensions. Specifically, data are simulated from the Gumbel, t-, Normal, and Frank copula in dimensions $d \in \{2, 3, 5\}$. Model parameters are chosen in such a way that they correspond to bivariate Spearman correlations equal to $\rho = 0.2$, $\rho = 0.5$ and $\rho = 0.8$ for weak, medium and strong dependence. Note that this implies that all pairwise dependencies are equal. This should not affect the performance of the nonparametric estimators. We focus on the nonparametric estimators since the parametric models that are currently developed are not flexible enough to accurately model tail dependencies in higher dimensions. Before reporting the simulation results, we briefly discuss the performance measures used to evaluate the estimators, the simulations parameters, and the simulation running times.

Estimator performance measures

In line with common practice, estimates of the STDF and the TDC are evaluated based on the squared bias, variance and mean squared error (MSE), or integrated versions thereof. To evaluate the d-dimensional estimated $\hat{\ell}$-function in one point $\mathbf{x} \in [0,1]^d$ these quantities are defined as follows,

$$\begin{aligned}
\text{Squared bias} &: \left(\mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right]\right)^2 \\
\text{Variance} &: \mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{x}) - \mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{x})\right]\right)^2\right] \\
\text{MSE} &: \mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right)^2\right].
\end{aligned}$$

Since the asymptotic distribution of all nonparametric estimators introduced above are Gaussian with a covariance matrix that depends on the true ℓ and its partial derivatives in a complicated way, it is not straightforward to retrieve the variance of the estimator with these expressions. Therefore, as in e.g. Bücher and Dette (2013), we approximate the limiting distribution using resampling methods. Following, e.g., Bücher and Dette (2013) and Einmahl et al. (2016b), we evaluate these expected values with $B = 500$ replications. To evaluate the global behavior of the estimated $\hat{\ell}$-function, integrated versions of the aforementioned performance measures are considered. These

are defined as follows,

$$\text{Integrated squared bias} : \int_{[0,1]^d} \left(\mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{x})\right] - \ell(\mathbf{x}) \right)^2 \mathrm{d}\mathbf{x}$$
$$\text{Integrated variance} : \int_{[0,1]^d} \mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{x}) - \mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{x})\right] \right)^2 \right] \mathrm{d}\mathbf{x}$$
$$\text{Integrated MSE} : \int_{[0,1]^d} \mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x}) \right)^2 \right] \mathrm{d}\mathbf{x}.$$

Several approaches can be taken to compute these integrated performance measures, such as grid-based or Monte Carlo-based numerical integration procedures. However, these procedures are computationally expensive. Alternatively, Segers et al. (2017) describe a method based on Fubini's theorem to combine the integral with the expectation over the random sample into a single expectation. Specifically,

$$\text{Integrated squared bias} : \mathbb{E}\left[\left(\hat{\ell}^{(1)}_{n,k}(\mathbf{U}) - \ell(\mathbf{U}) \right) \left(\hat{\ell}^{(2)}_{n,k}(\mathbf{U}) - \ell(\mathbf{U}) \right) \right]$$
$$\text{Integrated variance} : \mathbb{E}\left[\frac{1}{2} \left(\hat{\ell}^{(1)}_{n,k}(\mathbf{U}) - \hat{\ell}^{(2)}_{n,k}(\mathbf{U}) \right)^2 \right]$$
$$\text{Integrated MSE} : \mathbb{E}\left[\frac{1}{2} \left(\left(\hat{\ell}^{(1)}_{n,k}(\mathbf{U}) - \ell(\mathbf{U}) \right)^2 + \left(\hat{\ell}^{(2)}_{n,k}(\mathbf{U}) - \ell(\mathbf{U}) \right)^2 \right) \right],$$

with $\hat{\ell}^{(1)}_{n,k}$ and $\hat{\ell}^{(2)}_{n,k}$ estimators based on two independent random samples drawn from the same distribution. A proof of the equivalence of these expectations to the integrated performance measures is included in Appendix C.1. In accordance with, e.g., Segers et al. (2017) and Kiriliouk et al. (2018) we employ this method with $B = 20,000$ pseudo-random samples to evaluate the integrated behavior of the nonparametric STDF estimators.

Simulation parameters

Several parameters have to be specified to perform the simulations. It has already been mentioned that we simulate from four different copula distributions, being the Gumbel, t-, Normal, and Frank copula. As discussed in Chapter 3, these copulas exhibit different tail dependence structures: the Gumbel copula is tail dependent and belongs to its own domain of attraction; the t-copula converges to the tail dependent t-EV copula; the Normal copula converges quite slowly to Π; the Frank copula converges very quickly to Π. The model parameters of the copulas are chosen in such a way that the pairwise dependencies are either strong ($\rho = 0.8$), medium ($\rho = 0.5$), or weak ($\rho = 0.2$).

The appropriate number of tail observations k from a dataset of size N to use for estimation is determined by a trade-off between bias and variance. When using little tail observations, the bias of the estimators is generally low, but the variance explodes. Conversely, using more tail observations will result in a lower variance, but an increased bias. A common solution is to show results for a range of k-values, but since we aim to present a wide range of simulation results, this is not a suitable solution. Therefore, to determine the appropriate number of tail observations k, initial simulations are conducted. Sample sizes of $N \in \{1000, 2500, 5000\}$ are considered. For financial data, these are in generally reasonable sample sizes, corresponding to respectively ca. 4, 10, and 20 years of daily trading day observations. Results for the nonparametric estimator based on simulations from the Gumbel, t-, Normal and Frank copula with medium dependence ($\rho = 0.5$) are shown in Appendix C.2. The simulations are repeated for several dimensions $d \in \{2, 3, 5, 7, 10\}$ and the integrated MSE is based on $B = 20,000$ samples.

For the Gumbel copula, it seems optimal to use approximately 5% of the sample as tail observations for all considered dimensions. With a sample size of $N = 2500$ this corresponds to $k = 125$ tail observations. However, for the other copulas, the optimal percentage of tail observations equals approximately 1% of the sample size. With a sample size of $N = 2500$ this corresponds to $k = 25$ tail observations. It should be noted that other dependence strengths, dependence types, or other estimators for the STDF could lead to other optimal values for k and N. Especially for the bias-corrected estimators, finite sample performances might be better for a lower threshold, thus using more tail observations. Based on these initial results it is decided that subsequent simulations be

conducted with a sample size of $N = 2500$ of which either $k = 25$ or $k = 125$ tail observations are used. Furthermore, for the bias-corrected estimators and the kernel-smoothed estimator it is necessary to specify additional parameters. Following Fougères et al. (2015), Beirlant et al. (2016) and Kiriliouk et al. (2018) we employ the power kernel with parameter $\tau = 5$ and set $a = r = 0.4$ for the bias-corrected estimators.

Running times

Before conducting all simulations, the running times of the different estimation procedures are assessed for one sample ($B = 1$). The results are shown in Table 5.1.1. Since it would take more than two full days to evaluate the global performance of the bias-corrected estimator of Beirlant et al. (2016) for only one type of distribution to simulate from, we will only consider this bias-corrected estimator for the TDC estimation, and not for the STDF estimation. The kernel-smoothed estimator is also only considered for the TDC estimation because of long running times.

Table 5.1.1: Running times to compute the STDF estimators for one sample (in seconds) and hypothetically for $B = 20,000$ samples (in hours, between brackets).

	NP (*)	(**)	NP-BC1 (*)	NP-BC2 (*)	NP-BC3 (*)	(**)	NP-Beta (*)	NP-Kernel (*)
$d = 2$	0.002 (0.1h)	0.001 (0.1h)	0.031 (1.7h)	0.024 (1.3h)	37.089 (2060.5h)	0.536 (29.8h)	0.002 (0.1h)	0.069 (3.8h)
$d = 3$	0.005 (0.3h)	0.001 (0.1h)	0.029 (1.6h)	0.028 (1.6h)	55.238 (3068.8h)	0.583 (32.4h)	0.007 (0.4h)	0.274 (15.2h)
$d = 5$	0.007 (0.4h)	0.002 (0.1h)	0.041 (2.3h)	0.04 (2.2h)	74.079 (4115.5h)	0.709 (39.4h)	0.004 (0.2h)	0.168 (9.3h)

(*) Own implementation of the estimator (see Appendix E.1). (**) Implementation of the estimator from the `TailDepFun` package in `R`. The running times are retrieved from a Macbook Pro with 2,7 GHz Intel Core i5 processor using the statistical software program `R`. NP denotes the empirical STDF nonparametric estimator, NP-Beta the beta-smoothed estimator of Kiriliouk et al. (2018), NP-Kernel the kernel-smoothed estimator, NP-BC1 and NP-BC2 the bias-corrected estimators of Fougères et al. (2015) and NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

Note that we fashioned implementations of all estimation procedures in `R` ourselves (see Appendix E.1 for the codes). For the standard empirical STDF and the bias-corrected estimator of Beirlant et al. (2016) the `TailDepFun` package offers functions, of which running times are also shown for comparison. The implementation of the bias-corrected estimator of Beirlant et al. (2016) in the `TailDepFun` package defers the lengthy computations involved for this estimator from `R` to `C++`, making the running time of this implementation much shorter. Therefore, we employ the bias-corrected estimator function from the `TailDepFun` package but otherwise rely on our own implementations.

Simulation results

Results for all simulations can be found in Appendix C.3. The main findings are summarized here.

Gumbel. For the simulations from the Gumbel copula, the empirical STDF and smoothed versions thereof perform similarly. The beta-smoothed estimator tends to slightly outperform the standard and kernel-smoothed nonparametric estimators based on the MSE measures. This is mostly due to the decreased variance inherent to the smoothed version of the estimator. The kernel-smoothed estimator also reduces the variance, but to a lesser extent. Moreover, the kernel-smoothed estimator tends to increase the bias associated with the estimator, whereas the beta-smoothed estimator decreases the bias in some occasions. The bias-corrected versions of the empirical STDF do not offer significant improvements. The third bias-corrected version of Beirlant et al. (2016) performs quite similar to the standard empirical STDF and the smoothed versions in terms of the MSE for the TDC. Although this bias-corrected estimator tends to decrease the bias for the TDC estimate, the increase in variance leads to similar MSE results. The other two bias-corrected versions of Fougères et al. (2015) yield worse results than the standard and smoothed estimators.

Notably, these two estimators do not tend to decrease the bias but do significantly increase the variance. For all considered dimensions the estimators perform better for the higher threshold $k = 125$. However, compared to the performance of the estimators for the data simulated from other copulas, the performance of the estimators for the Gumbel copula with the lower threshold of $k = 25$ is still by far superior. The performance of the estimators for the multivariate TDC Λ based on data from the Gumbel copula with medium dependence are summarized in Table 5.1.2.

t. For the simulations from the t-copula, similar patterns as for the simulations from the Gumbel copula are noticeable although most values are higher than for the results of the Gumbel copula. In all cases, the performance of either the beta-smoothed or the kernel-smoothed nonparametric estimator is better than the performance of the standard nonparametric estimator regarding the MSE. This is mostly due to a reduction in the variance, but the bias also tends to be reduced for the smoothed versions of the estimator. For all considered dimensions the empirical STDF and smoothed estimators perform better for the lower threshold $k = 25$. In the bivariate case, bias-corrections only offer improvements if the higher threshold of $k = 125$ is employed. In this case, the first bias-corrected version of Fougères et al. (2015) effectively reduces the bias associated with the uncorrected estimators for light tail dependence, while the third bias-corrected version yields results similar to the uncorrected estimators. For medium and strong dependence, the third bias-corrected version of Beirlant et al. (2016) decreases bias more effectively. For higher dimensions, the third-bias-corrected version of Beirlant et al. (2016) outperforms consistently and offers a lower MSE for all dependence strengths but mostly for light tail dependence. Results based on the higher threshold of $k = 125$ yield better results, but employing the lower threshold of $k = 25$ also improves the MSE relative to the uncorrected estimators. The performance of the estimators for the multivariate TDC Λ based on data from the t-copula with medium dependence are summarized in Table 5.1.3.

Normal. For the Normal copula, extraordinarily high bias terms are observed. Differences between dependence strengths seem to be the determining factor of estimator performance. For simulations with strong dependence, the bias is exceptionally high, even for the bias-corrected nonparametric estimator. Both the first and third bias-corrected estimators improve the MSE by drastically decreasing the bias component: in most instances, the bias is more than halved. The third bias-corrected estimator of Beirlant et al. (2016) tends to yield the best results in terms of MSE for the TDC estimate in all dimensions. In some instances, the second bias-corrected estimator of Fougères et al. (2015) yields the lowest bias, but it is not consistently able to do so. For the smoothed estimators, similar patterns as for the tail dependent data can be observed: the smoothed versions yield slightly better results. Notably, the kernel-smoothed estimator tends to result in a slightly lower MSE compared to the beta-smoothed estimator. However, due to the substantial bias term for the Normal copula data, the bias-corrected versions outperform by far. Almost all results are better for the lower threshold $k = 25$, also for the bias-corrected versions. The bias-corrected results provide an exception for light tail dependence in the bivariate case. The performance of the estimators for the multivariate TDC Λ based on data from the Normal copula with medium dependence are summarized in Table 5.1.4.

Frank. For the simulations from the Frank copula, MSE values are quite low overall, especially for light tail dependence. The beta-copula version of the nonparametric estimator slightly outperforms the standard nonparametric estimator in terms of MSE, due to a decrease in variance, and, in some instances, also to a decrease in bias. The kernel-smoothed estimator tends to yield an MSE similar to the standard empirical STDF. Although it slightly reduces the variance in most cases, the bias is sometimes increased. In most cases, the bias-corrected estimators offer improvements relative to the empirical STDF and smoothed version thereof, especially for medium and light tail dependence. Both the first bias-corrected estimator of Fougères et al. (2015) and the third bias-corrected estimator of Beirlant et al. (2016) consistently offer an improved or similar MSE. The second bias-corrected estimator of Fougères et al. (2015) offers the best bias correction is some instances, but in other cases also shows extremely high MSE values. Therefore, we do not consider this estimator to be reliable. Most results are better for the lower threshold $k = 25$, also for the bias-corrected versions. The performance of the estimators for the multivariate TDC Λ based on data from the Frank copula with medium dependence are summarized in Table 5.1.5.

Table 5.1.2: Overview of performance measures (×1000) for multivariate TDC Λ-estimators for data simulated from the Gumbel copula with medium dependence.

	$d = 2$			$d = 3$			$d = 5$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
NP	0.028	7.7	7.7	0.00011	5.3	5.3	0.021	3.4	3.4
NP-Beta	0.089	6	6	0.2	4.8	5	0.02	3.3	3.3
NP-Kernel	0.21	7.9	8.1	0.0066	5.7	5.7	0.064	4.5	4.5
NP-BC1	1	27	28	0.49	20	20	0.61	18	19
NP-BC2	0.069	40	40	0.84	31	32	0.51	40	40
NP-BC3	0.029	7.3	7.3	0.0058	5.8	5.8	0.12	4.5	4.6

Results are based on $B = 500$ samples of sample size $N = 2500$ of which $k = 25$ tail observations are used. NP denotes the empirical STDF nonparametric estimator, NP-Beta the beta-smoothed estimator of Kiriliouk et al. (2018), NP-Kernel the kernel-smoothed estimator, NP-BC1 and NP-BC2 the bias-corrected estimators of Fougères et al. (2015) and NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

Table 5.1.3: Overview of performance measures (×1000) for multivariate TDC Λ-estimators for data simulated from the t-copula with medium dependence.

	$d = 2$			$d = 3$			$d = 5$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
NP	2	6.4	8.4	5.4	4.4	9.8	11	2.8	14
NP-Beta	1.7	5.4	7	4.8	3.8	8.5	10	2.4	13
NP-Kernel	3.1	6.7	9.8	5.5	4.4	9.9	11	2.9	14
NP-BC1	0.056	22	22	0.83	17	18	1.7	12	14
NP-BC2	0.9	39	40	0.42	34	34	0.42	57	58
NP-BC3	2.5	7.5	10	1.8	4.8	6.6	0.32	6.5	6.8

Results are based on $B = 500$ samples of sample size $N = 2500$ of which $k = 25$ tail observations are used. NP denotes the empirical STDF nonparametric estimator, NP-Beta the beta-smoothed estimator of Kiriliouk et al. (2018), NP-Kernel the kernel-smoothed estimator, NP-BC1 and NP-BC2 the bias-corrected estimators of Fougères et al. (2015) and NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

Table 5.1.4: Overview of performance measures (×1000) for multivariate TDC Λ-estimators for data simulated from the Normal copula with medium dependence.

	$d = 2$			$d = 3$			$d = 5$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
NP	17	3.9	20	30	2.9	33	59	1.8	61
NP-Beta	16	3.2	20	30	2.3	32	59	1.7	61
NP-Kernel	18	3.8	22	30	2.9	33	58	2.2	60
NP-BC1	6.3	7.8	14	8.5	7.3	16	21	6.3	27
NP-BC2	7.1	13	20	8	16	24	4.4	27	32
NP-BC3	9.6	4.5	14	5.1	3.7	8.8	0.26	2.7	3

Results are based on $B = 500$ samples of sample size $N = 2500$ of which $k = 25$ tail observations are used. NP denotes the empirical STDF nonparametric estimator, NP-Beta the beta-smoothed estimator of Kiriliouk et al. (2018), NP-Kernel the kernel-smoothed estimator, NP-BC1 and NP-BC2 the bias-corrected estimators of Fougères et al. (2015) and NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

Table 5.1.5: Overview of performance measures (×1000) for multivariate TDC Λ-estimators for data simulated from the Frank copula with medium dependence.

	$d=2$			$d=3$			$d=5$		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
NP	0.96	1.3	2.2	2.7	0.95	3.6	6.7	0.68	7.3
NP-Beta	1.1	0.92	2	2.7	0.65	3.3	6.5	0.64	7.1
NP-Kernel	2	1.2	3.3	2.8	0.73	3.5	6	0.6	6.5
NP-BC1	0.48	2	2.5	0.29	0.79	1.1	0.23	0.63	0.86
NP-BC2	0.82	5.7	6.5	0.92	6.4	7.3	86	182	267
NP-BC3	0.0015	0.086	0.087	0.17	0.68	0.85	5.3	0.67	6

Results are based on $B = 500$ samples of sample size $N = 2500$ of which $k = 25$ tail observations are used. NP denotes the empirical STDF nonparametric estimator, NP-Beta the beta-smoothed estimator of Kiriliouk et al. (2018), NP-Kernel the kernel-smoothed estimator, NP-BC1 and NP-BC2 the bias-corrected estimators of Fougères et al. (2015) and NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

We conclude that the beta-smoothed version of the empirical STDF yields slightly superior results compared to the standard empirical STDF in terms of (integrated) MSE. The bias-corrected version of the empirical STDF of Beirlant et al. (2016) offers a vast improvement when the copula convergence rate is slow. Furthermore, it has been observed that the estimators perform better for data simulated from the Gumbel copula than from the other copula models. This can be explained by the fact that the Gumbel copula belongs to its own maximum domain of attraction, whereas the t-copula converges to the t-EV copula. Especially the integrated MSE for the simulations from the Normal copula is much higher. Since the Normal copula converges relatively slow to the independent extreme value copula, this explains the relatively large error terms in the estimator. The Frank copula, on the other hand, converges relatively fast to the independent extreme value copula, which explains why the errors for the Frank copula are smaller than for the Normal copula.

We have established the relative behavior of estimators for different dependence structures, dependence strengths, and dimensions. A final note on the absolute performance of the estimators is given here since it is important to be aware of the bias and variance that can be expected for estimates regarding the STDF. Table 5.1.6 provides a quick overview of the bias and variance regarding estimates for the multivariate TDC Λ. Considering that the multivariate TDC is bounded as $0 \leq \Lambda \leq 1$, most bias terms are quite satisfactory. When using the bias-corrected estimator of Beirlant et al. (2016), the multivariate TDC Λ can be estimated with at least one decimal place accuracy. The standard empirical STDF yields a higher expected bias in most cases, with a maximum realized bias of 0.24 for the 5-dimensional Normal copula. The variance of the estimators is similar and approximately equal to 0.05. For tail dependent data, variances are a bit higher than for tail independent data.

Table 5.1.6: Overview of bias and standard error (between brackets) of multivariate TDC (Λ) estimates for a selection of estimators and simulated data.

	Gumbel		t		Normal		Frank	
	NP	NP-BC3	NP	NP-BC3	NP	NP-BC3	NP	NP-BC3
$d=2$	0.005 (0.088)	0.005 (0.085)	0.045 (0.080)	0.050 (0.087)	0.081 (0.062)	0.098 (0.067)	0.031 (0.036)	0.001 (0.029)
$d=3$	0.000 (0.073)	0.002 (0.076)	0.073 (0.066)	0.042 (0.069)	0.173 (0.054)	0.071 (0.061)	0.052 (0.031)	0.013 (0.026)
$d=5$	0.004 (0.058)	0.011 (0.067)	0.105 (0.053)	0.018 (0.081)	0.243 (0.052)	0.016 (0.052)	0.082 (0.026)	0.073 (0.026)

Results are based on $B = 500$ samples of sample size $N = 2500$ of which $k = 25$ tail observations are used. Data are simulated from the Gumbel, t-, Normal, and Frank copula with medium dependence (i.e., parameters are chosen such that $\rho = 0.5$ for all bivariate margins) in dimensions $d = 2, 3, 5$. NP denotes the empirical STDF estimator. NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

5.2 Testing tail independence

To determine whether data are asymptotically independent or asymptotically dependent, a hypothesis test can be conducted. It is instructive to test for tail independence before drawing any conclusions from estimations to prevent incorrect inferences on the strength of tail dependence, especially in the case of tail independence. Moreover, testing whether data are asymptotically independent can also be a good starting point for analysis of multivariate extremes since the trouble of determining the tail dependence structure can be saved when data turn out to be tail independent. This section is therefore devoted to the problem of testing tail independence.

Hypothesis testing setup

Recall that in the hypothesis testing framework, a null hypothesis H_0 is tested against an alternative hypothesis H_1. This is done by evaluating a test statistic for a given dataset and comparing the observed value to the theoretical distribution of the test statistic under the null hypothesis. The probability of observing the given or more extreme value of the test statistic under the null hypothesis is known as the p-value. The null hypothesis is rejected if the p-value is below the significance level $\alpha \geq \mathbb{P}(\text{reject } H_0|H_0)$. The significance level is chosen by the statistician to control the probability of making a type I error, which corresponds to rejecting the null hypothesis H_0 while H_0 is true. Conversely, a type II error corresponds to incorrectly retaining the null hypothesis H_0. The probability of making a type II error is measured by the power of the test, $1 - \beta = \mathbb{P}(\text{reject } H_0|H_1)$. Table 5.2.1 summarizes this hypothesis testing setup.

Table 5.2.1: Hypothesis testing framework.

	H_0 true	H_1 true
Retain H_0	Correct Size of the test: $1 - \alpha$	Type II error Probability: β
Reject H_0	Type I error Significance level: α	Correct Power of the test: $1 - \beta$

Generally, the testing setup is designed such that the type I error is more "expensive" to make than the type II error. Since it depends on the situation whether it is more appropriate to take tail independence or tail dependence as the null hypothesis, one would like to be able to conduct both versions of the testing problem.

Tail dependence testing literature

The hypothesis testing problem with tail dependence as null hypothesis has been addressed in multiple instances in literature, but only for the bivariate case. See for example Tawn and Ledford (1996), Coles et al. (1999), and Draisma et al. (2004). The tests are based on the tail order measure, which is equal to $\kappa = 2$ under the null hypothesis of tail dependence in dimension $d = 2$ (see Section 4.3.1). The tail order κ can be estimated with estimation procedures for the univariate tail index, e.g., the Hill estimator or the moment estimator, and test statistics are based on known asymptotic distributions of these estimators. Since a tail order of $\kappa = d$ does not uniquely correspond to tail dependence for higher dimensions, these tests cannot be extended for the multivariate problem under consideration.

The hypothesis testing problem with tail independence as null hypothesis, however, has remained largely untouched in literature. This probably has to do with the degenerate distribution of the empirical STDF under the null hypothesis of tail independence, which will be discussed in Section 5.2.1. Hüsler and Li (2009) introduce a method to handle the bivariate variant of this hypothesis testing problem by adjusting the empirical STDF estimator. The multivariate variant of the hypothesis testing problem has not been addressed, however. An exception is given by the testing method proposed in Falk and Michel (2006) who develop testing procedures for negative random variables based on their radial component. Negative random variables do not fit adequately into our framework and although they propose a testing procedure for higher dimensions, these tests are not implemented or further researched in their work.

In order to contribute to a solution of this gap in literature, we develop testing procedures for the tail dependence structure of *multivariate* random vectors. Specifically, we investigate the following problem,

$$H_0 : (X_1, \ldots, X_d) \sim C_F : C_F \in MDA(\Pi) \qquad H_1 : (X_1, \ldots, X_d) \sim C_F : C_F \notin MDA(\Pi).$$

That is, the null hypothesis of tail independence is tested against the alternative hypothesis of tail dependence. For this testing problem, a type I error corresponds to incorrectly assuming tail dependence, and a type II error corresponds to incorrectly assuming tail independence. We propose two methods to tackle the multivariate variant of the hypothesis testing problem: (1) conducting a multivariate test based on an extension of the bivariate test (Section 5.2.1) or (2) conducting multiple bivariate tests (Section 5.2.2). Most importantly, we propose a new test based on the multivariate tail dependence coefficient Λ. The different approaches to the testing problem are evaluated with a simulation study presented in Section 5.2.3.

5.2.1 Multivariate tests

In order to construct an hypothesis test for the testing problem, a test statistic with a known distribution under the null hypothesis of tail independence has to be constructed. Since the stable tail dependence function fully captures the extremal dependence structure of two or more random variables, the STDF is a natural candidate to use. Recall that under the null hypothesis of tail independence, the STDF is given by $\ell(x_1, \ldots, x_d) = x_1 + \cdots + x_d$ and that

$$\sup_{\mathbf{x} \in [0,1]^d} \left| \sqrt{k} \left(\hat{\ell}_{n,k}(x_1, \ldots, x_d) - \ell(x_1, \ldots, x_d) \right) - B_\ell(x_1, \ldots, x_d) \right| \to 0$$

(see Theorem 5.1), where the zero-mean Gaussian process B_ℓ defined in Equation 5.1.5, takes the following form under the null hypothesis of tail independence,

$$B_\ell(x_1, \ldots, x_d) = W_\ell(x_1, \ldots, x_d) - \sum_{i=1}^{d} W_{\ell,j}(x_j),$$

since the first order partial derivatives of the independent ℓ-function are all equal to 1. Consequently, the asymptotic variance is given by

$$\begin{aligned} \mathbb{E}\left(B_\ell(x_1, \ldots, x_d)^2\right) &= \ell(x_1, \ldots, x_d) - 2\left(\ell(x_1, 0, \ldots, 0) + \cdots + \ell(0, \ldots, 0, x_d)\right) \\ &\quad + \left(\ell(x_1, 0, \ldots, 0) + \ell(0, \ldots, 0, x_d)\right) \\ &= (x_1 + \cdots + x_d) - 2(x_1 + \cdots + x_d) + (x_1 + \cdots + x_d) = 0. \end{aligned}$$

Hence, based on the asymptotic properties of the nonparametric estimator of the stable tail dependence function, it can be established that under H_0,

$$\sup_{\mathbf{x} \in [0,1]^d} \sqrt{k} \left| \hat{\ell}_{n,k}(x_1, \ldots, x_d) - (x_1 + \cdots + x_d) \right| \to 0.$$

Apparently, the asymptotic distribution of the empirical STDF becomes degenerate under the null hypothesis of tail independence, making it impossible to construct a test statistic based on this STDF estimator. Hüsler and Li (2009) noted that the asymptotic variance vanishes because of the dependence between high threshold levels and data that are used to assess whether these thresholds are exceeded. This dependence stems from the notion that thresholds are based on ranks of the same data used to determine whether thresholds are exceeded. It is obvious that when based on the same data, high quantiles are dependent on the data observations. Hüsler and Li (2009) proposed to use different estimators for the tail quantiles which are independent of data observations used to estimate the empirical copula in order to prevent the asymptotic variance from vanishing. This can be accomplished, for example, by dividing the sample into two equally-sized subsamples and using one for estimating the quantiles and the other one for estimating the empirical copula. Since the two subsamples are independent, the asymptotic variance does not vanish. Extending the argument of Hüsler and Li (2009) to higher dimensions, this leads to the following adapted definition of the estimator for the STDF which does not converge to a degenerate asymptotic distribution.

Definition 5.5. *Let* $(X_{11},\dots,X_{1d}),\dots,(X_{n1},\dots,X_{nd})$, $(\tilde{X}_{11},\dots,\tilde{X}_{1d}),\dots,(\tilde{X}_{m1},\dots,\tilde{X}_{md})$ *be iid observations from a* d*-dimensional copula* C_F *such that* $m/n \to \theta > 0$ *and* $m/n - \theta = \mathcal{O}(n^{-1/2})$. *Let* $k := k_n \in \{1,\dots,n\}$ *be such that* $k \to \infty$, *and* $k/n \to 0$ *if* $n \to \infty$. *The adjusted empirical STDF is given by*

$$\overline{\ell}_{n,k}(x_1,\dots,x_d) = \frac{m}{k}\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{X_{i1} \geq \tilde{X}_{m+1/2-\lceil kx_1\rceil:m,1} \text{ or } \dots \text{ or } X_{id} \geq \tilde{X}_{m+1/2-\lceil kx_d\rceil:m,d}\right\}, \tag{5.2.1}$$

where $\tilde{X}_{m+1/2-\lceil kx_j\rceil:m,j}$ *are the order statistics of* $(\tilde{X}_{1,j},\dots,\tilde{X}_{m,j})$, $j = 1,\dots,d$, *and* $\lceil a\rceil$ *denotes the smallest integer larger than or equal to* a.

To get a sense of the behavior of the adjusted estimator, the STDF is estimated for data simulated from the bivariate Gumbel copula. The results are shown in Figure 5.2.1. Compared to the original empirical STDF, shown in Figure 5.1.1, the adjusted estimator looks much worse. Since data to determine ranks and data to determine threshold exceedances are no longer the same, additional variance is introduced and the estimator deviates from the theoretical bounds on the ℓ-function. Moreover, due to the division of the sample into two parts, there is less data available to estimate the dependence structure resulting in less accurate estimates.

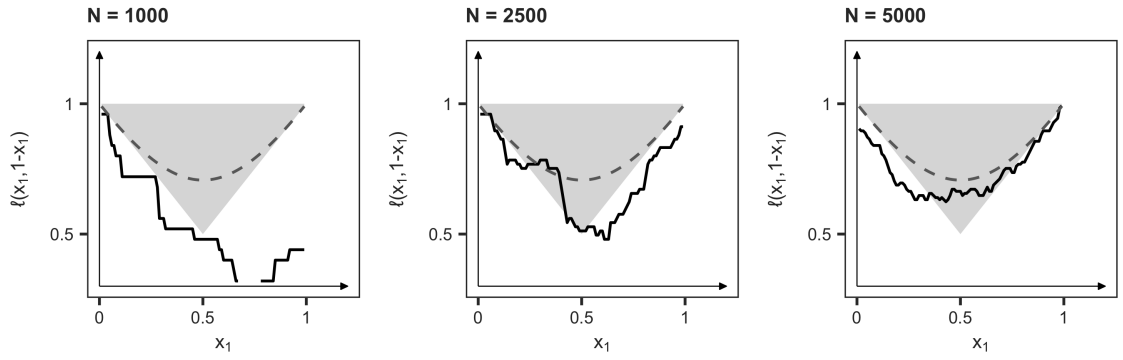

Figure 5.2.1: Estimated adjusted empirical STDF $\overline{\ell}$-function based on simulation results for the bivariate Gumbel copula with model parameter $\theta = 2$ using $k = 5\%$ of the $N \in \{1000, 2500, 5000\}$ observations.

It is clear that as an estimator for the STDF, the adjusted estimator is not an improvement relative to the original empirical STDF. However, due to the additional variance, the asymptotic distribution of this adjusted empirical STDF is no longer degenerate under the null hypothesis of tail independence and can therefore be used to construct a test statistic. Formally, the asymptotic behavior of this estimator is given by the following theorem.

Theorem 5.3. *Under the null hypothesis of tail independence and with the conditions of Theorem 5.1, supplemented with the conditions that* $m/n \to \theta > 0$ *and* $m/n - \theta = \mathcal{O}(n^{1/2})$, *it holds that*

$$\left\{\sqrt{k}\left(\overline{\ell}_{n,k}(x_1,\dots,x_d) - (x_1+\dots+x_d)\right), \mathbf{x} \in [0,1]^d\right\} \to \sum_{j=1}^{d} W_j\left((1+\theta)x_j\right) =: \overline{B}(\mathbf{x}) \tag{5.2.2}$$

as $n \to \infty$, *where* $W_j, j = 1,\dots,d$, *are* d *independent Brownian motions.*

Proof. The proof is shown in Appendix B.4.2. Note that we fashioned our own proof based on Theorem 5.3 and thereby employ a different approach to showing this result than Hüsler and Li (2009) for their bivariate results. ■

Based on this asymptotic result and inspired by the test statistics of Hüsler and Li (2009) for the bivariate problem, we propose two test statistics in the same fashion as the Cramer-von-Mises and Kolmogorov-Smirnov goodness-of-fit tests. The first one is based on the integrated squared difference between the estimated ℓ-function and the true independent ℓ-function,

$$I_n = \int_{[0,1]^d} k\left(\overline{\ell}_{n,k}(x_1,\dots,x_d) - (x_1+\dots+x_d)\right)^2 \mathrm{d}x_1 \dots \mathrm{d}x_d, \tag{5.2.3}$$

and the second one is based on the maximum absolute difference between the estimated ℓ-function and the true independent ℓ-function,

$$S_n = \sup_{\mathbf{x}\in[0,1]^d} \sqrt{k}\left|\overline{\ell}_{n,k}(x_1,\ldots,x_d) - (x_1+\cdots+x_d)\right|. \tag{5.2.4}$$

The asymptotic distributions of the test statistics I_n and S_n are given in the following proposition.

Proposition 5.3. *Under the null hypothesis of tail independence,*

$$\{I_n\} \to \int_{[0,1]^d} \left(\sum_{j=1}^{d} W_j\left((1+\theta)x_j\right)\right)^2 \mathrm{d}x_1\ldots\mathrm{d}x_d \quad \text{and} \quad \{S_n\} \to \sup_{\mathbf{x}\in[0,1]^d}\left|\sum_{j=1}^{d} W_j\left((1+\theta)x_j\right)\right|$$

as $n\to\infty$, where $W_j, j=1,\ldots,d$ are d independent Brownian motions and $\theta = m/n$ denotes the ratio between divided sample sizes.

Proof. Both convergences follow immediately from the continuous mapping theorem. ■

Since the distributions of the test statistics I_n and S_n cannot be easily evaluated analytically, a simulation procedure is employed to determine the quantiles of the distributions. Based on 100,000 realizations of the Wiener process, critical values are determined (Table 5.2.2). The simulated values for dimension $d = 2$ are very similar to the quantiles presented in Hüsler and Li (2009) who use 200,000 realizations of the Wiener process for the bivariate case, although not entirely the same. A disadvantage of this approach is that the distributions of the test statistics cannot be evaluated analytically and that simulations to determine distributions of the test statistics I_n and S_n are quite computationally expensive, especially for higher dimensions. For example, it took respectively 6.3, 14.7, and 29.6 hours to draw 100,000 samples from both the distribution of I_n and S_n for dimensions $d = 2, 3, 5$ (employing the statistical software program `R` on a Macbook Pro with 2,7 GHz Intel Core i5 processor).

Table 5.2.2: Simulated quantiles for the test statistics distributions in dimensions $d \in \{2,3,5\}$.

		10%	25%	50%	75%	90%	95%	97.5%	99%
$\mathbf{d=2}$	I_n	0.397	0.66	1.25	2.51	4.54	6.23	7.94	10.3
	S_n	1.10	1.58	2.28	3.13	4.00	4.57	5.08	5.71
$\mathbf{d=3}$	I_n	0.685	1.07	1.91	3.72	6.64	9.08	11.6	15.3
	S_n	1.45	2.08	2.97	4.05	5.11	5.82	6.45	7.18
$\mathbf{d=5}$	I_n	1.31	1.90	3.2	6.14	11.0	14.9	19.1	25.1
	S_n	2.05	2.89	4.1	5.52	6.94	7.86	8.68	9.67

Based on 100,000 realizations of the Brownian motion path. Samples are split into two equally sized subsamples ($\theta = 1$).

TDC-based testing

Since the STDF-based TDC measure Λ can be used to identify tail independence in all dimensions $d \geq 2$ (Main Theorem 2), this measure can also be used to test for tail independence. However, since the measure is based on the stable tail dependence function, it suffers from the same problematic characteristic of the asymptotic variance vanishing under the null hypothesis of tail independence. In order to circumvent this problem, the same strategy as for the ℓ-based tests is employed, which comes down to evaluating the adapted STDF estimator in the point $(1,\ldots,1)$. Hence, the test statistic based on the multivariate TDC Λ is given by,

$$T_n = \sqrt{k}\left(\overline{\ell}_{n,k}(1,\ldots,1) - d\right). \tag{5.2.5}$$

The asymptotic distribution of T_n follows from Theorem 5.3 and is captured by the following proposition.

Proposition 5.4. *Under the null hypothesis of tail independence and with the conditions of Theorem 5.3, we have that*

$$T_n \to N\left(0, d(1+\theta)\right), \tag{5.2.6}$$

as $n \to \infty$, with $\theta = m/n$ the ratio between divided sample sizes.

Proof. Theorem 5.3 immediately implies that

$$T_n = \sqrt{k}\left\{\overline{\ell}_{n,k}(1,\ldots,1) - d\right\} \to \left\{\sum_{i=1}^{d} W_i(1+\theta)\right\}, \tag{5.2.7}$$

as $n \to \infty$, where $W_i, i = 1, \ldots, d$ are d independent Brownian motions. The result follows by the definition of a Brownian motion. ■

Since $\ell(1,\ldots,1) = d$ is the theoretical upper bound, a one-sided test based on the lower tail of the asymptotic Normal distribution of T_n should be conducted. Note that this asymptotically Normal distribution is much easier to evaluate than asymptotic distributions of the STDF-based test statistic I_n and S_n.

Testing based on smoothed estimators

Since the simulation study on the performance of the nonparametric STDF estimators showed that the standard nonparametric estimator was outperformed in many instances by the smoothed and bias-corrected versions, a test statistic based on one of these improved versions of the nonparametric estimator might also yield superior testing performance. This section covers an adjusted kernel smoothed estimator. The next section covers an adjusted bias-corrected smoothed parameter. Since the beta-smoothed STDF is by definition determined based on the ranks of observations, the methodology cannot be easily split up in a similar way as proposed by Hüsler and Li (2009). The kernel-smoothed STDF can be adapted as follows.

$$\overline{\ell}^{K}_{n,k}(\mathbf{x}) = \frac{\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-1} \overline{\ell}_{n,k}(a_j \mathbf{x})}{\frac{1}{k}\sum_{j=1}^{k} K(a_j)}, \tag{5.2.8}$$

where $a_j = \frac{j}{k+1}$ and K is a kernel on $(0,1)$ such that $\int_0^1 K(u)\mathrm{d}u = 1$.

The behavior of the adjusted kernel-smoothed empirical STDF is shown in Figure 5.2.2. Compared to the estimates of the adjusted empirical STDF (see Figure 5.2.1), the kernel-smoothed version seems to perform better. The estimates are closer to the true dependence function and due to the smoothing, the estimates resemble the real STDF more closely. However, the estimator does not honor the theoretical bounds of the ℓ-function and requires a large sample size to yield acceptable results. Similar as for the standard adjusted estimator, this adjustment is not an improvement relative to the original kernel-smoothed estimator (see Figure 5.1.3).

The asymptotic behavior follows easily from Theorem 5.3 and is stated in the following proposition.

Proposition 5.5. *Under the null hypothesis of tail independence and with the assumptions of Theorem 5.3,*

$$\sqrt{k}\left\{\overline{\ell}^{K}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right\} \to \left(\int_0^1 K(u) u^{-1/2} \mathrm{d}u\right) \overline{B}(\mathbf{x}). \tag{5.2.9}$$

in $D([0,T]^d)$ for every $T > 0$ as $n \to \infty$. Here $\overline{B}(\mathbf{x})$ is a zero-mean Gaussian process defined in Equation 5.2.2.

Proof. We start from the asymptotic behavior of the adapted empirical STDF presented in Theorem 5.3, i.e.,

$$\sqrt{k}\left(\overline{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right) \to \overline{B}(\mathbf{x}).$$

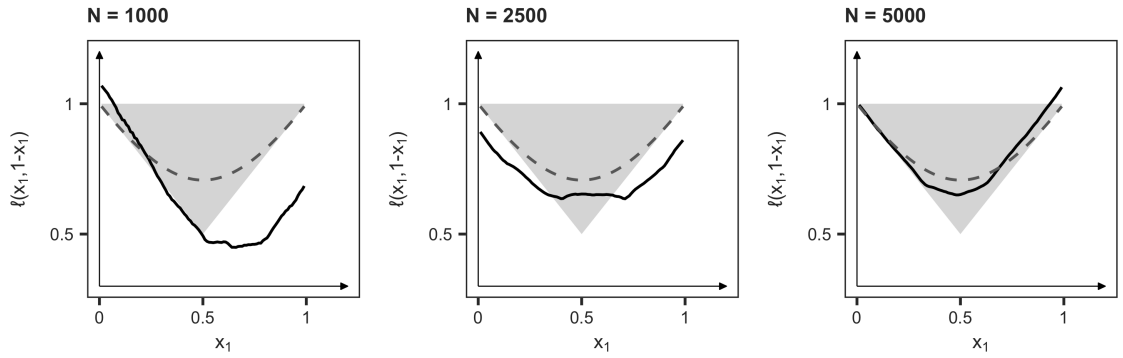

Figure 5.2.2: Estimated adjusted kernel-smoothed empirical STDF $\overline{\ell}^K$-function based on simulation results for the bivariate Gumbel copula with model parameter $\theta = 2$ using $k = 5\%$ of the $N \in \{1000, 2500, 5000\}$ observations.

By the continuous mapping theorem,

$$\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^{k} K(a_j)\overline{\ell}_{n,k}(\mathbf{x}) - \sum_{j=1}^{k} K(a_j)\ell(\mathbf{x})\right) \to \left(\int_0^1 K(u)\mathrm{d}u\right)\overline{B}(\mathbf{x}) = \overline{B}(\mathbf{x}).$$

By the homogeneity of ℓ and the fact that $\overline{B}(a\mathbf{x}) \sim \sqrt{a}\overline{B}(\mathbf{x})$ it next follows that

$$\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^{k} K(a_j)a_j^{-1}\overline{\ell}_{n,k}(a_j^{-1}\mathbf{x}) - \sum_{j=1}^{k} K(a_j)\ell(\mathbf{x})\right) \to \left(\int_0^1 K(u)u^{-1/2}\mathrm{d}u\right)\overline{B}(\mathbf{x}).$$

The result follows by dividing everything by $\sum_{j=1}^{k} K(a_j)$. ■

Recall that by taking the power kernel $K(t) = (\tau+1)t^\tau \mathbb{1}\{t \in (0,1)\}$, $\tau > -1/2$, we get

$$\sqrt{k}\left(\overline{\ell}_{n,k}^K(\mathbf{x}) - \ell(\mathbf{x})\right) \to 2\frac{1+\tau}{1+2\tau}\overline{B}(\mathbf{x}),$$

which is convenient because the asymptotic distribution then becomes a rescaled version of the asymptotic distribution of the non-bias-corrected estimator $\overline{\ell}_{n,k}$. The test statistics and their asymptotic distributions can now be defined similarly as before. Firstly, the integral-based test statistic is given by,

$$\begin{aligned} I_n^K &= \int_{[0,1]^d} k\left(\overline{\ell}_{n,k}^K(x_1,\ldots,x_d) - (x_1+\cdots+x_d)\right)^2 \mathrm{d}x_1\ldots\mathrm{d}x_d \\ &\to \left(2\frac{1+\tau}{1+2\tau}\right)^2 \int_{[0,1]^d} \overline{B}(\mathbf{x})^2 \mathrm{d}x_1\ldots\mathrm{d}x_d. \end{aligned}$$

The supremum-based test statistic is defined as,

$$\begin{aligned} S_n^K &= \sup_{\mathbf{x}\in[0,1]^d} \sqrt{k}\left|\overline{\ell}_{n,k}^K(x_1,\ldots,x_d) - (x_1+\cdots+x_d)\right| \\ &\to 2\frac{1+\tau}{1+2\tau}\sup_{\mathbf{x}\in[0,1]^d}\left|\overline{B}(\mathbf{x})\right|. \end{aligned}$$

Finally, the multivariate TDC-based test statistic is equal to the following,

$$\begin{aligned} T_n^K &= \sqrt{k}\left(\overline{\ell}_{n,k}^K(1,\ldots,1) - d\right) \\ &\to 2\frac{1+\tau}{1+2\tau}\overline{B}(1,\ldots,1) \sim N\left(0, \left(2\frac{1+\tau}{1+2\tau}\right)^2 d(1+\theta)\right). \end{aligned}$$

Note that the quantiles in Table 5.2.2 just have to be multiplied with a factor dependent on τ to retrieve the new quantiles of the asymptotic distribution of the test statistics.

Testing based on bias-corrected estimators

Moreover, we also show that the bias-corrected STDF estimator of Beirlant et al. (2016) can be adapted in a similar way to retrieve a bias-corrected estimator that does not become degenerate under the null hypothesis of tail independence.

Proposition 5.6. *Under the null hypothesis of tail independence and with the assumptions of Theorem 5.2,*

$$\sqrt{k}\left\{\overline{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x}) - \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\right\} \to \overline{B}(\mathbf{x}) \tag{5.2.10}$$

in $D([0,T]^d)$ *for every* $T > 0$ *as* $n \to \infty$. *Here* $\overline{B}(\mathbf{x})$ *is a zero-mean Gaussian process defined in Equation 5.2.2.*

Proof. The result follows by combining Theorem 5.2 and Theorem 5.3. ■

In order to remove the bias term, we employ the methodology of Beirlant et al. (2016) and arrive at the following adapted bias-corrected estimator.

$$\overline{\ell}^{BC3}_{n,k,\overline{k}}(\mathbf{x}) = \frac{\tilde{\ell}_{n,k}(\mathbf{x}) - \left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \tilde{\alpha}_{\overline{k}}(\mathbf{x}) \frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)}}{\frac{1}{k}\sum_{j=1}^{k} K(a_j)}. \tag{5.2.11}$$

The behavior of the estimator is explored in Figure 5.2.3. The figures indicate quite irregular behavior and show that a very large sample size is required to retrieve slightly reasonable results. Compared to the original bias-corrected estimator of Beirlant et al. (2016), illustrated in Figure 5.1.8, the adjusted version perform much worse. Relative to the other adjusted estimators, the bias-corrected version seems to suffer more heavily from the adjustment in the estimator.

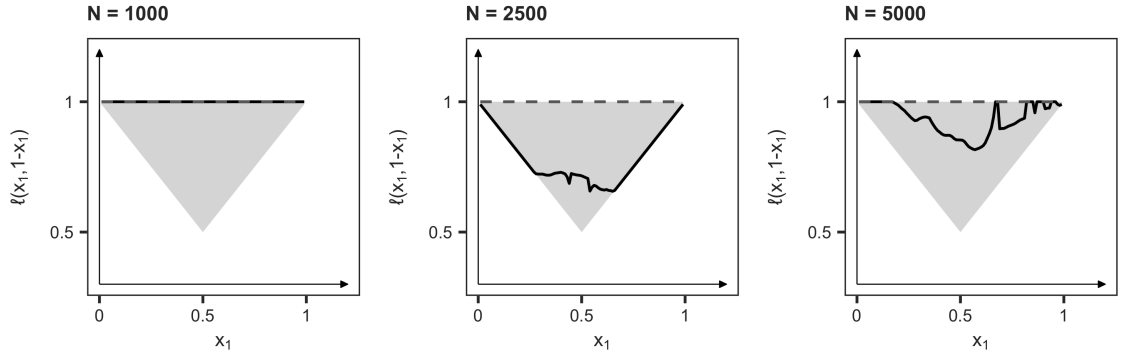

Figure 5.2.3: Estimated adjusted bias-corrected empirical STDF $\overline{\ell}^{BC3}$-function based on simulation results for the bivariate Normal copula with model parameter $\rho = 0.5$ using $k = 5\%$ of the $N \in \{1000, 2500, 5000\}$ observations.

Proposition 5.7. *Under the null hypothesis of tail independence and with the conditions of Theorem 5.2, satisfied for two intermediate sequences* k *and* $\overline{k}$ *such that* $k = \mathcal{O}(\overline{k})$, *we have*

$$\sqrt{k}\left(\overline{\ell}^{BC3}_{n,k,\overline{k}}(\mathbf{x}) - \ell(\mathbf{x})\right) \to \left(\int_0^1 K(u) u^{-1/2} \mathrm{d}u\right) \overline{B}(\mathbf{x}) \tag{5.2.12}$$

in $D([0,T]^d)$ *for every* $T > 0$.

Proof. We reformulate the proof of Beirlant et al. (2016) regarding the asymptotic behavior of the bias-corrected estimator ℓ using the adapted empirical STDF in Appendix B.4.3. Since the proof only relies on the Gaussianity of the process B_ℓ, and not on its specific form, the reformulation follows straightforwardly by replacing B_ℓ with $\overline{B}$ in the original proof. ■

Again, by taking the power kernel $K(t) = (\tau+1)t^\tau \mathbb{1}\{t \in (0,1)\}, \tau > -1/2$, we get

$$\sqrt{k}\left(\overline{\ell}^{BC3}_{n,k,\overline{k}}(\mathbf{x}) - \ell(\mathbf{x})\right) \to 2\frac{1+\tau}{1+2\tau}\overline{B}(\mathbf{x})$$

which is convenient because the asymptotic distribution then becomes a rescaled version of the asymptotic distribution of the non-bias-corrected estimator $\overline{\ell}$. The test statistics and their asymptotic distributions can now be defined similarly as before. The integral-based test statistic is given by,

$$\begin{aligned} I_n^{BC3} &= \int_{[0,1]^d} k\left(\overline{\ell}^{BC3}_{n,k,\overline{k}}(x_1,\ldots,x_d) - (x_1+\cdots+x_d)\right)^2 \mathrm{d}x_1\ldots\mathrm{d}x_d \\ &\to \left(2\frac{1+\tau}{1+2\tau}\right)^2 \int_{[0,1]^d} \overline{B}(\mathbf{x})^2 \mathrm{d}x_1\ldots\mathrm{d}x_d. \end{aligned}$$

The supremum-based test statistic is now equal to the following,

$$\begin{aligned} S_n^{BC3} &= \sup_{\mathbf{x}\in[0,1]^d} \sqrt{k}\left|\overline{\ell}^{BC3}_{n,k,\overline{k}}(x_1,\ldots,x_d) - (x_1+\cdots+x_d)\right| \\ &\to 2\frac{1+\tau}{1+2\tau} \sup_{\mathbf{x}\in[0,1]^d} \left|\overline{B}(\mathbf{x})\right|. \end{aligned}$$

Finally, the multivariate TDC-based test statistic is defined as,

$$\begin{aligned} T_n^{BC3} &= \sqrt{k}\left(\overline{\ell}^{BC3}_{n,k,\overline{k}}(1,\ldots,1) - d\right) \\ &\to 2\frac{1+\tau}{1+2\tau}\overline{B}(1,\ldots,1) \sim N\left(0, \left(2\frac{1+\tau}{1+2\tau}\right)^2 d(1+\theta)\right). \end{aligned}$$

Note that the quantiles in Table 5.2.2 just have to be multiplied with a factor dependent on τ to retrieve the new quantiles of the asymptotic distribution of the test statistics.

5.2.2 Multiple pairwise testing problem

Another way to extend the bivariate test for tail independence of Hüsler and Li (2009) to address the multivariate hypothesis testing problem is to conduct multiple bivariate tests. Recall that based on Theorem 3.3, it is sufficient to establish tail independence for all pairs in order to conclude joint tail independence. Therefore, the null hypothesis of multivariate tail independence can be investigated by testing the null hypothesis of tail independence for all bivariate pairs. That is, we propose to replace the original multivariate testing problem,

$$H_0 : (X_1,\ldots,X_d) \sim C_F : C_F \in MDA(\Pi) \qquad H_1 : (X_1,\ldots,X_d) \sim C_F : C_F \notin MDA(\Pi),$$

with the following multiple hypothesis testing problem,

$$\begin{aligned} H_0 &: \forall i,j \in \{1,\ldots,d\}, i \neq j : (X_i,X_j) \sim C_F : C_F \in MDA(\Pi) \\ H_1 &: \exists i,j \in \{1,\ldots,d\}, i \neq j : (X_i,X_j) \sim C_F : C_F \notin MDA(\Pi). \end{aligned}$$

As can be seen, this introduces a multiple testing problem where multiple hypotheses have to be evaluated simultaneously. When testing multiple hypotheses, the probability of rejecting the null hypothesis increases with every hypothesis test that is conducted. Consider, for example, the case where 2 independent hypotheses are to be tested together. If each test has a p-value of 0.1, the multiple hypothesis test has a p-value of 0.01. Usually, the relatively large p-values of the individual hypotheses would not be considered significant, while the relatively small p-value of the combined test would be considered significant. Moreover, the combined p-value decreases for every hypothesis test that is added to the multiple hypothesis testing problem, making it more likely to reject the combined null hypothesis with every individual hypothesis that is added to the problem. For our specific testing problem, this would imply that the probability of incorrectly rejecting tail independence increases with the dimension of the testing problem. Hence, for high-dimensional testing problems, tail dependence will almost always be concluded. This can be a costly mistake to make.

It is noteworthy that in literature, multivariate problems are almost always approached with pairwise tests. For example, Kiriliouk et al. (2014) explore the asymptotic dependence between returns of three large banks by considering pairwise tests. Due to the lack of multivariate testing procedures, this approach makes sense. However, to our knowledge, nobody acknowledges the multiple testing issue nor considers an adjustment of the p-values to prevent data from being classified as tail dependent too often.

In order to prevent the type I error to increase because of multiple hypotheses, the p-values (or the significance levels) have to be corrected for the number of hypotheses that are to be evaluated together. A common and straightforward method to adjust the p-values for a multiple testing problem is the Bonferroni method. This method sets the p-value for each test equal to np if there are n hypotheses to be tested. The Bonferroni procedure is known to be quite conservative and causes loss of power in the testing procedure. That is, the type II error rate increases, making it more likely to incorrectly assume tail independence.

A more powerful sequential version of the Bonferroni method has been proposed by Holm (1979). In Holm's sequential version, all pairwise tests need to be performed first in order to obtain their p-values. The tests are then ordered from the one with the smallest p-value to the one with the largest p-value. Denote these p-values by $p_{1:n} \leq \cdots \leq p_{n:n}$. The test with the lowest probability is tested first with a Bonferroni correction involving all tests. That is, the adjusted p-value is set equal to $p_1^{adj.} = np_{1:n}$. The second test is tested with a Bonferroni correction involving one less test, after which the adjust p-value is set equal to the maximum of the previously adjusted p-values and the newly adjusted p-value. That is, $p_2^{adj.} = \min\left\{(n-1)p_{2:n}, p_1^{adj.}\right\}$. This procedure is repeated until all original p-values have been corrected. Holm's approach is more powerful than the original Bonferroni approach but still keeps under control the increasing Type 1 error (Abdi, 2010). The Hochberg method is a reverse sequential version of the Bonferroni corrections. That is, the adjustments begin with the largest p-value and sets subsequent p-values equal to minimum of the appropriate Bonferroni correction for the step and the previously adjusted p-values. The approaches are illustrated with an example in Table 5.2.3. Although the Bonferroni method and to a lesser extent the Holm method are most frequently used, the Hochberg method tends to outperform the other two correction methods according to Blakesley et al. (2009).

Table 5.2.3: Example of methods to correct p-values in multiple testing problems.

	Uncorrected		**Bonferroni**		**Holm**		**Hochberg**	
	p-value	Reject	p-value	Reject	p-value	Reject	p-value	Reject
Test 1	0.01	TRUE	0.03	TRUE	0.03	TRUE	0.03	TRUE
Test 2	0.02	TRUE	0.06	FALSE	0.04	TRUE	0.03	TRUE
Test 3	0.03	TRUE	0.09	FALSE	0.04	TRUE	0.03	TRUE
Combined		TRUE		FALSE		TRUE		TRUE

The corrections for the p-values can be implemented in `R` with the `p.adjust` function from the `stats` package. An advantage of the pairwise multiple testing problem is that if the null hypothesis of tail independence is rejected, the testing results provide insights into which pairs contribute to the tail dependence structure and which are tail independent.

5.2.3 Simulation study

The performance of the different procedures to test for tail independence is assessed with a simulation study in this section. The aim is to analyze the behavior of the testing procedures for different dependence types, dependence strengths, and dimensions. In our opinion, this section presents a potentially important contribution to literature, since a thorough simulation study on the performance of tests for tail independence is to our knowledge lacking in the current literature. Moreover, the introduced testing procedures are evaluated in this section.

We start with the behavior of the STDF-based and TDC-based tests for the bivariate case and next compare the behavior of the multiple testing approach with the multivariate test approach for higher dimensions. The multivariate integrals inherent to the STDF-based test statistic I_n are determined with the `adaptIntegrate` function from the `cubature` package in `R`. The suprema inherent to the STDF-based test statistic S_n are determined with the `bobyqa` function from the `minqa` package for optimization problems. Before the simulation results are presented, simulation parameters and running times are briefly discussed.

Simulation parameters

Data are simulated from the Gumbel, t-, Normal, and Frank copula. First we consider the distributions in dimensions $d \in \{2, 3, 5\}$ with model parameters specified such that all pairwise dependencies equal $\rho = 0.2$ for weak dependence, $\rho = 0.5$ for medium dependence, and $\rho = 0.8$ for strong dependence. However, since the performance of the testing procedures might be different when the strength of pairwise dependencies vary, we also consider distributions with asymmetric pairwise dependencies. This is achieved by specifying different correlation parameters for the Normal and t-copula and with several D-vine copula models (see Figure 5.2.4 for the specifications of the trivariate D-vines).

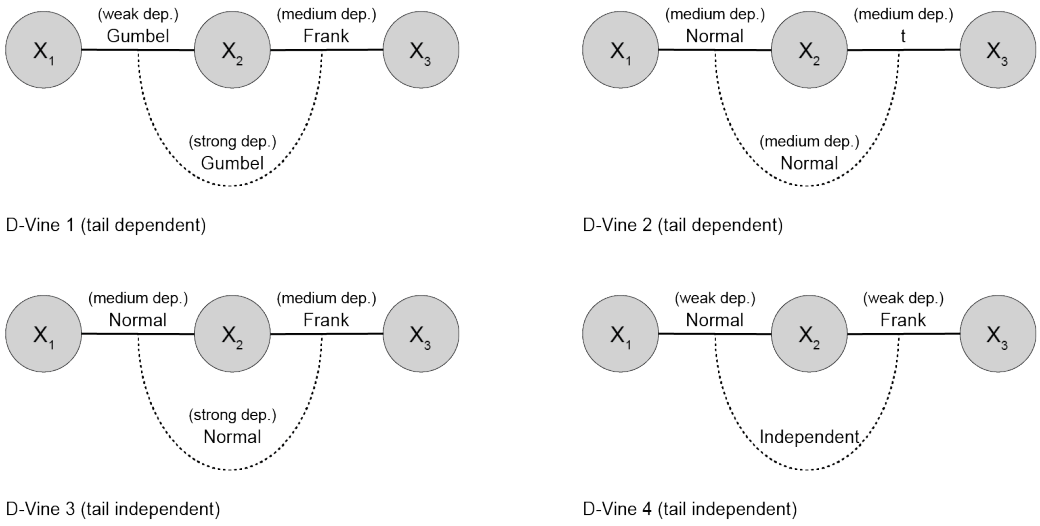

Figure 5.2.4: D-vine specifications for the testing simulations.

For each testing problem that is to be evaluated, we simulate $B = 500$ samples to run the test and report the rejection rate for significance levels of $\alpha = 5\%$ and $\alpha = 1\%$. Note that the number of samples considered here is significantly higher than in Hüsler and Li (2009) who only use 6 samples to evaluate the behavior of their test for the bivariate case. Since the estimator functions performed superior for a threshold value of $k = 1\%$ for most distributions, this threshold value is also used for the testing functions. In accordance with Hüsler and Li (2009), we split samples exactly in half to determine the test statistics ($\theta = 1$). Furthermore, for the tests based on the kernel-smoothed and bias-corrected estimators it is necessary to specify additional parameters. Following Fougères et al. (2015), Beirlant et al. (2016) and Kiriliouk et al. (2018) we employ the power kernel with parameter $\tau = 5$ and set $a = r = 0.4$ for the bias-corrected estimators.

Running times

Before the simulations are conducted, the running times are assessed. Table 5.2.4 shows running times for the different test statistics in dimensions $d \in \{2, 3, 5\}$. It is clear that the TDC-based test is significantly less computationally intensive than the other two test statistics that are based on the entire stable tail dependence function. Especially the integral-based test shows very long running times, making it unattractive to use this particular test statistic for higher dimensions. The supremum-based test can be evaluated faster than the integral-based test, but for the adjusted estimators, i.e., the kernel smoothed estimator and the bias-corrected estimator, and for higher

dimensions, this test statistic also shows quite long running times. The TDC-based tests for the bias-corrected estimator are also quite computationally expensive. However, this probably has to do with our implementation of the bias-corrected estimator. As mentioned in Section 5.1.5, the computation of the bias-corrected estimates involve quite a lot computations and without deferring the computations to `C++` it will take a long time to evaluate this estimator. Based on these running times, we decide to only assess the S_n and T_n test statistics for the tests based on the adjusted kernel estimator and only assess the T_n test statistic for the bias-corrected test.

Table 5.2.4: Running times to compute test statistics for one sample (in seconds) and hypothetically for $B = 500$ samples (in hours, between brackets).

	NP			NP-Kernel			NP-BC		
	I_n	S_n	T_n	I_n	S_n	T_n	I_n	S_n	T_n
$d = 2$	0.194 (0h)	0.22 (0h)	0.004 (0h)	11.626 (1.6h)	4.038 (0.6h)	0.079 (0h)	748.607 (104h)	478.245 (66.4h)	8.504 (1.2h)
$d = 3$	4.207 (0.6h)	0.108 (0h)	0.001 (0h)	17 (2.4h)	7.142 (1h)	0.112 (0h)	> 3600 (>500h)	> 3600 (>500h)	12.956 (1.8h)
$d = 5$	6.543 (0.9h)	0.106 (0h)	0.004 (0h)	382.666 (53.1h)	14.345 (2h)	0.198 (0h)	> 3600 (>500h)	> 3600 (>500h)	20.645 (2.9h)

The running times are retrieved from a Macbook Pro with 2,7 GHz Intel Core i5 processor using the statistical software program `R`. If computations of a test statistic for one sample took more than one hour, computations were prematurely terminated. All implementations of the test statistics are our own (see Appendix E.2).

Simulation results

All simulation results are reported in Appendix C.4. The main findings are discussed here. For the bivariate data, the results show that all test statistics tend to lead to rejection rates that are close to or lower than the significance level if the null hypothesis is true. The only notable exception is provided by TDC-based test for data simulated from the Normal copula with strong dependence. Here, the null hypothesis of tail independence is rejected 17% of the times based on a significance level of $\alpha = 5\%$. Overall, the TDC-based test statistic T_n has a higher rejection rate than the STDF-based test statistics I_n and S_n. The test statistics based on the kernel-smoothed estimator result in similar but slightly lower rejection rates. The test statistics based on the bias-corrected estimator lead to a zero-rejection rate. Overall, all test statistics behave very well in the bivariate case if the null hypothesis is true. See Table 5.2.5 for a quick overview.

Table 5.2.5: Rejection rates of the test statistics with a significance level of $\alpha = 5\%$ for bivariate data simulated from copulas with medium dependence.

		Tail dependent data		Tail independent data	
d = 2		Gumbel	t	Normal	Frank
Empirical STDF-based (NP)	I_n	0.076	0.044	0.046	0.022
	S_n	0.07	0.024	0.032	0.028
	T_n	0.19	0.1	0.068	0.03

Results are based on $B = 500$ samples with same size $N = 2500$. The sample is split into two equally sized subsamples and $k = 1\%$ of the remaining 1250 observations are used as tail observations.

However, if the null hypothesis is false, the rejection rates are still quite low. For example, for the Gumbel copula with strong dependence, the null hypothesis of tail independence is only rejected in 11% of the samples given a significance level of $\alpha = 5\%$ when using the supremum-based standard bivariate test of Hüsler and Li (2009). The rejection rates are even lower for tail dependent data with medium or light dependence. This indicates that the tests are not very powerful. Since the TDC-based test statistic T_n rejects more often than the STDF-based tests, this one is to be preferred. However, also for the TDC-based tests the power of the test remains quite low.

The kernel-smoothed tests and the bias-corrected test lead to lower rejection rates and therefore do not improve results in this scenario. See Table 5.2.5 for a quick overview of rejection rates for the bivariate copulas with medium dependence.

For the multivariate results ($d > 2$), both the multiple pairwise testing approach and the multivariate testing approach have been evaluated. For the trivariate case, the multiple pairwise testing approach has a much lower rejection rate than the multivariate testing approach in all instances. Whereas this leads to rejection rates close to the significance level for the tail independent simulations, the rejection rates are very low for the tail dependent simulations. A quick overview of rejection rates for the trivariate copulas with medium dependence is presented in Table 5.2.6. Similar to the bivariate results, the tests based on the adjusted kernel-smoothed estimator tend to reject the null hypothesis slightly less than the tests based on the standard adjusted empirical STDF, both for the STDF-based test and the TDC-based test, and both when the null hypothesis is true and when the null hypothesis is false. The bias-corrected 3-dimensional multivariate tests yield improved rejection rates relative to the uncorrected tests in some instances. For the tail independent data, the bias-corrected tests lead to lower rejection rates almost everywhere. The exception is presented by the trivariate Frank copula with strong dependence, where the rejection rate increases from 6.8% for the standard test to 14% for the bias-corrected test. However, for the tail dependent data, the bias-corrected tests also tend to lead to lower rejection rates, except for the Gumbel and t-copula with strong dependence. For the copulas with mixed pairwise tail dependencies, similar patterns are visible.

Table 5.2.6: Rejection rates of the test statistics with a significance level of $\alpha = 5\%$ for 3-dimensional data simulated from copulas with medium dependence.

			Tail dependent data		Tail independent data	
d = 3			Gumbel	t	Normal	Frank
Multiple tests	NP	T_n	0.098	0.038	0.02	0.006
	NP-Kernel	T_n	0.084	0.022	0.024	0.01
Multivariate test	NP	T_n	0.39	0.2	0.12	0.034
	NP-Kernel	T_n	0.32	0.17	0.086	0.034
	NP-BC	T_n	0.09	0.014	0.002	0

Results are based on $B = 500$ samples with same size $N = 2500$. The sample is split into two equally sized subsamples and $k = 1\%$ of the remaining 1250 observations are used as tail observations.

For higher dimensions ($d = 5$), the difference between the multiple testing and the multivariate testing approach becomes even more pronounced. The multiple testing approach is much less powerful compared to the multivariate tests. Even for data simulated from the Gumbel copula with strong dependence, the multiple pairwise test only rejects the null hypothesis of tail independence for 16% of the samples given a significance level of $\alpha = 5\%$ for the TDC-based test, which is the most powerful alternative. Again, the tests based on the adjusted kernel-smoothed estimator tend to reject the null hypothesis slightly less than the tests based on the standard adjusted empirical STDF. For example, for the simulations from the 5-dimensional Normal copula with equal medium bivariate dependencies, the kernel-based tests reject the null hypothesis of tail independence for 18% of the samples, versus 28% for the standard test. However, for the tail dependent simulations the rejection rate is also slightly lower, making this test less powerful. The multivariate TDC-based test derived from the bias-corrected estimator tends to yield improved rejection rates for the Normal copula, but not so much for the other copulas. It is questionable whether these modest improvements outweigh the additional complexity and computation costs associated with the bias-corrections. Table 5.2.7 shows an overview of rejection rates for the 5-dimensional copulas with medium dependence.

Table 5.2.7: Rejection rates of the test statistics with a significance level of $\alpha = 5\%$ for 5-dimensional data simulated from copulas with medium dependence.

			Tail dependent data		Tail independent data	
d = 5			Gumbel	t	Normal	Frank
Multiple tests	NP	T_n	0.044	0.018	0.004	0.004
	NP-Kernel	T_n	0.036	0.008	0.006	0.006
Multivariate test	NP	T_n	0.85	0.62	0.28	0.082
	NP-Kernel	T_n	0.77	0.52	0.18	0.068
	NP-BC	T_n	0.74	0.29	0.27	0.83

Results are based on $B = 500$ samples with same size $N = 2500$. The sample is split into two equally sized subsamples and $k = 1\%$ of the remaining 1250 observations are used as tail observations.

Overall, we conclude that the TDC-based test is a very viable alternative to the STDF-based tests. It is more powerful when the null hypothesis is false and yields rejection rates close to the significance level when the null hypothesis is true in most instances. Moreover, the TDC-based test statistic is much easier and faster to evaluate for a given dataset, especially if dimensions increase, and critical values stem from a known Normal distribution and do not have to be simulated.

Finally, we reran some of the simulations with a sample size of $N = 5000$ to investigate whether a larger sample size might lead to more powerful results of the tests. The results are presented in Table 5.2.8 below. An improvement in the rejection rate can be observed in all instances of the tail dependent data. Although the multiple testing approach is more powerful for the larger sample size, rejection rates are still significantly lower than those of the multivariate test. For the tail independent data, rejection rates are comparable or slightly higher than for the smaller sample size. The null hypothesis of tail independence is rejected too often for data simulated from the Normal copula. Although the bias-corrected testing procedure was not evaluated for this larger sample size due to computational limitations, bias-corrections might alleviate this problem.

Table 5.2.8: Rejection rates of the test statistics with a significance level of $\alpha = 5\%$ for data simulated from copulas with medium dependence in dimension $d = 3, 5$ for $N = 5000$.

			Tail dependent data		Tail independent data	
d = 3			Gumbel	t	Normal	Frank
Multiple tests	NP	T_n	0.228	0.092	0.042	0.03
	NP-Kernel	T_n	0.212	0.07	0.04	0.034
Multivariate test	NP	T_n	0.614	0.334	0.136	0.058
	NP-Kernel	T_n	0.566	0.278	0.11	0.05
d = 5						
Multiple tests	NP	T_n	0.156	0.06	0.046	0.022
	NP-Kernel	T_n	0.112	0.036	0.032	0.012
Multivariate test	NP	T_n	0.984	0.876	0.484	0.108
	NP-Kernel	T_n	0.974	0.796	0.408	0.084

Results are based on $B = 500$ samples with same size $N = 5000$. The sample is split into two equally sized subsamples and $k = 1\%$ of the remaining 2500 observations are used as tail observations.

5.3 Summary

This chapter studied statistical inference procedures to characterize the multivariate tail dependence structure. First, the focus was on estimating the stable tail dependence function (STDF) and, inherently, the multivariate STDF-based TDC Λ. Secondly, methods for testing tail independence were researched. These inference methods provide a comprehensive set of tools to characterize the tail dependence structure fully, to assess the strength of the tail dependence, or to classify data as either tail dependent or tail independent.

The STDF can be estimated nonparametrically and parametrically, but since parametric models lack the flexibility that is needed to accurately model tail dependence in higher dimensions, the focus was on the empirical STDF and adjusted versions thereof. Specifically, we considered smoothed versions of the empirical STDF based on the beta copula and based on a kernel-smoother. Furthermore, three bias-correction methods were discussed to mitigate the significant bias terms associated with uncorrected STDF estimators for data coming from copulas that converge slowly to their extreme value copula. The simulation study showed that both smoothed versions of the empirical STDF tend to yield slightly superior results compared to the standard empirical STDF in terms of (integrated) MSE. However, the main advantage of the smoothed estimators is presented by the more attractive visualizations, and not in the finite sample performance. The bias-corrected estimators do offer a vast improvement in the finite sample performance of the estimators when the convergence rate is low. Especially the third bias-corrected version of Beirlant et al. (2016) performs well. When using this bias-corrected estimator, the multivariate TDC Λ can be estimated with at least one decimal place accuracy in all considered dimensions $d = 2, 3, 5$. In contrast, the bias of the uncorrected estimator is greater than 0.2 for the 5-dimensional Normal copula with medium dependence.

Since the STDF is able to identify tail independence, it is natural to consider a testing procedure based on this dependence function. However, under the null hypothesis of tail independence, the asymptotic distribution of the empirical STDF becomes degenerate, making it impossible to construct a test statistic based on the distribution. Following the idea of Hüsler and Li (2009), we adjust estimators by using separate datasets for estimating high quantiles and for counting exceedances. The asymptotic distribution of this adjusted empirical STDF is no longer degenerate, and we derived its specific form in the d-dimensional case from Theorem 5.3. In addition, it is shown that the kernel-smoothed and bias-corrected versions of the empirical STDF can be adjusted similarly and asymptotic results for these estimators are also derived. Since the multivariate TDC Λ is also able to identify tail independence (Main Theorem 2), we propose a new testing procedure that only requires the evaluation of the STDF in the point $(1, \ldots, 1)$, simplifying estimation and testing greatly. An extensive simulation study showed that this TDC-based test tends to be more powerful than the STDF-based tests and yields rejection rates close to the significance level when the null hypothesis is true in most instances. Tests based on the kernel-smoothed or bias-corrected estimators do not offer significant improvements in the rejection rates that outweigh the additional computation costs associated with these adjusted testing procedures. Besides the multivariate testing approach, a multiple pairwise testing approach is considered, where p-values were adjusted to correct for the multiple testing problem. However, the simulation study showed that the multiple testing approach is significantly less powerful than the multivariate tests. Larger sample sizes might alleviate this problem to some extent, but when the data collection does not permit this, the multivariate TDC-based testing procedure is to be preferred.

Chapter 6

Tail dependence in FX markets

In this chapter, previously discussed methods to characterize multivariate tail dependence structures are applied to assess extremal dependencies in FX (foreign exchange) markets. The aim of this chapter is to illustrate how the introduced theory and statistical methods can be applied to real data. The FX market serves as an interesting example to use for this purpose since it is by far the largest and most liquid market in terms of trading volume and tends to exhibit extreme jumps in reaction to political and financial crises. Moreover, the trading activities of buying and selling currencies against an uncertain exchange rate expose many international companies and investors to substantial financial risks.

The exchange rate indicates how much units of a particular domestic currency are required to buy one unit of a particular foreign currency and is quoted as foreign-domestic (FORDOM, FOR-DOM or FOR/DOM) (Wystup, 2006). That is, an EURUSD (Euro-US Dollar) rate of 1.18 indicates that 1 Euro is worth 1.18 US Dollars. The term domestic is unrelated to the location of the trader or any country. It merely means the numeraire currency (unit of account). An increase (decrease) in an FX rate corresponds to an increasing (decreasing) value of the foreign currency relative to the domestic currency. Hence, if the EURUSD rate increases, the Euro becomes more valuable relative to the US Dollar, and if the EURUSD rate decreases, the US Dollar becomes more valuable relative to the Euro. Changes in demand or supply of either the domestic currency, the foreign currency, or both, lead to movements in the spot rate (i.e., the current exchange rate). Such changes in demand can be driven by interest rates, economic developments, or politics, for example. See, e.g., Della Corte et al. (2016) for research on the drivers of FX rates.

As a corporation operating in the global economy, it is often inevitable to deal with foreign currencies. For example, a company based in Europe with a client base in the UK will have revenues in British Pounds which will have to be converted to Euros with the EURGBP conversion rate. The uncertainty in this conversion rate, or foreign exchange (FX) rate, is inherited by the EUR-denoted profits and costs of the company. This uncertainty in the profits and losses presents the company with financial risks. Imagine for example that the EURGBP rate increases, indicating that the British Pound becomes less valuable relative to the Euro due to some political crisis (e.g., the Brexit referendum). Profits made in the UK will now be less valuable in Euros than with the lower FX rate. This represents a loss for the company. Movements in the FX rates can, therefore, be considered as risk factors for a multinational company, and for any investor exposed to one or more foreign currencies.

The illustrative application of extreme value theory to foreign exchange rates has been popular in literature, but only for bivariate dependencies. See, for example, Resnick (2004) for an application of methods to estimate the spectral measure for tail dependence between the German Mark and the French Franc (before the introduction of the Euro). In this chapter, however, the focus is on illustrating methods to assess *multivariate* tail dependence. We first assess the extremal dependence structure between three European currency pairs, being the EURGBP (Euro-British Pound), EURSEK (Euro-Swedish Krona), and EURNOK (Euro-Norwegian Krone). Next, a higher dimensional problem is considered by looking at the tail dependence structure of 5 worldwide currency pairs: EURUSD (Euro-US Dollar), GBPUSD (British Pound-US Dollar), USDCAD (US Dollar-Canadian Dollar), USDTRY (US Dollar-Turkish Lira), and USDRUB (US Dollar-Russian Ruble).

6.1 Preliminary data analysis

The required FX rates are retrieved from Bloomberg for a 10-year timeframe ranging from January 2007 to January 2017. This leads to 2632 daily observations of the FX rates. Note that the size of the dataset is very similar to the sample size of $N = 2500$ employed for the simulation studies in the previous chapter. Appendix D.1 presents an overview of the time series of all considered FX rates. Since the spot rates are not stationary (see Appendix D.2 for test results), log returns are used to model the spot behavior: if $s_1, \ldots, s_T$ is a series of daily spot FX rates, the daily log returns are

$$r_t = \log\left(\frac{s_t}{s_{t-1}}\right), \quad t = 2, \ldots, T.$$

Log returns are often used because of modeling advantages and they are approximately the same as regular returns if returns are small. Moreover, in the context of FX rates, log returns provide the attractive property that changing the side of the exposure (i.e., buyer/seller of the foreign currency) corresponds to changing the sign of the log return. The analysis can be based on daily returns, as is commonly done in literature, but since this is not necessarily the time horizon used by corporations for accounting or risk management purposes, some results are also considered for alternative return horizons, taking either 1-month, 3-month, 6-month, or annual log returns. Appendix D.2 shows that the ADF (Augmented Dickey Fuller) test indicates that the log return series are mostly stationary, with some exceptions to be found for the 6-month and 12-month return horizons. Figure 6.1.1 shows an example of the original time series, the daily log returns and the monthly log returns for the EURGBP rate. Whereas the raw timeseries clearly exhibits trends, the log return series look stationary, especially the daily log returns.

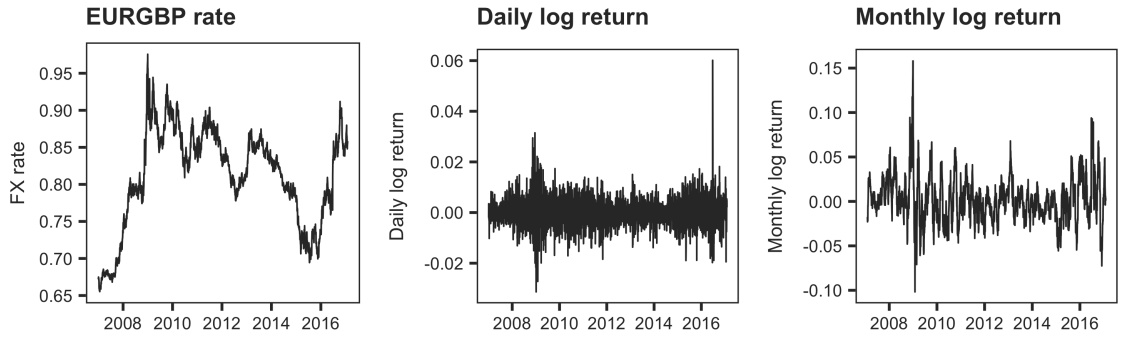

Figure 6.1.1: Time series of the EURGBP FX rate and corresponding daily and monthly log returns.

Usually, the log returns of financial series exhibit temporal dependence, which violates the iid assumption that underpins most models used to characterize tail dependence. Indeed, the Ljung-Box test indicates that most return series exhibit temporal dependence (see Appendix D.2 for the testing results). A non-straightforward solution would be to adjust the considered theory on multivariate tail dependence structures such that it can be applied to data that are time dependent instead of iid. Einmahl et al. (2016a) and de Haan et al. (2016) conducted research in this direction for univariate extreme value theory but it is beyond the scope of this data example to extend their efforts to higher dimensions. A solution that can be employed more readily is to adjust data such that they do satisfy the iid assumption. This can be achieved by applying a filter to the log return series. In literature, this is a common approach to apply extreme value theory, copula theory, or in general, models that rely on iid data, to time series that exhibit serial dependence. For example, McNeil and Frey (2000) employed this method to model univariate tails, Wang et al. (2010) apply a GARCH model to filter daily log returns of FX rates and to assess the marginal extremal behavior, and Czado et al. (2012) apply an ARMA(1,1)-GARCH(1,1) filter before fitting copulas to USD-exchange rates.

Hence, in order to account for time dependence between the marginal FX returns, an ARMA model is used for the conditional mean and a GARCH (Bollerslev, 1986) or GJR-GARCH (Glosten et al., 1993) model for the conditional variance. The GJR-GARCH model can be preferred to the standard GARCH model because of the leverage term that allows for asymmetry in the reaction to

positive and negative shocks. The ARMA(2,0)-GJR-GARCH(1,1) model yields satisfactory results for the FX returns and is therefore employed in this analysis. Formally, this model is given by

$$r_{j,t+1} = a_{j,0} + a_{j,1} r_{j,t} + a_{j,2} r_{j,t-1} + \sigma_{j,t+1} \varepsilon_{j,t+1},$$
$$\sigma^2_{j,t+1} = \omega_j + (\alpha_j + \gamma_j \mathbb{1}\{\varepsilon_{j,t} < 0\}) \varepsilon^2_{j,t} + \beta_j \sigma^2_{j,t},$$

where $r_{j,t}$ is the log return of the j-th FX rate at time t and $\sigma_{j,t}$ the volatility of the j-th FX rate at time t. To determine the model fit, residuals are assumed to follow a t-distribution. This is a common assumption when applying the ARMA-GARCH filter to financial returns. See Appendix Appendix D.2 for goodness-of-fit testing results that confirm that residuals of the filtered FX return series are approximately following a t-distribution. After applying an ARMA-GJR-GARCH filter to the time series data, standardized residuals can be retrieved as follows

$$\tilde{\varepsilon}_{jt} = \frac{r_{j,t} - \mathbb{E}\left(r_{j,t} | \mathcal{F}_{t-1}\right)}{\sqrt{\mathrm{Var}\left(r_{j,t} | \mathcal{F}_{t-1}\right)}},$$

where $\mathbb{E}\left(r_{jt} | \mathcal{F}_{t-1}\right)$ is the conditional expectation of $r_{j,t}$ given the information available one timestep before (denoted by the sigma algebra $\mathcal{F}_{t-1}$) and $\mathrm{Var}\left(r_{j,t} | \mathcal{F}_{t-1}\right)$ the conditional variance of the log return at time t given the information available at time $t-1$. See Patton (2013) for an elaborate discussion on this filtering approach. To illustrate the effect of the filter, Figure 6.1.2 shows the autocorrelation pattern of daily and monthly log returns of the EURGBP rate before and after applying the ARMA-GJR-GARCH filter. Whereas the unfiltered log returns exhibit serial dependence, especially for monthly returns, the autocorrelations of the filtered log returns suggest that the filtered series are time independent. See Appendix D.1 for an overview of the results of the filtering process for the timeseries of all considered FX rates. After applying the filter, standardized residuals are tested for serial dependence. The results show that the filter effectively removes temporal dependence (Appendix D.2).

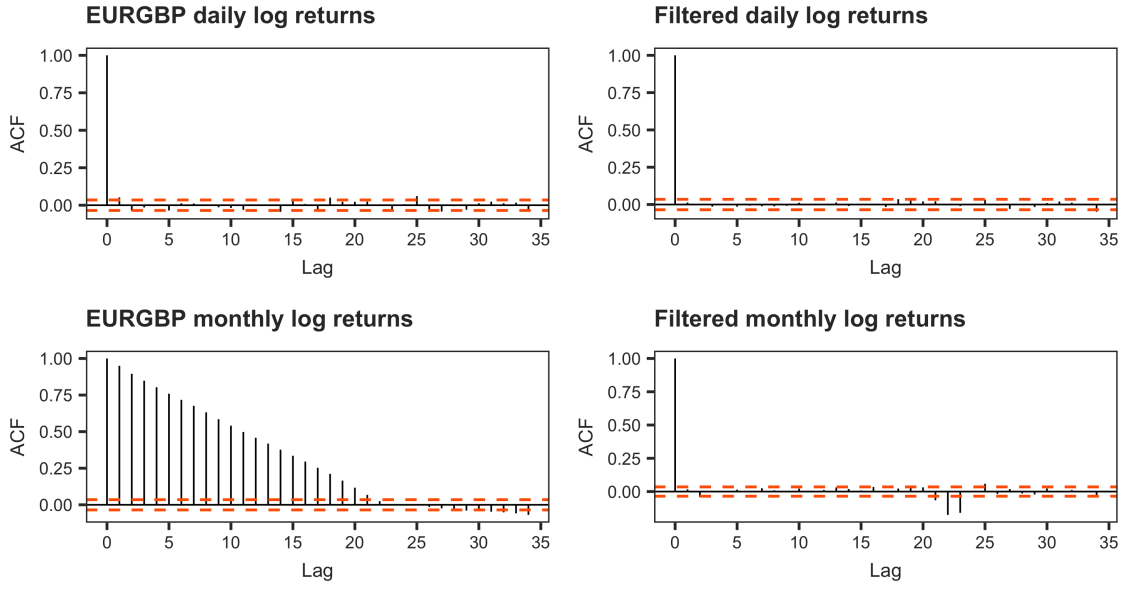

Figure 6.1.2: Autocorrelation for different lags for daily and monthly log returns of the EURGBP rate before and after applying the ARMA(2,0)-GJR-GARCH(1,1) filter.

The standardized residuals of the ARMA-GJR-GARCH filtered log return series are ultimately used to assess the extremal dependence structure between several FX rates in the next sections. It is important to note that since copulas are invariant under monotone transformations, the marginal transformations applied to the original return series do not change the dependence structure between the FX rates. That is, the copula of the standardized residuals is the same as the copula of the log return series. However, since the residuals do not contain temporal dependence (in theory), it is easier to distill the true dependence structure of the data. Figure 6.1.3 illustrates the importance of the filtering procedures. The left-hand panels contain scatterplots for the raw EURUSD and EURGBP rates on their original scale (top) and on the copula scale (bottom). Due to serial

dependence and nonstationarity of the time series, the dependence structure is heavily distorted and does not resemble any known parametric copula model. The rank correlation is estimated to be $\rho = 0.25$, and the bivariate TDC is estimated as $\lambda = 0$, indicating tail independence. The log return series presented in the middle panels yield significantly higher dependence measures: the estimated rank correlation equals $\rho = 0.51$, and the estimated bivariate TDC equals $\lambda = 0.41$. By taking the log returns the time series are less heavily distorted. Finally, the ARMA-GJR-GARCH filtered monthly log returns are plotted in the right-hand panels. In theory, these series do not contain time dependence, and, consequently, the copula plot (bottom) should only reflect the dependence between the two FX rates. The dependence measures of the filtered log returns are slightly lower than for the unfiltered log returns: the rank correlation is estimated to be $\rho = 0.47$, and the estimated TDC equals $\lambda = 0.30$. Hence, in the presence of serial dependence, the (tail) dependence structure might be overestimated.

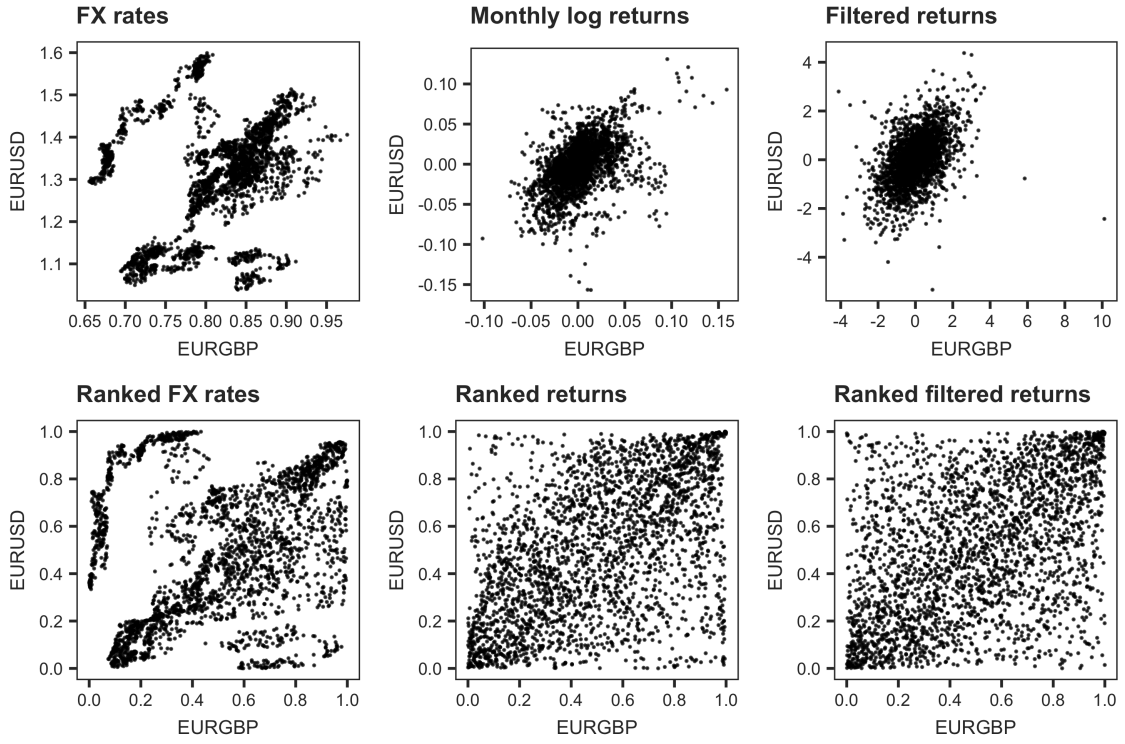

Figure 6.1.3: Scatterplots for the FX rates (left panels), monthly log returns (mid panels), and ARMA-GJR-GARCH filtered monthly log returns (right panels) of the EURGBP and EURUSD rates on the original scale (top) and the copula scale (bottom).

Finally, it should be noted that we implicitely assume that the tail dependence structure is constant over the considered timeframe. This is not necessarily a trivial assumption since evidence has been presented that the overall dependence structure between financial-asset returns is time-dependent. See, for example, Hartmann et al. (2006) and Peng and Lon (2012). One way to account for time-varying dependence is to impose a parametric model with an autoregressive structure for the parameters. See, e.g., Manner and Reznikova (2012), Almeida and Czado (2010), and Almeida et al. (2012) for research on such approaches. Alternatively, (Markov) regime switching models can be employed to account for structural breaks in the dependence structure. See, e.g., Chollete et al. (2009), Stöber and Czado (2014) and Fink et al. (2016) for applications of copula regime switching models. However, data constraints in estimating extreme value structures make it very challenging to test for different dependence structures. Camilo et al. (2014) present a parametric approach to model a time-dependent bivariate tail dependence structure. Since this is not the scope of this thesis, we simply assume that the dependence structure is time independent.

6.2 Tail dependence in European FX markets (3D)

This section is concerned with the trivariate tail dependence structure between the following three European FX rates: EURGBP (Euro-British Pound), EURSEK (Euro-Swedish Krona), and EURNOK (Euro-Norwegian Krone). The time series of these exchange rates are shown in Figure 6.2.1. Some noticeable trends can be observed from these timeseries. For example, the value of the Euro compared to the other three currencies increased significantly in 2008-2009. This can be explained by the fact that during crises, people tend to prefer larger, more stable currencies over relatively smaller currencies. As such, the Euro would be one of the go-to currencies when the economy takes a turn for the worse and can be considered a safe haven currency during these times (Ranaldo and Söderlind, 2010). As of 2012, the Swedish economy proved to recover stronger from the past economic crisis than the Norwegian economy, looking at indicators such as GDP growth. However, due to decreasing oil prices and fears for a housing crisis in both Norway and Sweden, both currencies lost value against the Euro in the past years. Lastly, the British Pound lost value against the Euro after the Brexit referendum in 2016. This may explain the quick rise of the EURGBP rate in 2016.

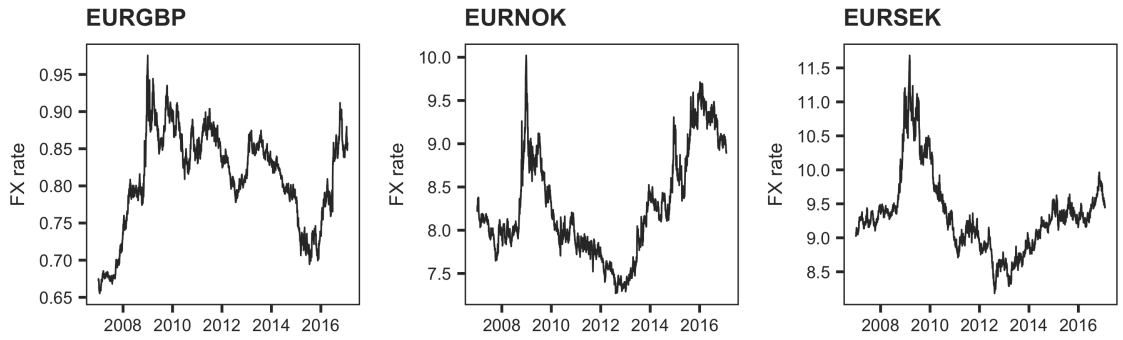

Figure 6.2.1: Time series of the considered EUR-exchange rates.

Following the methodology described in Section 6.1, the tail dependence structure between the EURGBP, EURNOK, and EURSEK rates is investigated based on standardized residuals of filtered marginal log returns. To get an initial overview of the data, pairwise dependence structures between these residuals are shown on the copula unit scale in Figure 6.2.2. Overall, there seems to be a positive dependence structure between the considered variables, which is also reflected by the positive pairwise rank correlations that are equal to 0.16, 0.09, and 0.47, for respectively the EURGBP-EURNOK, EURGBP-EURSEK, and the EURNOK-EURSEK combinations. Especially the EURNOK and EURSEK exchange rates exhibit quite a strong positive dependence structure. This can be explained by the fact that the two Scandinavian countries have a strong trading relationship and a large overlap in their import and export products (most notably: petroleum). The clustering of observations in the upper and lower orthants suggests that the EURSEK and EURNOK exchange rates might be tail dependent, whereas the dependence structures between the other two pairs look close to independent.

Note that it depends on the position of an FX exposure whether positive or negative returns correspond to losses. We first assume a short EUR position for all exposures. This corresponds to the scenario where a company makes profits denoted in another currency and then buys Euros to retrieve its EUR-denoted profits. In this case, a cheap Euro is favorable and positive returns on the EUR-exchange rate correspond to losses. Extreme losses are therefore represented by observations in the upper right orthant of the copula plot. In order to magnify these large values, the variables are further transformed to unit-Pareto margins (see Figure 6.2.3). Recall from Section 3.2.2 that these figures can be interpreted based on the spectral measure: if points are mainly concentrated on the x- and y-axis, as can be observed for the EURGBP-EURNOK and EURGBP-EURSEK plots, only one main variable contributes to large norms of vector-observations. This indicates tail independence: variables are not usually large together. Points that are scattered between the axes correspond to vector-observations for which both components are large, indicating tail dependence. This pattern can be seen in the EURSEK-EURNOK Pareto plot.

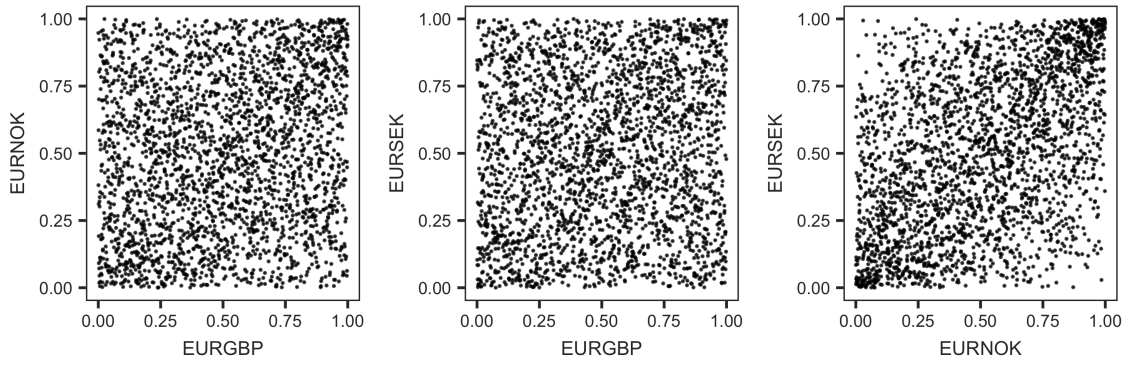

Figure 6.2.2: Pairwise copula plots for the standardized residuals of ARMA-GJR-GARCH filtered daily log returns of three EUR-exchange rates.

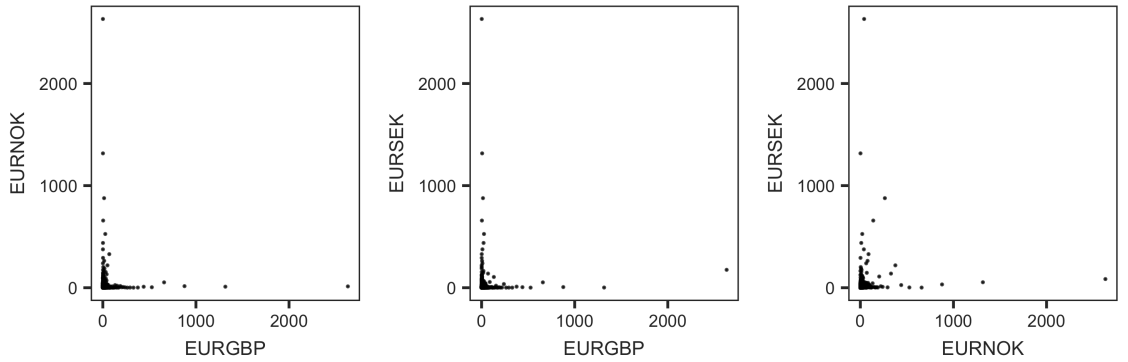

Figure 6.2.3: Pairwise Pareto plots for the residuals of ARMA-GJR-GARCH filtered daily log returns of three EUR-exchange rates assuming a short EUR position.

The tail dependence strength can be assessed with the multivariate TDC Λ (Section 4.2.3). Furthermore, test statistics developed in Section 5.2 can be evaluated to classify the tail dependence structure as either tail dependent or tail independent. Table 6.2.1 presents the estimated multivariate TDCs based on the empirical STDF and the bias-corrected version of Beirlant et al. (2016), combined with test statistic and p-values for the corresponding multivariate TDC-based tests. Since the time horizon of corporations is usually longer than one day, results are determined for a range of different time horizons.

Table 6.2.1: Trivariate TDCs and test statistics with p-values for different return horizons of EUR-exchange rates given a short EUR position.

		1-day	1-month	3-months	6-months	12-months
Empirical STDF (NP)	TDC (Λ)	0.14	0.14	0.12	0.14	0.19
	Test statistic (T_n)	1.81	-0.60	-4.60	-0.99	-3.06
	p-value	0.77	0.40	0.03	0.34	0.11
Bias-corrected (NP-BC3)	TDC (Λ)	0	0	0	0.09	0
	Test statistic (T_n)	-0.29	-0.73	0.13	-0.18	-0.56
	p-value	0.46	0.39	0.52	0.47	0.42

Calculations are based on $k = 1\%$ of the standardized residuals of the filtered log return observations. NP denotes the empirical STDF estimator, NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

The results show that the standard estimate of the TDC is approximately equal to $\Lambda = 0.15$ for all time horizons, indicating that there might be a similar light tail dependence structure for these different return series. However, only for the 3-month horizon does the TDC-based test indicate that the extremal dependence structure is significantly different from being independent. The bias-corrected estimates indicate that the FX rates are tail independent. These findings suggest that

the considered EUR-exchange rates are in fact asymptotically independent but exhibit residual dependence above high but finite thresholds.

To get a better picture of the tail dependence structure between the three considered FX rates, the full extremal dependence structure is assessed based on pairwise estimates of the STDF. Figure 6.2.4 shows the beta-copula smoothed estimated ℓ-functions for 3-month log returns. As explained in Section 3.4.2, the stable tail dependence function restricted to the unit simplex contains all information on the function and is also commonly called the Pickands dependence function. It can be seen that the EURGBP and EURNOK rates are estimated to be asymptotically independent, whereas the other two FX pairs exhibit some very small deviations from tail independence.

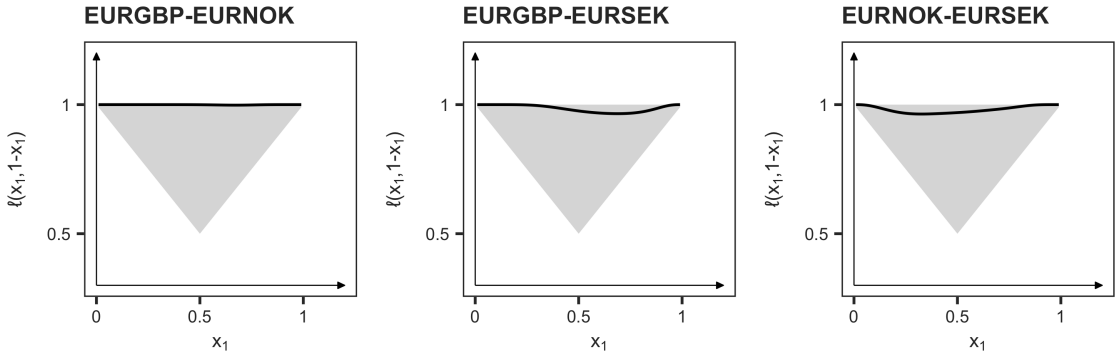

Figure 6.2.4: Estimated pairwise STDFs for EUR-exchange rates given a short EUR position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 3-month log returns.

However, pairwise dependence functions might not show the full picture of the extremal dependence structure since higher order interaction terms might also play a role. Therefore, the trivariate STDF is shown in Figure 6.2.5 below for both 3-month and annual log returns. The deviation from the tail independent structure is very small for both return horizons and the dependence structures look almost identical. Based on the visualizations of the data and the estimated tail dependence functions and coefficients, it can be concluded that the three European FX rates might exhibit very light tail dependence, mostly coming from the tail dependence between the EURSEK and the EURNOK rates, but are most likely asymptotically independent. This indicates that the considered EUR-exchange rates do not tend to show simultaneous positive shocks, limiting the tail risks associated with a short EUR position for these exposures.

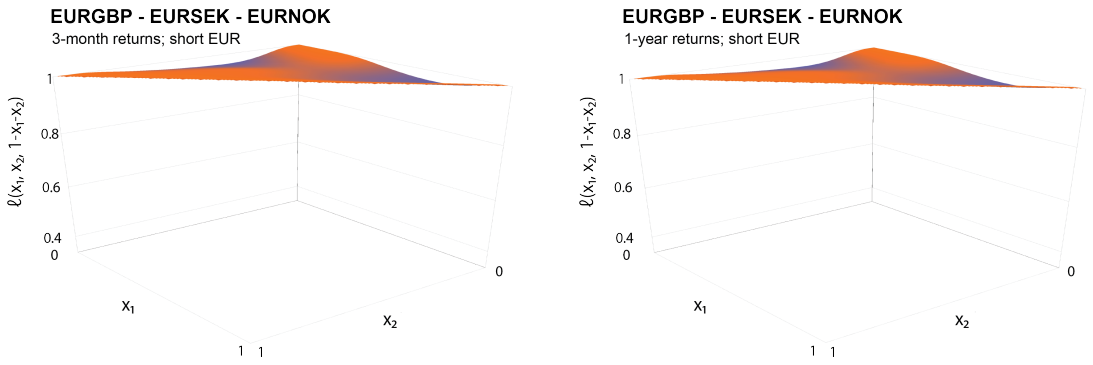

Figure 6.2.5: Estimated trivariate STDFs for EUR-exchange rates given a short EUR position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 3-month (left-hand-side) and 12-month (right-hand-side) log returns.

In contrast to the short EUR position, a long EUR position is now assumed for all European FX exposures. This corresponds to the scenario where a company makes expenses denoted in another currency and therefore has to sell Euros to retrieve the other currency. In this case, a valuable Euro is favorable and negative returns on the EUR exchange rate correspond to losses. Extreme losses are now represented by observations in the lower left orthant of the copula plot. By taking

negative log returns, these observations are moved to the upper right orthant and the analysis can be conducted in exactly the same way as before. Table 6.2.2 shows estimated TDC-values and testing results. The results suggest that for all considered time horizons of one month or longer, the data exhibit light tail dependence with an approximate TDC of $\Lambda = 0.13$. However, only for the 1-month horizon does the TDC-based hypothesis test indicate that the tail dependence structure is significantly different from the independent dependence structure. The bias-corrected results indicate again that the EUR-exchange rates are tail independent.

Table 6.2.2: Trivariate TDCs and test statistics with p-values for different return horizons of EUR-exchange rates given a long EUR position.

		1-day	1-month	3-months	6-months	12-months
Empirical STDF (NP)	TDC (Λ)	0.04	0.14	0.12	0.12	0.13
	Test statistic (T_n)	-1.50	-4.19	-1.81	-2.69	-2.19
	p-value	0.27	0.04	0.23	0.14	0.19
Bias-corrected (NP-BC3)	TDC (Λ)	0	0	0	0.07	0
	Test statistic (T_n)	0.02	-1.11	-0.95	-0.79	-0.27
	p-value	0.50	0.34	0.36	0.38	0.46

Calculations are based on $k = 1\%$ of the standardized residuals of the filtered log return observations. NP denotes the standard empirical STDF estimator, NP-BC3 denotes the bias-corrected estimator of Beirlant et al. (2016).

To get a better picture of the tail dependence structure between the three long EUR exposures, the full extremal dependence structure is assessed based on estimates of the pairwise STDF. Figure 6.2.6 shows the estimated pairwise ℓ-functions for the 1-month log returns based on the beta-copula empirical STDF. In contrast to the extremal dependence structure of losses for the short EUR position, all pairwise dependencies seem to contribute to the full dependence structure here. That is, all estimated pairwise STDFs deviate slightly from the tail independent STDF.

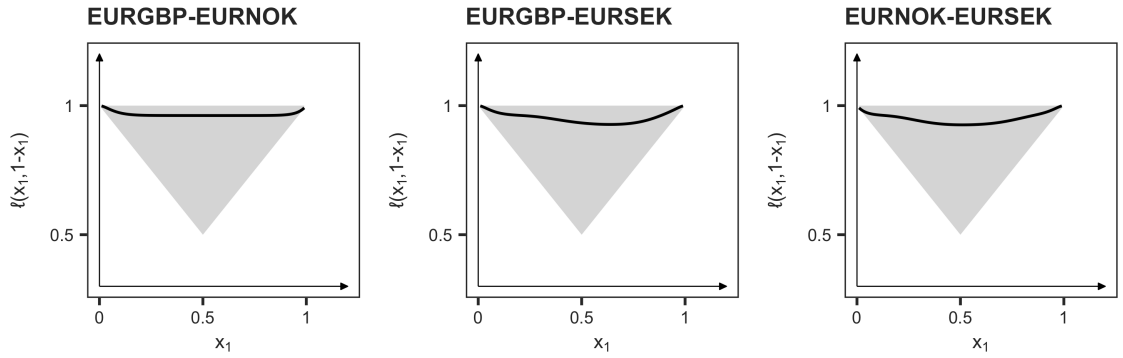

Figure 6.2.6: Estimated pairwise STDFs for EUR-exchange rates given a long EUR position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month log returns.

The full trivariate tail dependence function is shown in Figure 6.2.7 for 1-month and 1-year log returns. The tail dependence structures looks similar for both time horizons and deviates only slightly from the tail independent STDF. Compared to the short EUR position considered before (Figure 6.2.5), the tail dependence between losses on the three EUR-exchange rates is stronger for the long EUR position considered here. This could indicate that negative shocks to the relative value of the Euro are mainly driven by adverse shocks to the Euro, while positive shocks to the relative value of the Euro are mainly driven by approximately independent adverse shocks to the value of the other currencies under consideration. However, the tail dependence is very modest and based on the testing results and bias-corrected estimates it can be concluded that losses associated with both long and short EUR positions are tail independent for the considered European FX rates.

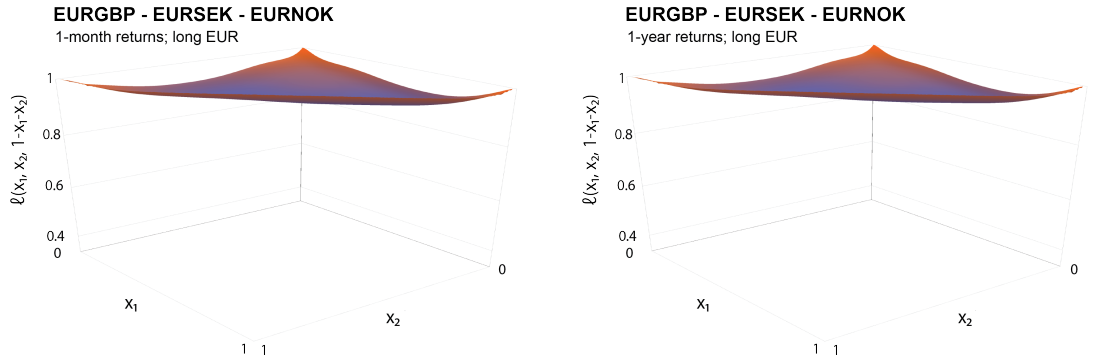

Figure 6.2.7: Estimated trivariate STDFs for EUR-exchange rates given a long EUR position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month (left-hand-side) and 12-month (right-hand-side) log returns.

6.3 Tail dependence in worldwide FX markets (5D)

This section concerns the five-dimensional tail dependence structure between the following five worldwide FX rates: EURUSD (Euro-US Dollar), GBPUSD (British Pound-US Dollar), USDCAD (US Dollar-Canadian Dollar), USDTRY (US Dollar-Turkish Lira), and USDRUB (US Dollar-Russian Ruble). The time series of these exchange rates are shown in Figure 6.3.1.

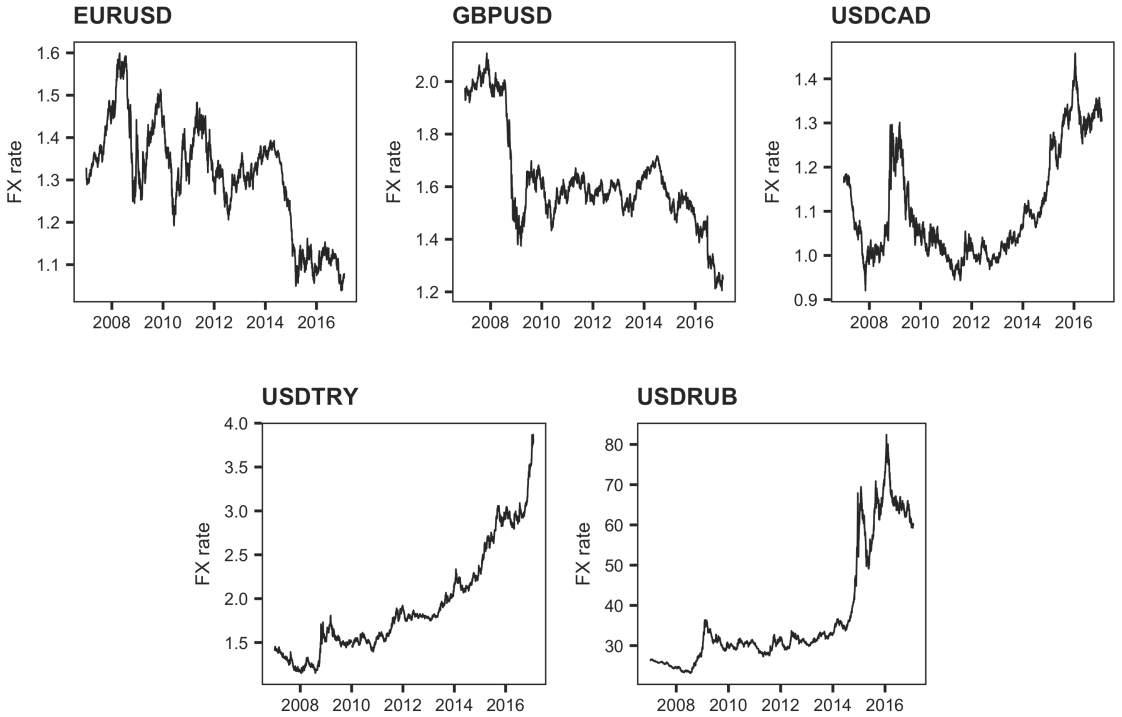

Figure 6.3.1: Time series of the considered USD-exchange rates.

Several trends can be observed from these time series. First of all, similar to the Euro, the US Dollar is a relatively large and stable currency and can be considered a safe haven currency during times of economic distress (Ranaldo and Söderlind, 2010). This might explain the increase in value of the USD relative to the GBP, CAD, TRY and RUB currencies during the past financial crisis. A sharp decline of the Euro against the Dollar can be observed in 2014. Factors that may have contributed to this movement are the quantitative easing policy of the ECB and the Greek sovereign debt crisis. The Russian Ruble also plunged against the US Dollar in this timeframe. Decreasing oil prices played a large role here, Russia being one of the main oil exporters in the world. The Turkish Lira has consistenly lost value against the US Dollar during the past decade.

This is mostly due to political instability causing a Turkish capital flight to more stable currencies such as the USD. Finally, the Brexit effect is also visible for the GBPUSD rate in 2016.

Following the methodology described in Section 6.1, the tail dependence structure between the FX rates is investigated based on standardized residuals of filtered marginal log returns. To get an initial overview of the data, the pairwise dependence structures are shown on the copula unit scale in Appendix D.3. A selection of pairwise copula plots is shown in Figure 6.3.2 below.

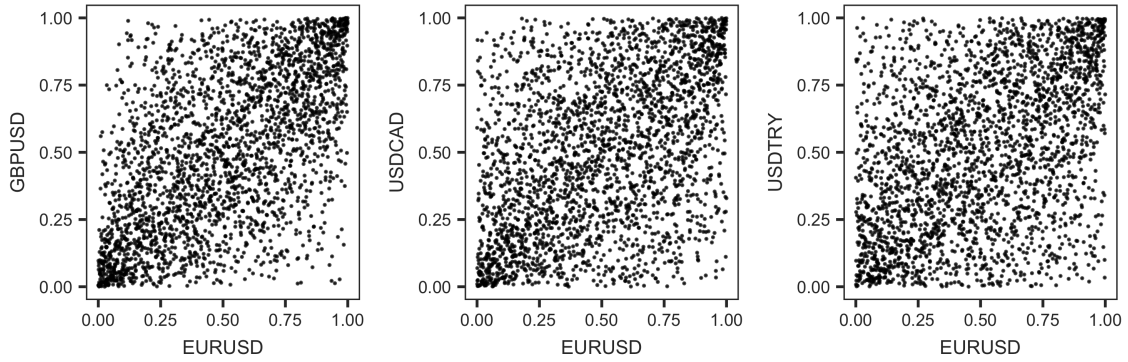

Figure 6.3.2: Selection of pairwise copula plots for the standardized residuals of ARMA-GJR-GARCH filtered daily log returns of five USD-exchange rates.

Overall, there seems to be a positive dependence structure between the considered variables, which is also reflected by the pairwise rank correlations (see Table 6.3.1). The presence of tail dependence is indicated by the clustering of observations in the upper right and lower left orthants of the copula plots. Generally speaking, extremal dependencies between the USD-exchange rates can be caused by either a single shock in the currency value of the common denominator, i.e., the US Dollar, or by a joint shock in two or more of the other currency values. Interlinkages between currency regions based on trading relationships or based on similar import or export products can cause such joint shocks in multiple currency values.

Table 6.3.1: Rank correlations between the considered USD-exchange rates.

	EURUSD	USDCAD	USDTRY	USDRUB	USDGBP
EURUSD	1.00	0.51	0.43	0.28	0.60
USDCAD	0.51	1.00	0.52	0.37	0.50
USDTRY	0.43	0.52	1.00	0.32	0.39
USDRUB	0.28	0.37	0.32	1.00	0.25
USDGBP	0.60	0.50	0.39	0.25	1.00

The correlations are computed for the ranked standardized residuals of the ARMA-GJR-GARCH filtered daily log returns of the FX rates. Positive returns correspond to an increasing value of the USD. The signs of the returns on the EURUSD and GBPUSD rates have been changed.

To continue the analysis, we first assume a short USD position for all FX exposures. This corresponds to the scenario where a company makes profits denoted in another currency and then buys US Dollars to retrieve its USD-denoted profits. Alternatively, this scenario corresponds to the unwinding of a currency carry trade. This investment strategy exploits differences between interest rates by lending money against a low interest rate in one currency (in this case the USD) and then receiving a higher interest rate in another currency (e.g., the Turkish Lira). At the end of the currency carry trade, the foreign currency has to be converted back to the USD to repay the loan. Since the future exchange rate at this point is uncertain, the strategy involves FX risk. A cheap USD is favorable and positive returns on the USD correspond to losses. Extreme losses are therefore represented by observations in the upper right orthant of the copula plot. By transforming the variables to the Pareto scale, we zoom in on these upper orthants. The figures are shown in Appendix D.3. A selection of the Pareto plots is shown in Figure 6.3.3 below. Recall that a concentration of points on the x- and y-axis indicates tail independence, whereas points scattered

between the axes indicates tail dependence. Since some observations are scattered between the axes, these figures support the hypothesis of tail dependence between the losses on a short USD exposure.

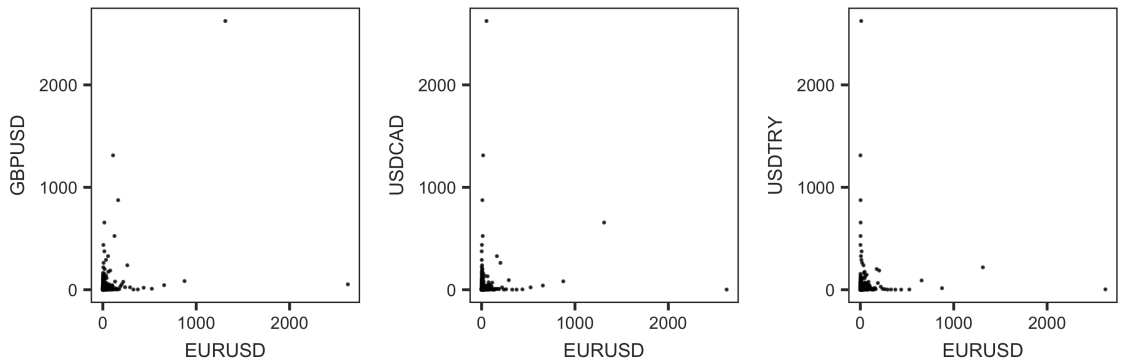

Figure 6.3.3: Selection of pairwise Pareto plots for the standardized residuals of ARMA-GJR-GARCH filtered daily log returns of five USD-exchange rates assuming a short USD position.

To quantify the strength of the tail dependence between the five USD-exchange rates, the multivariate tail dependence coefficient Λ and the related test statistics are determined (Table 6.3.2). Similar as for the trivariate European FX example, the results are shown for a range of different return horizons and for both the standard empirical STDF and the bias-corrected version thereof. The TDC-values that are estimated with the standard empirical STDF indicate that the five FX rates are tail dependent with a TDC-value ranging from $\Lambda = 0.21$ for a 1-day return horizon to $\Lambda = 0.36$ for a 1-month return horizon. On the other hand, the TDC-values that are estimated with the bias-corrected estimator indicate that the considered FX rates are tail independent. This is also suggested by most testing results; only for the 6-month return horizon does the TDC-based test reject the null hypothesis of tail independence.

Table 6.3.2: Multivariate TDCs and test statistics with p-values for different return horizons of USD-exchange rates given a short USD position.

		1-day	1-month	3-months	6-months	12-months
Empirical STDF	TDC (Λ)	0.21	0.36	0.25	0.24	0.23
(NP)	Test statistic (T_n)	-4.02	-1.68	-2.50	-5.76	1.15
	p-value	0.10	0.30	0.21	0.03	0.64
Bias-corrected	TDC (Λ)	0	0	0	0	0
(NP-BC3)	Test statistic (T_n)	0.48	-4.98	12.11	7.79	-1.37
	p-value	0.56	0.07	1.00	0.99	0.35

Calculations are based on $k = 1\%$ of the residuals of the filtered log return observations. NP denotes the standard empirical STDF estimator, NP-BC3 denotes the bias-corrected empirical STDF estimator of Beirlant et al. (2016).

To further assess the tail dependence structure, the STDF is estimated for all pairs of the considered FX rates. Figure 6.3.4 shows the estimated ℓ-function for a selection of FX pairs. See Appendix D.4 for all figures. All estimated dependence functions deviate quite a bit from the tail independent STDF. This indicates that all pairs contribute to the multivariate tail dependence structure between the five USD-exchange rates. The full tail dependence structure cannot be visualized since the problem is 5-dimensional. However, the trivariate ℓ-functions can be explored to further assess extremal dependence between the FX rates. Figure 6.3.5 shows the estimated trivariate STDF based on the beta-copula for the EURUSD-GBPUSD-USDRUB and the USDCAD-USDTRY-USDRUB. Both estimated dependence functions deviate significantly from the tail independent STDF. However, since the hypothesis tests and the bias-corrected results indicate tail independence, this might be due to strong residual dependence above high but finite thresholds.

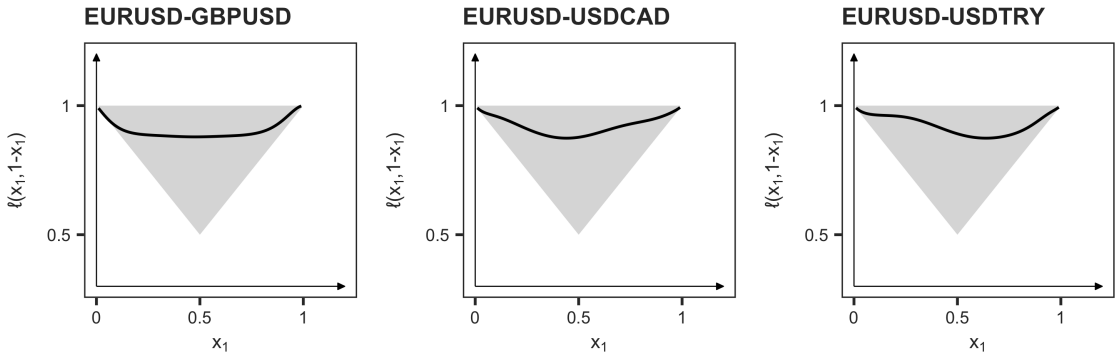

Figure 6.3.4: Selection of estimated pairwise STDFs for USD-exchange rates given a short USD position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month log returns.

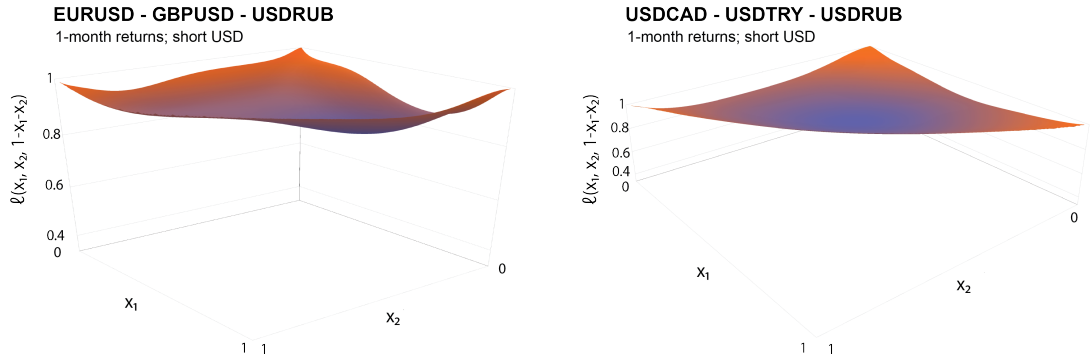

Figure 6.3.5: Selection of estimated trivariate STDFs for USD-exchange rates given a short USD position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month log returns.

Next, a long USD position is assumed for all FX exposures. Recall that this corresponds to the scenario where a company makes expenses in another currency and has to sell US Dollars. A strong USD is favorable and negative returns on the USD exchange rates correspond to losses. Extreme losses are represented by observations in the lower left orthant of the copula plots, but by taking negative log returns, these observations are shifted to the upper right orthant. Table 6.3.3 present the estimated multivariate TDCs and the test statistics. Whereas the standard empirical STDF indicate that the multivariate TDC is approximately equal to $\Lambda = 0.2 - 0.3$, the bias-corrected STDF estimator indicates again that the data are in fact tail independent.

Table 6.3.3: Multivariate TDCs and test statistics with p-values for different return horizons of USD-exchange rates given a long USD position.

		1-day	1-month	3-months	6-months	12-months
Empirical STDF (NP)	TDC (Λ)	0.28	0.29	0.20	0.21	0.25
	Test statistic (T_n)	-2.64	-7.78	-1.66	-3.21	-2.34
	p-value	0.20	0.01	0.30	0.15	0.23
Bias-corrected (NP-BC3)	TDC (Λ)	0	0	0	0	0
	Test statistic (T_n)	-2.05	-0.73	1.08	8.77	2.77
	p-value	0.28	0.42	0.62	0.99	0.79

Calculations are based on $k = 1\%$ of the residuals of the filtered log return observations. NP denotes the standard empirical STDF estimator, NP-BC3 denotes the bias-corrected empirical STDF estimator of Beirlant et al. (2016).

The tail dependence structure between losses on a long USD exposure for the considered USD-exchange rates can be further assessed based on estimates of the STDF. Figure 6.3.6 shows a selection of the estimated pairwise ℓ-functions and we refer to Appendix D.4 for all figures. Although

all pairs deviate slightly from the tail independent ℓ-function, the bivariate tail dependencies are weaker than for the short USD position. The estimated trivariate STDFs shown in Figure 6.3.7 also suggest this. In contrast to our findings for the EUR exposures, positive shocks to the value of the USD exhibit stronger tail dependence than negative shocks in the value of the USD relative to the other five considered currencies. The safe-haven property of the USD can offer an explanation for this phenomenon.

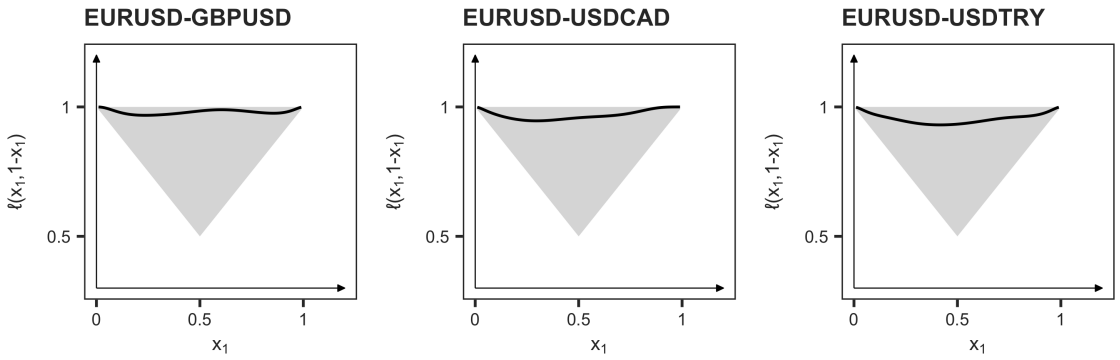

Figure 6.3.6: Selection of estimated pairwise STDFs for USD-exchange rates given a long USD position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month log returns.

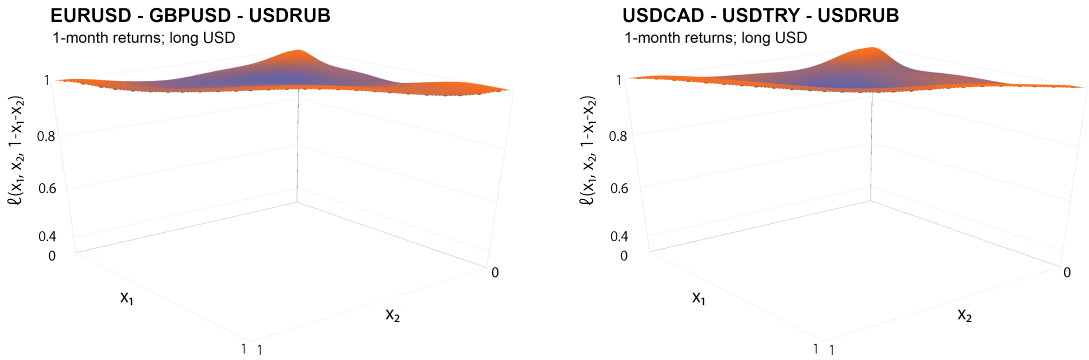

Figure 6.3.7: Selection of estimated trivariate STDFs for USD-exchange rates given a long USD position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month log returns.

Chapter 7

Conclusion

Multivariate tail dependence captures the extent to which extremes of a vector of random variables are likely to occur together. In the context of finance, simultaneous extreme events can have a significant impact on companies, investors, or the economy as a whole. Therefore, this thesis aimed to find suitable characterizations of the multivariate tail dependence structure, both theoretically and statistically. Based on such characterizations, tail risk can be better understood, and informed decisions regarding risk management can be made. Three types of tail dependence characterizations were distinguished: (1) the full tail dependence structure; (2) a summary measure of the strength of tail dependence; and (3) a classification into either tail dependence or tail independence. Our main findings regarding these characterizations are summarized.

The full tail dependence structure

There is a plethora of functions to characterize the full multivariate tail dependence structure. The multivariate extreme value distribution (MEVD) captures both the extremal behavior of the margins and the tail dependence structure. The exponent and spectral measure are alternative representations of the MEVD, but these measures also contain information on the marginal tails. A copula approach can be taken to disentangle the marginal behavior from the dependence structure. The copula of the MEVD is a particular type of copula, being an extreme value copula. Based on the extreme value copula, three dependence functions were considered: the stable tail dependence function (STDF), Pickands dependence function, and the tail copula. Whereas the STDF captures the conditional probability that at least one component of a random vector is large, the tail copula captures the conditional probability that all components of a random vector are large. The tail copula therefore only captures part of the multivariate extremal behavior, while the STDF fully describes it. Other attractive properties of the STDF are its bounds between tail independence and tail comonotonicity, homogeneity of order one, and convexity. Because of its homogeneity, it is sufficient to define the STDF on the unit simplex. This restriction is called the Pickands dependence function and allows for intuitive visualizations of the tail dependence structure for bivariate and trivariate problems.

The STDF can be estimated nonparametrically with the empirical STDF or with smoothed or bias-corrected versions thereof. Some parametric estimation methods are also available, but especially for higher dimensions, models that are currently developed lack flexibility. The simulation study presented in Section 5.1.5 showed that both the beta-copula and kernel-smoothed versions of the empirical STDF slightly outperform the standard estimator in terms of (integrated) mean squared error, but mainly offer advantages because visualizations of the smoothed STDF are more intuitive to interpret than the original jumpy empirical STDF. For copula distributions that converge slowly to their extreme value copula, the bias-corrected versions of the empirical STDF offer significant reductions in bias and mean squared error. The kernel-smoothed bias-corrected version of Beirlant et al. (2016) tends to outperform, but a disadvantage is that this method is computationally expensive.

Summary measures of tail dependence strength

Summary measures for the strength of tail dependence have been considered in literature almost exclusively for the bivariate case. The primary statistic is the tail dependence coefficient (TDC) λ defined as the limiting conditional probability of observing an extreme value of variable X_2 given that X_1 is extreme. The TDC is directly connected to the tail copula and the STDF, and, inherently, also to the other dependence characterizations. Extensions of the TDC to capture the strength of tail dependence for arbitrary dimensions $d \geq 2$ are considered based on the connection with the tail copula and the stable tail dependence function. The STDF-based extension Λ conveys more information than the tail copula-based extension τ and has some attractive properties. Most importantly, the multivariate TDC Λ is able to identify tail independence in all dimensions $d \geq 2$ which is captured by Main Theorem 2. Because of its relationship with the STDF, the multivariate TDC Λ can be estimated by evaluating the estimators for the full STDF in the point $(1, \ldots, 1)$. The bias-corrected estimates offer one decimal place accuracy for estimates concerning up to 5 random variables. Besides these first-order tail dependence measures, second order measures have been suggested to characterize the tail dependence structure in tail independent cases further. The tail order describes the residual dependence that is left above high but finite thresholds and can identify tail independence in the bivariate case, but not in higher dimensions.

Classification into tail dependence and tail independence

Finally, the classification of data as either tail dependent or tail independent can be made based on a hypothesis test. Since both the stable tail dependence function ℓ and the multivariate TDC Λ can identify tail independence and can be estimated nonparametrically, these characterizations are natural candidates to fashion an appropriate test statistic. However, under the null hypothesis of tail independence, the variance of the STDF estimators vanishes, and the distribution of the test statistic becomes degenerate. Hüsler and Li (2009) suggest to adjust the empirical STDF in such a way that high thresholds and exceedances are determined with independent datasets, resulting in a non-degenerate asymptotic distribution of the estimator under the null hypothesis of tail independence. Several new STDF-based test statistics were developed to evaluate the null hypothesis of tail independence by extending this idea to higher dimensions and other estimators. Specifically, we consider a supremum- and an integral-based test statistic for the empirical STDF, the kernel-smoothed version, and the bias-corrected version of Beirlant et al. (2016). Although these are all newly developed tests for the multivariate tail independence testing problem, our main contribution here is a test statistic based on the multivariate TDC Λ which is also able to identify dependence (Main Theorem 2) and can be estimated with one of the adjusted STDF estimators. It is much easier to evaluate the TDC-based test statistic, and since this test statistic is asymptotically Normal, critical values do not have to retrieved via simulations. Moreover, the simulation study in Section 5.2.3 indicates that the TDC-based test is more powerful than the test statistics based on the entire STDF. The smoothed and bias-corrected versions did not offer consistent improvements. As an alternative to the multivariate testing approach, a multiple pairwise testing approach was also considered, where the p-values for the bivariate tests were adjusted to correct for the multiple testing problem at hand. This multiple testing approach is significantly less powerful than the multivariate testing approach.

Suggestions for further research

Several directions for further research have been encountered. As mentioned in Section 5.2, relatively little is known on testing procedures to classify data as either tail dependent or tail independent, especially for multivariate data. We have contributed to existing literature by introducing several new multivariate test statistics based on the STDF and the multivariate TDC Λ for evaluating the null hypothesis of tail independence. The performance of these new testing procedures can be further assessed by taking into consideration higher-dimensional data or by exploring the effect of specifying other model parameters. Especially the parameters for bias-corrected testing procedures might improve the performance of those tests. The multiple testing approach presents another line of research. Methods to adjust the p-values for the multiple testing problem currently cause the tests to lose a significant amount of power. Alternative methods to undertake the multiple testing problem can be researched to improve the pairwise testing approach. More generally, efforts could be made to develop an improved testing procedure by adapting the STDF estimator

in such a way that the asymptotic distribution does not become degenerate *without splitting the sample*. Since the currently employed method does not use data efficiently, the adjusted empirical STDF has a relatively high variance and needs massive amounts of data to yield satisfactory results. An STDF-based testing statistic that uses data more efficiently might improve the accuracy of the hypothesis tests significantly.

Further research can also be undertaken concerning several applications of the multivariate tail dependence structure. First of all, bounds on tail risk measures such as the VaR and the ES can be developed based on our characterizations of tail dependence. As mentioned in Chapter 1, currently developed bounds on these tail risks measures entail too much uncertainty. By incorporating information on, for example, the classification of (groups of) data as tail dependent or tail independent, uncertainty can be reduced. The multivariate TDC Λ might also be applied to retrieve a more accurate estimation of tail risk. Many other applications in finance can also be considered for further research. For example, it would be exciting to incorporate the extremal dependence behavior into the pricing of basket options or credit default swaps. Another application could be to optimize portfolio compositions based on minimal tail risk constraints as measured by the multivariate TDC Λ.

Appendix A

Multivariate regular variation

Univariate regular variation

A key concept in extreme value analysis is regular variation. See for example Chapter 2 of Resnick (2007) for an exposition on the subject. Another standard reference on the topic of regular variation is Bingham et al. (1987). Resnick (2004) provides an accessible overview.

Definition A.1. *A measurable function $U : R_+ \to R_+$ is regularly varying at ∞ with index α (notation: $U \in \mathrm{RV}_\alpha$) if for any $x > 0$,*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\alpha. \tag{A.0.1}$$

For $\alpha = 0$ the function U is said to be slowly varying at ∞, written as $U \in \mathrm{RV}_0$.

The definition for (slow) regular variation at 0 can be defined similarly. A slowly varying function will be represented by $\mathscr{L}(x)$. Note that since $U(x)/x^\alpha \in RV_0$ for $U \in RV_\alpha$, a regularly varying function U can be written as $U(x) = x^\alpha \mathscr{L}(x)$. Hence, regularly varying functions behave asymptotically like power functions. It turns out, that regularly varying functions also behave like power functions when it comes to integrating and differentiating. This is given formally by Karamata's theorem. The proof is given in Resnick (2007).

Theorem A.1. ***(Karamata's theorem).*** *Suppose $\alpha \geq -1$ and $U \in RV_\alpha$. Then $\int_0^x U(t)\mathrm{d}t \in RV_{\alpha+1}$ and*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)\mathrm{d}t} = \alpha + 1.$$

If $\alpha < -1$ (or if $\alpha = -1$ and $\int_x^\infty U(s)\mathrm{d}s < \infty$), then $U \in RV_\alpha$ implies that $\int_x^\infty U(t)\mathrm{d}t$ is finite, $\int_x^\infty U(t)\mathrm{d}t \in RV_{\alpha+1}$, and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)\mathrm{d}t} = -\alpha - 1.$$

Furthermore, if U satisfies

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)\mathrm{d}t} = \beta \in (0, \infty),$$

then $U \in RV_{\beta-1}$. If $\int_x^\infty U(t)\mathrm{d}t < \infty$ and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)\mathrm{d}t} = \beta \in (0, \infty),$$

then $U \in RV_{-\beta-1}$.

For regularly varying distributions, Karamata's theorem implies a relationship between the distribution function and the density function. This is captured by the following theorem.

Theorem A.2. *Suppose that the distribution function $F : R_+ \to [0,1]$ is absolutely continuous with density f, i.e.,*

$$F(x) = \int_0^x f(t)\mathrm{d}t, \quad x \geq 0.$$

If $\overline{F} = 1 - F \in \mathrm{RV}_{-\alpha}, \alpha > 0$, and f is monotone, then $f \in \mathrm{RV}{-}\alpha - 1$.

Multivariate regular variation

Multivariate regular variation (MRV) can be most easily delt with on cones in $\mathbb{R}^d$. We say that $\mathbb{C} \subset \mathbb{R}^d$ is a cone if the following holds: $\mathbf{x} \in \mathbb{C}$ if and only if $t\mathbf{x} \in \mathbb{C}$ for every $t > 0$. In general, multivariate regularly varying functions are defined as follows (Resnick, 2007).

Definition A.2. *Suppose $h \geq 0$ is a measurable function defined on a cone $\mathbb{C}$. Suppose $\mathbf{1} = (1, \ldots, 1) \in \mathbb{C}$. We call h multivariate regularly varying with limit function ζ, provided $\zeta(\mathbf{x}) > 0$ for $\mathbf{x} \in \mathbb{C}$, and for all $\mathbf{x} \in \mathbb{C}$, we have*

$$\lim_{t\to\infty} \frac{h(t\mathbf{x})}{h(t\mathbf{1})} = \zeta(\mathbf{x}). \tag{A.0.2}$$

The limit function ζ necessarily is homogeneous of order α with $\alpha \in \mathbb{R}$. That is, $\zeta(s\mathbf{x}) = s^\alpha \zeta(\mathbf{x})$ for all $s > 0$. Furthermore, $h(t\mathbf{x}) \in RV_\alpha$.

For random variables, multivariate regular variation can be defined based on multivariate regular variation of the distribution function.

Definition A.3. *Let F be a d-variate distribution function with support $[0,\infty)^d$ and let $\mathbf{1} = (1, \ldots, 1)$ be a vector in $\mathbb{R}^d$. Then F is said to be regularly varying on $(0,\infty)^d$ if there exists a function $\lambda : (0,\infty)^d \to (0,\infty)$ such that*

$$\lim_{t\to\infty} \frac{1 - F(t\mathbf{x})}{1 - F(t\mathbf{1})} = \lambda(\mathbf{x}), \quad \mathbf{x} \in (0,\infty)^d, \tag{A.0.3}$$

with $\lambda(s\mathbf{x}) = s^{-\alpha}\lambda(\mathbf{x})$. That is, the limit function λ has to be homogeneous of order α.

This definition indicates that MRV describes the relative decay rates of joint tail probabilities of a random vector. For large values of the index α, the probability of an extreme event decays much faster than for small values of the index α.

Finally, an alternative definition of multivariate regular variation can be given as follows (Rvačeva, 1962).

Definition A.4. *A random vector $\mathbf{X} \geq \mathbf{0}$ is multivariate regularly varying if there exist an index $\alpha > 0$, and a Radon probability measure Φ (i.e., finite on compact sets) on $\Theta = \{\mathbf{z} \in \mathbb{R}^d : ||\mathbf{z}|| = 1\}$, the unit hypersphere, such that*

$$\lim_{t\to\infty} \frac{\mathbb{P}(||\mathbf{X}|| \geq tx, \mathbf{X}/||\mathbf{X}|| \in A)}{\mathbb{P}(||\mathbf{X}|| \geq t)} = x^{-\alpha}\Phi(A), \tag{A.0.4}$$

for every $x > 0$ and Borel set A in Θ with $\Phi(\partial A) = 0$, with $||\mathbf{X}||$ the L_2-norm of $\mathbf{X}$.

Equivalently, a random vector $\mathbf{X} \geq \mathbf{0}$ is said to be multivariate regularly varying if

$$\lim_{t\to\infty} \frac{\mathbb{P}(||\mathbf{X}|| \geq tx)}{\mathbb{P}(||\mathbf{X}|| \geq t)} = x^{-\alpha}, \tag{A.0.5}$$

for an index $\alpha > 0$ and for every $x > 0$, and if there exists a Radon measure μ (i.e., finite on compact sets) on $E := [\mathbf{0},\infty]\backslash\{\mathbf{0}\}$ such that,

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{X} \in tB)}{\mathbb{P}(||\mathbf{X}|| > t)} = \mu(B), \tag{A.0.6}$$

for any relatively compact set $B \subset E$ with $\mu(\partial B) = 0$, where $||\cdot||$ denotes a norm on $\mathbb{R}^d$ (Resnick, 1987). Note that μ is homogeneous of order α and is a probability measure on $\{\mathbf{z} \in \mathbb{R}^d : ||\mathbf{z}|| \geq 1\}$. The intensity measure μ that satisfies this scaling property of order $-\alpha$, contains the extremal dependence information of $\mathbf{X}$. For large values of α, the conditional probability of an extreme set decays much faster than for small values of α.

Appendix B

Proofs

B.1 D-vine tail density

Theorem 3.7. *(Li and Wu, 2013). Suppose C_F is a d-dimensional* D*-vine copula and assume that all bivariate linking copulas $C_{F;ij}$ have continuous densities and satisfy the uniform convergence properties. If the baseline linking copula $C_{F;i,i+1}$ are all upper tail dependent, then*

$$\frac{\Upsilon_{\{1,\dots,d\}}(\mathbf{w})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w}_{\{2,\dots,d-1\}})} = \frac{\Upsilon_{\{1,\dots,d-1\}}(\mathbf{w}_{\{1,\dots,d-1\}})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w}_{\{2,\dots,d-1\}}} \frac{\Upsilon_{\{2,\dots,d\}}(\mathbf{w}_{\{2,\dots,d\}})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w}_{\{2,\dots,d-1\}})} \cdot c_{1,d|2,\dots,d-1}\left(1 - t_{1|2,\dots,d-1}\left(w_1|\mathbf{w}_{\{2,\dots,d-1\}}\right), 1 - t_{d|2,\dots,d-1}\left(w_d|\mathbf{w}_{\{2,\dots,d-1\}}\right)\right),$$

where for $j \in \{1,\dots,d\}$, and $S \subseteq \{1,\dots,d\}\backslash\{j\}$,

$$\begin{aligned}\bar{t}_{j|S}\left(w_j|\mathbf{w}_S\right) &= \lim_{u\downarrow 0} \overline{C}_{j|S}\left(1 - uw_j | 1 - uw_i, i \in S\right) \\ &= \int_0^{w_j} \frac{\Upsilon_{S\cup\{j\}}(\tilde{w}_j, \mathbf{w}_S)}{\Upsilon_S(w_S)} \mathrm{d}\tilde{w}_j.\end{aligned}$$

Proof. (Different from the proof presented by Li and Wu (2013)). Recall from Section 2.4 that the density c of the d-dimensional D-vine copula C is given by

$$c(x_1,\dots,x_d) = \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j|i+1,\dots,i+j-1}\Big(C_{i|i+1,\dots,i+j-1}(x_i|x_{i+1},\dots,x_{i+j-1}), C_{i+j|i+1,\dots,i+j-1}(x_{i+j}|x_{i+1},\dots,x_{i+j-1})\Big).$$

The tail density for the D-vine copula is therefore defined as,

$$\begin{aligned}\Upsilon_{\{1,\dots,d\}}(\mathbf{w}) &= \lim_{t\downarrow 0} t^{d-1} c(1 - tw_i, 1 \le i \le d) \\ &= \lim_{t\downarrow 0} t^{d-1} \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j|i+1,\dots,i+j-1}\Big(C_{i|i+1,\dots,i+j-1}(1 - tw_i | 1 - tw_{i+1},\dots,1 - tw_{i+j-1}), \\ &\qquad C_{i+j|i+1,\dots,i+j-1}(1 - tw_{i+j} | 1 - tw_{i+1},\dots,1 - tw_{i+j-1})\Big).\end{aligned}$$

Note that for a subset $S \subseteq \{1,\dots,d\}$ the following holds,

$$\begin{aligned}\lim_{t\downarrow 0} C_{i|S}(1 - tw_i | 1 - tw_j, j \in S) &= \lim_{t\downarrow 0} \mathbb{P}\left(U_i > 1 - tx_i | U_j = 1 - tx_j, j \in S\right) \\ &= \lim_{t\downarrow 0} \int_{1-tx_i}^{1} c_{i|S}(w_i | 1 - tx_j, j \in S) \mathrm{d}w_i \\ &= \lim_{t\downarrow 0} \int_0^{x_i} t \cdot c_{i|S}(1 - tv_i | 1 - tx_j, j \in S) \mathrm{d}v_i \\ &= \int_0^{x_i} \frac{\Upsilon_{i\cup S}(v_i, x_j, j \in S)}{\Upsilon_S(x_j, j \in S)} \mathrm{d}v_i,\end{aligned}$$

where Υ_S denotes the marginal copula tail density of the variables $X_i : i \in S$.

For notational convenience, the arguments of the conditional copula densities are dropped in the following, but they are always defined as above. Similar expressions for partial copula tail density functions can then be derived as

$$\Upsilon_{\{1,\dots,d-1\}}(\mathbf{w}) = \lim_{t\downarrow 0} t^{d-2} c(1-tw_i, 1 \le i \le d-1) = \lim_{t\downarrow 0} t^{d-2} \prod_{j=1}^{d-2} \prod_{i=1}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1},$$

$$\Upsilon_{\{2,\dots,d\}}(\mathbf{w}) = \lim_{t\downarrow 0} t^{d-2} c(1-tw_i, 2 \le i \le d) = \lim_{t\downarrow 0} t^{d-2} \prod_{j=1}^{d-2} \prod_{i=2}^{d-j} c_{i,i+j|i+1,\dots,i+j-1},$$

$$\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w}) = \lim_{t\downarrow 0} t^{d-3} c(1-tw_i, 2 \le i \le d-1) = \lim_{t\downarrow 0} t^{d-3} \prod_{j=1}^{d-3} \prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1}.$$

Using these expressions, it can be deduced that

$$\begin{aligned}
\frac{\Upsilon_{\{1,\dots,d\}}(\mathbf{w})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w})} &= \lim_{t\downarrow 0} t^2 \frac{\left(\prod_{j=1}^{d-3}\prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1} \cdot c_{1,1+j|2,\dots,j} \cdot c_{d-j,d|d-j+1,\dots,d-1}\right)}{\prod_{j=1}^{d-3}\prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1}} \\
&\qquad \times \frac{c_{1,d-1|2,\dots,d-2} \cdot c_{2,d|3,\dots,d-1} \cdot c_{1,d|2,\dots,d-1}}{\prod_{j=1}^{d-3}\prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1}} \\
&= \lim_{t\downarrow 0} t^2 \cdot c_{1,d-1|2,\dots,d-2} \cdot c_{2,d|3,\dots,d-1} \cdot c_{1,d|2,\dots,d-1} \prod_{j=1}^{d-3} c_{1,1+j|2,\dots,j} \cdot c_{d-j,d|d-j+1,\dots,d-1}, \\
\frac{\Upsilon_{\{1,\dots,d-1\}}(\mathbf{w})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w})} &= \lim_{t\downarrow 0} t \frac{\left(\prod_{j=1}^{d-3}\prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1} \cdot c_{1,1+j|2,\dots,j}\right) \cdot c_{1,d-1|2,\dots,d-2}}{\prod_{j=1}^{d-3}\prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1}} \\
&= \lim_{t\downarrow 0} t \cdot c_{1,d-1|2,\dots,d-2} \prod_{j=1}^{d-3} c_{1,1+j|2,\dots,j}, \\
\frac{\Upsilon_{\{2,\dots,d\}}(\mathbf{w})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w})} &= \lim_{t\downarrow 0} t \frac{\left(\prod_{j=1}^{d-3}\prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1} \cdot c_{d-j,d|d-j+1,\dots,d-1}\right) \cdot c_{2,d|3,\dots,d-1}}{\prod_{j=1}^{d-3}\prod_{i=2}^{d-1-j} c_{i,i+j|i+1,\dots,i+j-1}} \\
&= \lim_{t\downarrow 0} t \cdot c_{2,d|3,\dots,d-1} \prod_{j=1}^{d-3} c_{d-j,d|d-j+1,\dots,d-1},
\end{aligned}$$

after which it is obvious that

$$\frac{\Upsilon_{\{1,\dots,d\}}(\mathbf{w})}{\lambda^U_{\{2,\dots,d-1\}}(\mathbf{w}_{\{2,\dots,d-1\}})} = \lim_{t\downarrow 0} c_{1,d|2,\dots,d-1} \cdot \frac{\Upsilon_{\{1,\dots,d-1\}}(\mathbf{w}_{\{1,\dots,d-1\}})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w}_{\{2,\dots,d-1\}}} \frac{\Upsilon_{\{2,\dots,d\}}(\mathbf{w}_{\{2,\dots,d\}})}{\Upsilon_{\{2,\dots,d-1\}}(\mathbf{w}_{\{2,\dots,d-1\}})}.$$

This finalizes the proof. ■

Proposition 3.3. *Based on the representation of the copula tail density function for a* D*-vine copula, the following results can be deduced:*

1. *If some baseline linking copulas $C_{i,i+1}$ are tail independent (i.e., $\Upsilon_{i,i+1} = 0$ for some $1 \leq i \leq d$), then $\Upsilon(\mathbf{w}) = 0$ (Li and Wu, 2013).*
2. *If all baseline linking copulas $C_{i,i+1}$ are tail dependent (i.e., $\Upsilon_{i,i+1} > 0$ for some $1 \leq i \leq d$) and if all linking copulas $C_{\{i,j|i+1,\ldots,j-1\}}$ have a non-zero density everywhere on $[0,1]^2$, then the* D*-vine copula is tail dependent (i.e., $\Upsilon(\mathbf{w}) > 0$).*

Proof. Note that the full copula density can also be decomposed as follows,

$$\begin{aligned}
\Upsilon_{\{1,\ldots,d\}}(\mathbf{w}) &= \lim_{t\downarrow 0} t^{d-1} c(1-tw_i, 1 \leq i \leq d) \\
&= \lim_{t\downarrow 0} t^{d-1} \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}\Big(C_{i|i+1,\ldots,i+j-1}(1-tw_i|1-tw_{i+1},\ldots,1-tw_{i+j-1}), \\
&\qquad\qquad C_{i+j|i+1,\ldots,i+j-1}(1-tw_{i+j}|1-tw_{i+1},\ldots,1-tw_{i+j-1})\Big) \\
&= \lim_{t\downarrow 0} t^{d-1} \left(\prod_{i=1}^{d-1} c_{i,i+1}\left(1-tw_i, 1-tw_{i+1}\right)\right) \\
&\quad \cdot \left(\prod_{j=2}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}\Big(C_{i|i+1,\ldots,i+j-1}(1-tw_i|1-tw_{i+1},\ldots,1-tw_{i+j-1}), \right. \\
&\qquad\qquad \left. C_{i+j|i+1,\ldots,i+j-1}(1-tw_{i+j}|1-tw_{i+1},\ldots,1-tw_{i+j-1})\Big)\right),
\end{aligned}$$

where

$$\begin{aligned}
\lim_{t\downarrow 0} t^{d-1} \prod_{i=1}^{d} c_{i,i+1}(1-tw_i, 1-tw_{i+1}) &= \prod_{i=1}^{d-1} \lim_{t\downarrow 0} t \cdot c_{i,i+1}\left(1-tw_i, 1-tw_{i+1}\right) \\
&= \prod_{i=1}^{d-1} \Upsilon_{\{i,i+1\}}(w_i, w_{i+1}).
\end{aligned}$$

Hence, the D-vine copula density can be decomposed in the product of the bivariate copula densities of the baseline linking copulas and terms coming from the higher level conditional copulas (see Section 2.4 for terminology). It is straightforward to see from this decomposition that if some of the baseline linking copulas $C_{i,i+1}$ are tail independent (i.e., $\Upsilon_{i,i+1} = 0$ for some $1 \leq i \leq d$), then the full copula density equals $\Upsilon_{\{1,\ldots,d\}} = 0$. In contrast, if all baseline linking copulas $C_{i,i+1}$ are tail dependent (i.e., $\Upsilon_{i,i+1} > 0$ for all $1 \leq i \leq d$) and all bivariate linking copulas have a non-zero mass everywhere on $[0,1]^2$, then the full copula tail density is positive (i.e., $\Upsilon_{\{1,\ldots,d\}} > 0$) indicating that there is tail dependence. ■

B.2 Euler representation of the tail copula

Because of the homogeneity of the tail copula, Euler's formula provides us with the following representation,

$$b(x_1,\dots,x_d)=\sum_{j=1}^{d}x_j\frac{\partial b}{\partial x_j}.$$

Since

$$\begin{aligned}\mathbb{P}\left(X>x|Y=y\right)&=\int_x^\infty f_{X|Y}(\omega|y)\mathrm{d}\omega=\int_x^\infty\frac{f_{XY}(\omega,y)}{f_Y(y)}\mathrm{d}\omega\\&=-\frac{1}{f_Y(y)}\frac{\partial}{\partial y}\int_y^\infty\int_x^\infty f_{XY}(\omega,z)\mathrm{d}\omega\mathrm{d}z=-\frac{1}{f_Y(y)}\frac{\partial}{\partial y}\mathbb{P}\left(X>x,Y>y\right),\end{aligned}$$

it follows that

$$\begin{aligned}\frac{\partial b}{\partial x_j}&=\frac{\partial}{\partial x_j}\lim_{t\downarrow0}\frac{\overline{C}_F(1-tx_1,\dots,1-tx_d)}{t}=\frac{\partial}{\partial x_j}\lim_{t\downarrow0}\frac{\mathbb{P}(U_1>1-tx_1,\dots,U_d>1-tx_d)}{t}\\&=\lim_{t\downarrow0}\mathbb{P}\left(U_i>1-tx_i,i\in I_j|U_j=1-tx_j\right)=\lim_{t\downarrow0}\overline{C}_{F;I_J|\{j\}}\left(1-tx_i,i\in I_j|1-tx_j\right).\end{aligned}$$

Defining

$$\begin{aligned}\bar{t}_{S_1|S_2}(x_i,i\in S_1|x_j,j\in S_2)&=\lim_{t\downarrow0}\mathbb{P}(U_i>1-tx_i,i\in S_1|U_j=1-tx_j,j\in S_2)\\&=\lim_{t\downarrow0}\overline{C}_{F;S_1|S_2}\left(1-tx_i,i\in S_1|1-tx_j,j\in S_2\right),\end{aligned}$$

we can write

$$b(x_1,\dots,x_d)=\sum_{j=1}^{d}x_j\frac{\partial b}{\partial x_j}=\sum_{j=1}^{d}x_j\bar{t}_{I_j|\{j\}}. \tag{B.2.1}$$

Based on this representation, Joe et al. (2010) present several interesting results that can ultimately be used to write the tail copula recursively in terms of the copula density.

Proposition B.1. *For all* $1\leq j\leq d$

$$b(x_1,\dots,x_d)=\int_0^{x_j}\bar{t}_{I_j|\{j\}}(x_i,i\in I_j|x)\mathrm{d}x=\int_0^{x_j}\lim_{t\downarrow0}\overline{C}_{F;I_j|\{j\}}(1-tx_i,i\in I_j|1-tx_j)\mathrm{d}x. \tag{B.2.2}$$

Proof. Firstly, note that

$$\int_0^{x_j}\frac{\partial b(x_1,\dots,\omega_j,\dots,x_d)}{\partial\omega_j}\mathrm{d}\omega_j=[b(x_1,\dots,\omega_j,\dots,x_d]_0^{x_j}=b(x_1,\dots,x_d),$$

where we have used that b is grounded. Secondly,

$$\begin{aligned}\frac{\partial b(x_1,\dots,x_d)}{\partial x_j}&=\frac{\partial}{\partial x_j}\left(\sum_{k=1}^{d}x_k\bar{t}_{I_k|\{k\}}\right)=\frac{\partial}{\partial x_j}\left(\sum_{k\in\{1,\dots,d\}\setminus\{j\}}x_k\bar{t}_{I_k|\{k\}}+w_j\bar{t}_{I_j|\{j\}}\right)\\&=\sum_{k\in\{1,\dots,d\}\setminus\{j\}}\frac{\partial}{\partial x_j}x_k\lim_{t\downarrow0}\mathbb{P}\left(U_i>1-tx_i,i\in I_k|U_k=1-tx_k\right)\\&\qquad+\frac{\partial}{\partial x_j}x_j\lim_{t\downarrow0}\mathbb{P}\left(U_i>1-tx_i,i\in I_j|U_j=1-tx_j\right)\\&=\sum_{k\in\{1,\dots,d\}\setminus\{j\}}x_k\lim_{t\downarrow0}(-t)\mathbb{P}\left(U_i>1-tx_i,i\in I_k\setminus\{j\}|U_k=1-tx_k,U_j=1-tx_j\right)\\&\qquad+\lim_{t\downarrow0}\mathbb{P}\left(U_i>1-tx_i,i\in I_j|U_j=1-tx_j\right)\\&\qquad+x_j\lim_{t\downarrow0}(-t)\frac{\partial}{\partial x}\mathbb{P}\left(U_i>1-tx_i,i\in I_j|U_j=x\right)\\&=\lim_{t\downarrow0}\mathbb{P}\left(U_i>1-tx_i,i\in I_j|U_j=1-tx_j\right)=\bar{t}_{I_j|\{j\}}.\end{aligned}$$

It follows that

$$b(x_1,\dots,x_d) = \int_0^{x_j} \frac{\partial b(x_1,\dots,x,\dots,x_d)}{\partial x}\mathrm{d}x = \int_0^{x_j} \bar{t}_{I_j|\{j\}}(x_i, i\in I_j|x)\mathrm{d}x$$
$$= \int_0^{x_j} \lim_{t\downarrow 0} \overline{C}_F(1-tx_i, i\in I_j|1-tx)\mathrm{d}x.$$

■

This result leads to the following representation of the tail copula.

Theorem B.1. *(Joe et al., 2010). Let $S_1, S_2 \subseteq I$ be two non-empty subsets of indices with $S_1 \cap S_2 = \emptyset$ and $S_1 \cup S_2 = I$. W.l.o.g., write $S_1 = \{1,\dots,s\}$ and $S_2 = \{s+1,\dots,d\}$ with $1 \le s \le d$. Assume that the copula C has continuous partial derivatives of order $(|S_2|+1)$. Then the tail copula is given by*

$$b(x_1,\dots,x_d) = \int_0^{x_{s+1}} \cdots \int_0^{x_d} t_{S_1|S_2}(x_1,\dots,x_s|\omega_{s+1},\dots,\omega_d)\frac{\partial^{d-s} b_{S_2}(\omega_{s+1},\dots,\omega_d)}{\partial\omega_{s+1}\dots\partial\omega_d}\mathrm{d}\omega_{s+1}\dots\mathrm{d}\omega_d. \tag{B.2.3}$$

Proof. (Adapted from the proof in Joe et al. (2010) given for the lower tail copula.) Writing $t_i = tx_i$, we first derive an expression for $t_{S_1|S_2}$ as follows (using the rules for differentiating under the integral sign),

$$\begin{aligned}
\bar{t}_{S_1|S_2}(x_1,\dots,x_s|x_{s+1},\dots,x_d) &= \lim_{t\downarrow 0}\overline{C}_{F;S_1|S_2}(1-tx_i, i\in S_1|1-tx_j, j\in S_2)\\
&= \lim_{t\downarrow 0}\mathbb{P}(U_i > 1-t_i, i\in S_1|U_j = 1-t_j, j\in S_2)\\
&= \lim_{t\downarrow 0}\int_{1-t_1}^{\infty}\cdots\int_{1-t_s}^{\infty}\frac{c_F(\omega_1,\dots,\omega_s,1-t_{s+1},\dots,1-t_d)}{c_{F;S_2}(1-t_{s+1},\dots,1-t_d)}\mathrm{d}\omega_1\dots\mathrm{d}\omega_s\\
&= \lim_{t\downarrow 0}\frac{\partial^{d-s}\overline{C}_F(1-t_1,\dots,1-t_d)/\partial t_{s+1}\dots\partial t_d}{c_{F;S_2}(1-t_{s+1},\dots,1-t_d)}\\
&= \lim_{t\downarrow 0}\frac{\partial^{d-s}\overline{C}_F(1-t_1,\dots,1-t_d)/\partial t_{s+1}\dots\partial t_d}{\partial^{d-s}\overline{C}_{F;S_2}(1-t_{s+1},\dots,1-t_d)/\partial t_{s+1}\dots\partial t_d}
\end{aligned}$$

Since by a simple application of the chain rule it follows that

$$\begin{aligned}
\frac{\partial^{d-s}\overline{C}_F(1-t_1,\dots,1-t_d)}{\partial t_{s+1}\dots\partial t_d} &= (-t)^{-d+s}\frac{\partial^{d-s}\overline{C}_F(1-tx_1,\dots,1-tx_d)}{\partial x_{s+1}\dots\partial x_s}\\
&\approx t(-t)^{-d+s}\frac{\partial^{d-s}b(x_1,\dots,x_d)}{\partial x_{s+1}\dots\partial x_d},
\end{aligned}$$

as $t\downarrow 0$, we can write

$$\bar{t}_{S_1|S_2}(x_1,\dots,x_s|x_{s+1},\dots,x_d) = \frac{\partial^{d-s}b(x_1,\dots,x_d)/\partial x_{s+1}\dots\partial x_d}{\partial^{d-s}b_{S_2}(x_{s+1},\dots,x_d)/\partial x_{s+1}\dots\partial x_d}.$$

By applying Proposition B.1 multiple times, it follows that

$$b(x_1,\dots,x_d) = \int_0^{x_{s+1}} \cdots \int_0^{x_d} t_{S_1|S_2}(x_1,\dots,x_s|\omega_{s+1},\dots,\omega_d)\frac{\partial^{d-s} b_{S_2}(\omega_{s+1},\dots,\omega_d)}{\partial\omega_{s+1}\dots\partial\omega_d}\mathrm{d}\omega_{s+1}\dots\mathrm{d}\omega_d.$$

■

If the conditional tail dependence functions $\bar{t}_{I_j|\{j\}}(\cdot|w_j), 1 \leq j \leq d$, are proper distribution functions, then the marginal tail dependence functions are given by evaluation of the tail copula in ∞ (Joe et al., 2010).

Theorem B.2. *(Joe et al., 2010). If all bivariate linking copulas of a D-vine C have continuous second-order partial derivatives, then the tail copula is given by the recursions (for $1 \leq i < i+l \leq d, l \geq 2$):*

$$b_{\{i,\dots,i+l\}}(x_i,\dots,x_{i+l}) = \int_0^{x_{i+1}} \cdots \int_0^{x_{i+l-1}} \bar{t}_{\{i,i+l\}|\{i+1,\dots,i+l-1\}}(x_i,\dots,x_{i+l}|\omega_{i+1},\dots,\omega_{i+l-1}) \times \frac{\partial^{l-1} b_{\{i+1,\dots,i+l-1\}}(\omega_{i+1},\dots,\omega_{i+l-1})}{\partial\omega_{i+1}\dots\partial\omega_{i+l-1}} \mathrm{d}\omega_{i+1}\dots\mathrm{d}\omega_{i+l-1},$$

up to

$$b(x_1,\dots,x_d) = \int_0^{x_2} \cdots \int_0^{x_{d-1}} \bar{t}_{\{1,d\}|\{2,\dots,d-1\}}(x_1,x_d)|\omega_2,\dots,\omega_{d-1}) \times \frac{\partial^{d-2} b_{\{2,\dots,d-1\}}(\omega_2,\dots,\omega_{d-1})}{\partial\omega_2\dots\partial\omega_{d-1}} \mathrm{d}\omega_2\dots\mathrm{d}\omega_{d-1}.$$

If, furthermore, the supports of the bivariate linking copulas are the entire $[0,1]^2$ and the baseline copulas are all tail dependent, then C is upper tail dependent.

Proof. The recursions follow immediately from theorem B.1. ■

B.3 TDC-based identification of tail independence

B.3.1 Main Theorem 1

Main Theorem 1. *Let X_1 and X_2 be two continuous random variables joint with a copula C_F that belongs to the maximum domain of attraction of an extreme value copula. The variables X_1 and X_2 are asymptotically independent if and only if their tail dependence coefficient $\lambda(C_F)$ is equal to zero. That is,*

$$C_F \in MDA(\Pi) \Leftrightarrow \lambda(C_F) = 0.$$

Proof. By the relationship between the TDC and the STDF (Equation 4.2.3) it follows immediately that if $C_F \in MDA(\Pi)$, then $\lambda(C_F) = 2 - \ell(1,1) = 2 - (1+1) = 0$. Conversely, we have to show that if $\lambda(C_F) = 0$, then $C_F \in MDA(\Pi)$. By the relationship between the extreme value copula and the STDF (Proposition 3.5), it is sufficient to show that if $\lambda(C_F) = 0$, then $\ell(x_1, x_2) = x_1 + x_2$ for any $x_1, x_2 \in \mathbb{R}_+$. Alternatively, by the Pickands representation (Section 3.4.2), it is sufficient to show that if $\lambda(C_F) = 0$, then $A(t) = A(t, 1-t) = 1$ for any $t \in [0,1]$. Without making use of the maximimum principle for convex functions, this is done as follows.

- First of all, note that if $\lambda(C_F) = 0$, the relationship between the TDC and the Pickands dependence function implies that

$$A(1/2) = 1.$$

- Recall the properties of the Pickands dependence function from Section 3.4.2. The A-function is convex and is equal to 1 if it is evaluated in one of the unit vectors. Hence,

$$A(1) = A(0) = 1.$$

  Furthermore, the bounds of the A-function imply that $A(t) \leq 1$ for all $t \in [0,1]$. Hence, it is sufficient to show that $A(t) \geq 1$ for all $t \in [0,1]$ in order to prove that $A(t) = 1$ for all $t \in [0,1]$.

- The proof relies on the convexity of the Pickands dependence function. Formally, the convexity of A implies that for all $t_1, t_2 \in [0,1]$ and $\alpha \in [0,1]$ the following defining relationship holds,

$$A\left(\alpha t_1 + (1-\alpha)x_2\right) \leq \alpha A(t_1) + (1-\alpha)A(t_2).$$

  This implies that for any $t \in [0,1]$, $t_2 \in [0,1]$ and $\alpha \in [0,1]$,

$$A(t) \geq \frac{1}{\alpha}\left(A\left(\alpha t + (1-\alpha)t_2\right) - (1-\alpha)A(t_2)\right).$$

- The concept of the proof is to show that due to the convexity of A, the only line that can pass through the three given points, is the horizontal line $A(t) = 1$. This is done by picking a point $t \in (0,1)$ and by picking a boundary point t_2 (i.e., $t_2 = 0$ or $t_2 = 1$) such that the mid point $1/2$ can be retrieved from a linear combination of t and t_2. Note that it is always possible to pick a boundary point t_2 such that this is possible. In order for the convexity relationship to hold, we must have that for the point $t \in (0,1)$, $A(t) \geq 1$. Hence, $A(t) = 1$.

- Formally, let $t \in (0,1)$ and consider specifying α and t_2 such that $\alpha t + (1-\alpha)t_2 = 1/2$ with $t_2 \in \{0,1\}$.

  - If $t \in (0, 1/2)$, we find by putting $t_2 = 1$ and $\alpha = 1/(2-2t) \in (0,1)$ that

$$\begin{aligned} A(t) &\geq \frac{1}{\alpha}\left(A\left(\frac{t}{2-2t} + \frac{1-2t}{2-2t}\right) - (1-\alpha)A(1)\right) \\ &= \frac{1}{\alpha}\left(A\left(\frac{1}{2}\right) - (1-\alpha)A(1)\right) = \frac{1}{\alpha}(1-(1-\alpha)) = 1. \end{aligned}$$

– Similarly, if $t \in (1/2, 1)$, we find by putting $t_2 = 0$ and $\alpha = 1/(2t) \in (0, 1)$ that

$$\begin{aligned} A(t) &\geq \frac{1}{\alpha}\left(A\left(\frac{t}{2t}\right) - (1-\alpha)A(0)\right) \\ &= \frac{1}{\alpha}\left(A\left(\frac{1}{2}\right) - (1-\alpha)A(0)\right) = \frac{1}{\alpha}(1-(1-\alpha)) = 1. \end{aligned}$$

- Hence, for any $t \in [0, 1]$ we have that $1 \leq A(t) \leq 1$, indicating that $A(t) = 1$.

■

B.3.2 Main Theorem 2

Main Theorem 2. *Let $X_1, \ldots, X_d$ be continuous random variables joint with a copula C_F that belongs to the maximum domain of attraction of an extreme value copula. The variables $X_1, \ldots, X_d$ are asymptotically independent if and only if their STDF-based multivariate tail dependence coefficient $\Lambda(C_F)$ is equal to zero. That is,*

$$C_F \in MDA(\Pi_d) \Leftrightarrow \Lambda(C_F) = 0.$$

Proof. If $C_F \in MDA(\Pi_d)$, the STDF is equal to $\ell(x_1, \ldots, x_d) = x_1 + \cdots + x_d$. Hence, in that case it follows immediately that

$$\Lambda(C_F) = \frac{d - \ell(1, \ldots, 1)}{d - 1} = \frac{d - d}{d - 1} = 0.$$

Conversely, we have to show that if $\Lambda(C_F) = 0$, then $C_F \in MDA(\Pi_d)$. By the relationship between the extreme value copula and the STDF (Proposition 3.5), it is sufficient to show that if $\Lambda(C_F) = 0$, then $\ell(\mathbf{x}) = x_1 + \cdots + x_d$ for any $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}_+^d$. Alternatively, by using the Pickands representation, it is sufficient to show that if $\Lambda(C_F) = 0$, then $A(\mathbf{x}) = 1$ for any $\mathbf{x} \in \Delta_d$. This is done as follows.

- First of all, note that if $\Lambda(C_F) = 0$, the relationship between the multivariate TDC Λ and the STDF implies that $\ell(1, \ldots, 1) = d$. Hence, for the Pickands dependence function we must have that

  $$A(1/d, \ldots, 1/d) = 1.$$

- Recall from Section 3.4.2 that the A-function is convex and is equal to 1 if it is evaluated in one of the unit vectors:

  $$A(\mathbf{e}_j) = 1,$$

  for $\mathbf{e}_j$ the j-th unit vector, $1 \leq j \leq d$. Since $A(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \Delta^d$, it is sufficient to show that $A(\mathbf{x}) \geq 1$ in order to prove that $A(\mathbf{x}) = 1$. The aim is therefore to show that $A(\mathbf{x}) \geq 1$ for all $\mathbf{x} \in \Delta_d$ given that $A(1/d, \ldots, 1/d) = 1$ and $A(\mathbf{e}_j) = 1$ for all $j = 1, \ldots, d$.

- The proof hinges on the convexity of the Pickands dependence function. Formally, the convexity of A implies that for all $\mathbf{x}, \mathbf{v} \in \Delta_d$, and $\alpha \in [0, 1]$, the following defining relationship holds,

  $$A\left(\alpha\mathbf{x} + (1 - \alpha)\mathbf{v}\right) \leq \alpha A(\mathbf{x}) + (1 - \alpha)A(\mathbf{v}).$$

  This implies that for any $\mathbf{x}, \mathbf{v} \in \Delta_d$ and $\alpha \in [0, 1]$,

  $$A(\mathbf{x}) \geq \frac{1}{\alpha}\left(A\left(\alpha\mathbf{x} + (1 - \alpha)\mathbf{v}\right) - (1 - \alpha)A(\mathbf{v})\right).$$

- In the bivariate case, we showed that if we could find a boundary point such that for any point $\mathbf{x} \in \Delta_2$, the mid point is a linear combination of the chosen point $\mathbf{x}$ and an unit vector, the A-function had to be greater than or equal to 1 in the point $\mathbf{x}$. Hence, trying to apply the same technique as for the bivariate proof, we would like to consider as a special case the above inequality with $\mathbf{v}$ and α such that $\alpha\mathbf{x} + (1 - \alpha)\mathbf{v} = (1/d, \ldots, 1/d)$ with $\mathbf{v} = \mathbf{e}_j$ for some $j \in \{1, \ldots, d\}$. However, for a general $\mathbf{x}$ it is impossible to always find a suitable unit vector such that the mid point, $(1/d, \ldots, 1/d)$ can be retrieved as a linear combination. Therefore, we need an extra step.

- For the 3-dimensional case the logic is as follows: we pick a point $\mathbf{x} \in \Delta_3$ and connect it to an unit vector $\mathbf{e}_j$, $j \in \{1, 2, 3\}$, *such that* the mid point $(1/3, 1/3, 1/3)$ is a linear combination of another unit vector $\mathbf{e}_k$, $k \in \{1, 2, 3\}$, and a point $\mathbf{y}$ on the line segment between $\mathbf{x}$ and $\mathbf{e}_j$. We will show below that it is always possible to pick unit vectors $\mathbf{e}_j$ and $\mathbf{e}_k$ such that this is true. Similar to the bivariate proof, the convexity of the Pickands dependence function yields that $A(\mathbf{y}) = 1$ in this setting. But then, applying the bivariate argument again, we also find that $A(\mathbf{x}) = 1$. The concept for the multivariate case is similar, but then we need more unit vectors to connect a general point $\mathbf{x} \in \Delta_d$ to the mid point $(1/d, \ldots, 1/d)$ with linear line segments. The argument explained here is made concrete with the following computations for $d = 3$ and for $d \geq 2$, in general.

d = 3

We first consider the proof for the trivariate case. For a given point $\mathbf{x} \in \Delta_3$, a coefficient $\alpha \in [0,1]$, and an index $j \in \{1,2,3\}$,

$$A(x_1,x_2,x_3) \geq \frac{1}{\alpha}\left(A\left(\alpha\mathbf{x} + (1-\alpha)\mathbf{e}_j\right) - (1-\alpha)A(\mathbf{e}_j)\right).$$

Now define $\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{e}_j$. Then $\mathbf{z} \in \Delta_3$. Therefore, for some $\beta \in [0,1]$ and some $k \in \{1,2,3\}$, we have

$$A(z_1,z_2,z_3) \geq \frac{1}{\beta}\left(A\left(\beta\mathbf{z} + (1-\beta)\mathbf{e}_k\right) - (1-\beta)A(\mathbf{e}_k)\right).$$

Substitution gives us the following expression

$$\begin{aligned}
A(x_1,x_2,x_3) &\geq \frac{1}{\alpha}\left(A\left(\mathbf{z}\right) - (1-\alpha)A(\mathbf{e}_j)\right) \\
&\geq \frac{1}{\alpha}\left(\frac{1}{\beta}\left(A\left(\beta\mathbf{z} + (1-\beta)\mathbf{e}_k\right) - (1-\beta)A(\mathbf{e}_k)\right) - (1-\alpha)A(\mathbf{e}_j)\right) \\
&= \frac{1}{\alpha}\left(\frac{1}{\beta}\left(A\left(\beta(\alpha\mathbf{x} + (1-\alpha)\mathbf{e}_j) + (1-\beta)\mathbf{e}_k\right) - (1-\beta)A(\mathbf{e}_k)\right) - (1-\alpha)A(\mathbf{e}_j)\right).
\end{aligned}$$

We now wish to find $\alpha, \beta \in [0,1]$ such that

$$\beta(\alpha\mathbf{x} + (1-\alpha)\mathbf{e}_j) + (1-\beta)\mathbf{e}_k = (1/3, 1/3, 1/3).$$

If we can show that this is in fact possible, it follows that

$$\begin{aligned}
A(x_1,x_2,x_3) &\geq \frac{1}{\alpha}\left(\frac{1}{\beta}\left(A\left(\beta(\alpha\mathbf{x} + (1-\alpha)\mathbf{e}_j) + (1-\beta)\mathbf{e}_k\right) - (1-\beta)A(\mathbf{e}_k)\right) - (1-\alpha)A(\mathbf{e}_j)\right) \\
&= \frac{1}{\alpha}\left(\frac{1}{\beta}\left(A\left(1/3,1/3,1/3\right) - (1-\beta)A(\mathbf{e}_k)\right) - (1-\alpha)A(\mathbf{e}_j)\right) \\
&= \frac{1}{\alpha}\left(\frac{1}{\beta}\left(1 - (1-\beta)\right) - (1-\alpha)\right) = \frac{1}{\alpha}(1-(1-\alpha)) = 1,
\end{aligned}$$

which was to be shown (for $d = 3$ at least). To establish the existence of such $\alpha, \beta \in [0,1]$, consider the following system of equations that needs to be solved,

$$\begin{cases}
\beta\alpha x_1 + \beta(1-\alpha)\mathbb{1}\{j=1\} + (1-\beta)\mathbb{1}\{k=1\} &= 1/3 \\
\beta\alpha x_2 + \beta(1-\alpha)\mathbb{1}\{j=2\} + (1-\beta)\mathbb{1}\{k=2\} &= 1/3 \\
\beta\alpha x_3 + \beta(1-\alpha)\mathbb{1}\{j=3\} + (1-\beta)\mathbb{1}\{k=3\} &= 1/3
\end{cases}$$

Without loss of generality, assume that $x_1 \geq x_2 \geq x_3$. In this case $x_1 \geq 1/3$ and $x_3 \leq 1/3$. Since $\alpha\beta \in [0,1]$, the last equation definitely needs an extra term coming from the unit vector. In contrast, α and β can be chosen in such a way that the first equation does not need an extra term coming from the unit vector. Therefore, we choose $j = 2$ and $k = 3$. This leads to the following system of equations:

$$\begin{cases}
\beta\alpha x_1 & & & = 1/3 \\
\beta\alpha x_2 & +\beta(1-\alpha) & & = 1/3 \\
\beta\alpha x_3 & & +(1-\beta) & = 1/3
\end{cases}$$

Solving the sytem yields

$$\alpha = \frac{x_1 + x_2 + x_3}{2x_1 + x_3}, \quad \beta = \frac{2x_1 + x_3}{3x_1}.$$

Note that $\alpha \in [0,1]$ because $x_1 \geq x_2$ and that $\beta \in [0,1]$ because $x_1 \geq x_3$. Also note that we do not lose generality because for a different ordering of the x's, j and k can be specified according to the same logic, leading to similar solutions and the same conclusion.

d ≥ 2

Now we consider the proof for a general $d \geq 2$. For a vector $\mathbf{x} = (x_1, \ldots, x_d) \in \Delta_d$, a coefficient $\alpha \in [0,1]$ and an index $j \in \{1, \ldots, d\}$,

$$A(x_1, \ldots, x_d) \geq \frac{1}{\alpha} \left(A\left(\alpha \mathbf{x} + (1-\alpha)\mathbf{e}_j\right) - (1-\alpha)A(\mathbf{e}_j) \right).$$

Performing the same trick as for the trivariate case, we find that for some $\alpha_1, \ldots, \alpha_{d-1} \in [0,1]$ and for $j_1, \ldots, j_{d-1} \in \{1, \ldots, d\}$,

$$A(\mathbf{x}) \geq \frac{1}{\alpha_1}\left(\frac{1}{\alpha_2}\left(\ldots \frac{1}{\alpha_{d-2}}\left(\frac{1}{\alpha_{d-1}}\Big(A\big(\alpha_{d-1}\mathbf{y}^{d-2} + (1-\alpha_{d-1})\mathbf{e}_{j_{d-1}}\big) - (1-\alpha_{d-1})A(\mathbf{e}_{j_{d-1}})\Big) - (1-\alpha_{d-2})A(\mathbf{e}_{j_{d-2}})\right) \cdots - (1-\alpha_2)A(\mathbf{e}_{j_2})\right) - (1-\alpha_1)A(\mathbf{e}_{j_1})\right),$$

where $\mathbf{y}^{d-2} = \alpha_{d-2}\mathbf{y}^{d-3} + (1-\alpha_{d-2})\mathbf{e}_{j_{d-2}}$ is defined recursively with starting value $\mathbf{y}^0 = \mathbf{x}$. Again, we aim to find $\alpha_1, \ldots, \alpha_{d-1} \in [0,1]$ such that

$$\alpha_{d-1}\mathbf{y}^{d-2} + (1-\alpha_{d-1})\mathbf{e}_{j_{d-1}} = (1/d, \ldots, 1/d).$$

If we can show that this is in fact possible, it follows that

$$\begin{aligned}
A(\mathbf{x}) &\geq \frac{1}{\alpha_1}\left(\frac{1}{\alpha_2}\left(\ldots \frac{1}{\alpha_{d-2}}\left(\frac{1}{\alpha_{d-1}}\Big(A\big(\alpha_{d-1}\mathbf{y}^{d-2} + (1-\alpha_{d-1})\mathbf{e}_{j_{d-1}}\big) - (1-\alpha_{d-1})A(\mathbf{e}_{j_{d-1}})\Big) - (1-\alpha_{d-2})A(\mathbf{e}_{j_{d-2}})\right) \cdots - (1-\alpha_2)A(\mathbf{e}_{j_2})\right) - (1-\alpha_1)A(\mathbf{e}_{j_1})\right) \\
&= \frac{1}{\alpha_1}\left(\frac{1}{\alpha_2}\left(\ldots \frac{1}{\alpha_{d-2}}\left(\frac{1}{\alpha_{d-1}}\Big(A\big(1/d, \ldots, 1/d\big) - (1-\alpha_{d-1})A(\mathbf{e}_{j_{d-1}})\Big) - (1-\alpha_{d-2})A(\mathbf{e}_{j_{d-2}})\right) \cdots - (1-\alpha_2)A(\mathbf{e}_{j_2})\right) - (1-\alpha_1)A(\mathbf{e}_{j_1})\right) \\
&= \frac{1}{\alpha_1}\left(\frac{1}{\alpha_2}\left(\ldots \frac{1}{\alpha_{d-2}}\left(\frac{1}{\alpha_{d-1}}\Big(1 - (1-\alpha_{d-1})\Big) - (1-\alpha_{d-2})\right) \cdots - (1-\alpha_2)\right) - (1-\alpha_1)\right) \\
&= \frac{1}{\alpha_1}\left(\frac{1}{\alpha_2}\left(\ldots \frac{1}{\alpha_{d-2}}\left(1 - (1-\alpha_{d-2})\right) \cdots - (1-\alpha_2)\right) - (1-\alpha_1)\right) = \cdots = 1,
\end{aligned}$$

which gives the desired result for a general $d \geq 2$. To establish the existence of such $\alpha_1, \ldots, \alpha_{d-1} \in [0,1]$, consider the following system of equations that needs to be solved,

$$\begin{cases}
\alpha_{d-1}\alpha_{d-2}\ldots\alpha_2\alpha_1 x_1 + \alpha_{d-1}\alpha_{d-2}\ldots\alpha_2(1-\alpha_1)\mathbb{1}\{j_1 = 1\} + \alpha_{d-1}\alpha_{d-2}\ldots\alpha_3(1-\alpha_2)\mathbb{1}\{j_2 = 1\} \\
\qquad + \cdots + \alpha_{d-1}(1-\alpha_{d-2})\mathbb{1}\{j_{d-2} = 1\} + (1-\alpha_{d=1})\mathbb{1}\{j_{d-1} = 1\} = 1/d \\
\ldots \\
\alpha_{d-1}\alpha_{d-2}\ldots\alpha_2\alpha_1 x_d + \alpha_{d-1}\alpha_{d-2}\ldots\alpha_2(1-\alpha_1)\mathbb{1}\{j_1 = d\} + \alpha_{d-1}\alpha_{d-2}\ldots\alpha_3(1-\alpha_2)\mathbb{1}\{j_2 = d\} \\
\qquad + \cdots + \alpha_{d-1}(1-\alpha_{d-2})\mathbb{1}\{j_{d-2} = d\} + (1-\alpha_{d=1})\mathbb{1}\{j_{d-1} = d\} = 1/d
\end{cases}$$

Without loss of generality, assume that $x_1 \geq x_2 \geq \cdots \geq x_d$. In this case $x_1 \geq 1/d$ and $x_d \leq 1/d$. Since $\alpha_1 \ldots \alpha_{d-1} \in [0,1]$, the last equation definitely needs an extra term coming from the unit vectors. In contrast, $\alpha_1, \ldots, \alpha_{d-1}$ can be chosen in such a way that the first equation does not need an extra term coming from the unit vectors. Therefore, we choose $j_i = i+1$ for $1 \leq i \leq d-1$. This leads to the following system of equations,

$$\begin{cases}
\alpha_{d-1}\alpha_{d-2}\ldots\alpha_2\alpha_1 x_1 & & = 1/d \\
\alpha_{d-1}\alpha_{d-2}\ldots\alpha_2\alpha_1 x_2 & + \alpha_{d-1}\alpha_{d-2}\ldots\alpha_2(1-\alpha_1) & = 1/d \\
\ldots & & \\
\alpha_{d-1}\alpha_{d-2}\ldots\alpha_2\alpha_1 x_{d-1} & + \alpha_{d-1}(1-\alpha_{d-2}) & = 1/d \\
\alpha_{d-1}\alpha_{d-2}\ldots\alpha_2\alpha_1 x_d & + (1-\alpha_{d-1}) & = 1/d
\end{cases}$$

Solving the system from the bottom up yields the following expressions for $\alpha_1, \ldots, \alpha_{d-1}$,

$$\begin{array}{lll}
1-\alpha_{d-1} &= \frac{1}{d} - \alpha_{d-1}\ldots\alpha_1 x_d & = \frac{x_1-x_d}{dx_1} \quad \Rightarrow \alpha_{d-1} = \frac{dx_1-(x_1-x_d)}{dx_1} \\
1-\alpha_{d-2} &= \frac{1}{\alpha_{d-1}}\left(\frac{1}{d} - \alpha_{d-1}\ldots\alpha_1 x_{d-1}\right) & = \frac{1}{\alpha_{d-1}}\left(\frac{x_1-x_{d-1}}{dx_1}\right) \quad \Rightarrow \alpha_{d-2} = \frac{\alpha_{d-1}dx_1-(x_1-x_{d-1})}{\alpha_{d-1}dx_1} \\
\ldots & & \\
1-\alpha_2 &= \frac{1}{\alpha_{d-1}\ldots\alpha_3}\left(\frac{1}{d} - \alpha_{d-1}\ldots\alpha_1 x_3\right) & = \frac{1}{\alpha_{d-1}\ldots\alpha_3}\left(\frac{x_1-x_3}{dx_1}\right) \quad \Rightarrow \alpha_2 = \frac{\alpha_{d-1}\ldots\alpha_3 dx_1-(x_1-x_3)}{\alpha_{d-1}\ldots\alpha_3 dx_1} \\
1-\alpha_1 &= \frac{1}{\alpha_{d-1}\ldots\alpha_2}\left(\frac{1}{d} - \alpha_{d-1}\ldots\alpha_1 x_2\right) & = \frac{1}{\alpha_{d-1}\ldots\alpha_2}\left(\frac{x_1-x_2}{dx_1}\right) \quad \Rightarrow \alpha_1 = \frac{\alpha_{d-1}\ldots\alpha_2 dx_1-(x_1-x_2)}{\alpha_{d-1}\ldots\alpha_2 dx_1}
\end{array}$$

Since $x_1 \geq \cdots \geq x_d$, it is obvious that all $\alpha_1, \ldots, \alpha_{d-1} \leq 1$. To see that $\alpha_1, \ldots, \alpha_{d-1} \geq 0$, first observe that

$$\frac{x_1 - x_d}{dx_1} \leq \frac{1}{d} \qquad \Rightarrow \qquad \alpha_{d-1} = 1 - \frac{x_1 - x_d}{dx_1} \geq \frac{d-1}{d},$$

since $1/d \leq x_1 \leq 1$ and $0 \leq x_d \leq 1/d$. Applying this inequality to the next equation,

$$\frac{1}{\alpha_{d-1}}\left(\frac{x_1 - x_{d-1}}{dx_1}\right) \leq \frac{1}{\alpha_{d-1}} \cdot \frac{1}{d} \leq \frac{d}{d-1} \cdot \frac{1}{d} = \frac{1}{d-1} \Rightarrow \alpha_{d-2} = 1 - \frac{1}{\alpha_{d-1}}\left(\frac{x_1 - x_{d-1}}{dx_1}\right) \geq \frac{d-2}{d-1}.$$

Repeating these steps recursively yields that

$$\begin{aligned}
\frac{1}{\alpha_{d-1}\alpha_{d-2}}\left(\frac{x_1 - x_{d-2}}{dx_1}\right) \leq \frac{1}{\alpha_{d-1}\alpha_{d-2}} \cdot \frac{1}{d} \leq \frac{d}{d-1} \cdot \frac{d-1}{d-2} \cdot \frac{1}{d} = \frac{1}{d-2} \\
\Rightarrow \alpha_{d-3} = 1 - \frac{1}{\alpha_{d-1}\alpha_{d-2}}\left(\frac{x_1 - x_{d-2}}{dx_1}\right) \geq \frac{d-3}{d-2}
\end{aligned}$$

$$\ldots$$

$$\begin{aligned}
\frac{1}{\alpha_{d-1}\ldots\alpha_3}\left(\frac{x_1 - x_3}{dx_1}\right) \leq \frac{1}{\alpha_{d-1}\ldots\alpha_3} \cdot \frac{1}{d} \leq \frac{d}{d-1} \cdot \frac{d-1}{d-2} \cdots \frac{4}{3} \cdot \frac{1}{d} = \frac{1}{3} \\
\Rightarrow \alpha_2 = 1 - \frac{1}{\alpha_{d-1}\ldots\alpha_3}\left(\frac{x_1 - x_3}{dx_1}\right) \geq \frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\alpha_{d-1}\ldots\alpha_2}\left(\frac{x_1 - x_2}{dx_1}\right) \leq \frac{1}{\alpha_{d-1}\ldots\alpha_2} \cdot \frac{1}{d} \leq \frac{d}{d-1} \cdot \frac{d-1}{d-2} \cdots \frac{3}{2} \cdot \frac{1}{d} = \frac{1}{2} \\
\Rightarrow \alpha_1 = 1 - \frac{1}{\alpha_{d-1}\ldots\alpha_2}\left(\frac{x_1 - x_2}{dx_1}\right) \geq \frac{1}{2},
\end{aligned}$$

showing that for all $\alpha_1, \ldots, \alpha_{d-1}$ the required bounds hold. Therefore, for any $\mathbf{x} \in \Delta_d$, it holds that $1 \leq A(\mathbf{x}) \leq 1$, indicating that $A(\mathbf{x}) = 1$. This is what needed to be shown. ■

B.4 Asymptotical behavior of STDF estimators

B.4.1 Asymptotics of the empirical STDF

Theorem 5.1. *(Einmahl et al., 2012). Assuming that*

1. $\lim_{t\downarrow 0} t^{-1}\left(1 - C_F(1-tx_1,\ldots,1-tx_d)\right)$ *exists and converges uniformly to* $\ell(x_1,\ldots,x_d)$ *on* $[0,T]^d$ *for* $T>0$ *(first order condition);*
2. $t^{-1}\left(1 - C_F(1-tx_1,\ldots,1-tx_d)\right) - \ell(\mathbf{x}) = \mathcal{O}\left(t^{\alpha}\right)$*, uniformly in* $\mathbf{x}\in[0,1]^d$ *as* $t\downarrow 0$*, for some* $\alpha>0$ *(second order condition);*
3. $k = \mathcal{O}\left(n^{2\alpha/(1+2\alpha)}\right)$ *for the positive number* α *used in the assumption above, and* $k\to\infty$ *as* $n\to\infty$*;*
4. *for all* $j = 1,...,d$*, the first-order partial derivative of* ℓ *with respect to* x_j *exists and is continuous on the set of points* $\mathbf{x}$ *such that* $x_j>0$*;*

we have

$$\sup_{x\in[0,T]^d}\left|\sqrt{k}\left(\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right) - B_\ell(\mathbf{x})\right| \to 0,$$

for $T>0$ *as* $n\to\infty$*. Here* $B_\ell(\mathbf{x})$ *is a zero-mean Gaussian process defined in Equation 5.1.5.*

Proof. We produce an outline of the proof and refer to the original proof for the multivariate case in Einmahl et al. (2012) for details. Also see Theorem 7.2.2 in de Haan and Ferreira (2006) for a detailed proof in the bivariate case. The idea of the proof is to capture the distributions of the differences between the theoretical ℓ-function and the respective approximations made in the process to retrieve the empirical STDF $\hat{\ell}_{n,k}$. To formalize this idea, let

$$\mathbf{U}_i = (U_{i1},\ldots,U_{id}) = (1-F_1(X_{i1}),\ldots,1-F_d(X_{id}))$$

for $i=1,\ldots,n$, denote by $U_{1:n,j}\le\cdots\le U_{n:n,j}$ the order statistics of $U_{1j},\ldots,U_{nj}$, $j=1,\ldots,d$, and denote by $\lceil a\rceil$ the smallest integer larger than or equal to a. Define

$$\begin{aligned}
S_n(\mathbf{x}) &= \left(\frac{n}{k}U_{\lceil kx_1\rceil:n,1},\ldots,\frac{n}{k}U_{\lceil kx_d\rceil:n,d}\right),\\
V_n(\mathbf{x}) &= \frac{n}{k}\mathbb{P}\left(F_1(X_1) > 1-\frac{kx_1}{n} \text{ or } \ldots \text{ or } F_d(X_d) > 1-\frac{kx_d}{n}\right)\\
&= \frac{n}{k}\mathbb{P}\left(U_1 \le \frac{kx_1}{n} \text{ or } \ldots \text{ or } U_d \le \frac{kx_d}{n}\right),\\
T_n(\mathbf{x}) &= \frac{1}{k}\sum_{i=1}^{n}\mathbb{1}\left\{F_1(X_{i1}) > 1-\frac{kx_1}{n} \text{ or } \ldots \text{ or } F_d(X_{id}) > 1-\frac{kx_d}{n}\right\}\\
&= \frac{1}{k}\sum_{i=1}^{n}\mathbb{1}\left\{U_{i1} \le \frac{kx_1}{n} \text{ or } \ldots \text{ or } U_{id} \le \frac{kx_d}{n}\right\},
\end{aligned}$$

with $\mathbf{x}=(x_1,\ldots,x_d)\in[0,1]^d$. Note that $S_n(\mathbf{x})$ contains the approximate high quantile levels based on the marginal empirical distribution functions (rank-based) and that $\hat{\ell}_{n,k}(\mathbf{x}) = T_n(S_n(\mathbf{x}))$. Now we can write,

$$\begin{aligned}
\sqrt{k}\left(\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right) &= \sqrt{k}\left(T_n\left(S_n(\mathbf{x})\right) - V_n\left(S_n(\mathbf{x})\right)\right) && (=: D_1(\mathbf{x}))\\
&\quad + \sqrt{k}\left(V_n\left(S_n(\mathbf{x})\right) - \ell\left(S_n(\mathbf{x})\right)\right) && (=: D_2(\mathbf{x}))\\
&\quad + \sqrt{k}\left(\ell\left(S_n(\mathbf{x})\right) - \ell(\mathbf{x})\right). && (=: D_3(\mathbf{x}))
\end{aligned}$$

Here D_1 represents the part introduced by the approximation of the multivariate probability (i.e., the dependence part), D_2 represents the part introduced by the approximation of the limiting value $t\downarrow 0$ by the finite value k/n, and D_3 represents the part introduced by the approximation of the marginal distributions with the ranked data (i.e., the marginal part).

It is a known result that for $T > 0$,

$$\sup_{\mathbf{x}\in[0,2T]^d} \left|\sqrt{k}\left(T_n(\mathbf{x}) - V_n(\mathbf{x})\right) - W_\ell(\mathbf{x})\right| \to 0,$$

see Einmahl et al. (1997), Einmahl et al. (2012) or Prop. 7.2.3 in de Haan and Ferreira (2006). Because of the uniform continuity of W_ℓ, this can be used to show that

$$\sup_{\mathbf{x}\in[0,T]^d} |D_1(\mathbf{x}) - W_\ell(\mathbf{x})| \to 0.$$

The D_2-term represents the error made from the approximation of the limiting value with a finite value. Because of the assumptions 2 and 3, it can be shown that

$$\sup_{\mathbf{y}\in[0,2T]^d} \sqrt{k}\,|V_n(\mathbf{y}) - \ell(\mathbf{y})| = \sqrt{k}\mathcal{O}\left(\left(\frac{k}{n}\right)^\alpha\right) = \mathcal{O}\left(\left(\frac{k}{n^{2\alpha/(1+2\alpha)}}\right)^{1/2+\alpha}\right) = \mathcal{O}(1).$$

and, hence, that

$$\sup_{x\in[0,T]^d} |D_2(\mathbf{x})| \to 0.$$

Finally, the Vervaat (1972) lemma can be used similarly to the proof of de Haan and Ferreira (2006) (Thm. 7.2.2) for the bivariate case to ensure the marginal convergences for $j = 1, \ldots, d$,

$$\sup_{\mathbf{x}\in[0,T]} \left|\sqrt{k}\left(\frac{n}{k}U_{\lceil kx_j\rceil:n,j} - x_j\right) + W_\ell(x_j\mathbf{e}_j)\right| \to 0 \quad \text{a.s.},$$

which, combined with an application of the mean value theorem yields the result that

$$\sup_{\mathbf{x}\in[0,T]^d} \left|D_3(\mathbf{x}) + \sum_{j=1}^d \ell_j(\mathbf{x})W_\ell(x_j\mathbf{e}_j)\right| \to 0 \quad \text{a.s.}.$$

Putting everything together, the result is retrieved:

$$\sup_{\mathbf{x}\in[0,T]^d} \left|\sqrt{k}\left(\hat{\ell}_{n,k}(\mathbf{x}) - \ell(\mathbf{x})\right) - \left(W_\ell(\mathbf{x}) - \sum_{j=1}^d \ell_j(\mathbf{x})W_\ell(x_j\mathbf{e}_j)\right)\right| \to 0 \quad \text{a.s.}.$$

■

B.4.2 Asymptotics of the adjusted empirical STDF

Theorem 5.3. *Under the null hypothesis of tail independence and with the conditions of Theorem 5.1, supplemented with the conditions that $m/n \to \theta > 0$ and $m/n - \theta = \mathcal{O}(n^{1/2})$, it holds that*

$$\left\{\sqrt{k}\left(\overline{\ell}_{n,k}(x_1,\ldots,x_d) - (x_1+\cdots+x_d)\right), \mathbf{x} \in [0,1]^d\right\} \to \sum_{j=1}^{d} W_j\left((1+\theta)x_j\right) =: \overline{B}(\mathbf{x}) \qquad (5.2.2)$$

as $n \to \infty$, where $W_j, j = 1,\ldots,d$, are d independent Brownian motions.

Proof. The proof can be derived similarly to the proof of Theorem 5.1. Note that this is a different approach than taken by Hüsler and Li (2009) to prove the result for the bivariate case. Let

$$\tilde{\mathbf{U}}_i = (\tilde{U}_{i1},\ldots,\tilde{U}_{id}) = \left(1 - F_1(\tilde{X}_{i1}),\ldots,1 - F_d(\tilde{X}_{id})\right),$$

for $i = 1,\ldots,m$, denote by $\tilde{U}_{1:m,j} \le \cdots \le \tilde{U}_{m:m,j}$ the order statistics of $\tilde{U}_{1j},\ldots,\tilde{U}_{mj}$, $j = 1,\ldots,d$, and denote by $\lceil a \rceil$ the smallest integer larger than or equal to a. Define

$$\begin{aligned}
\tilde{S}_m(\mathbf{x}) &= \left(\frac{m}{k}\tilde{U}_{\lceil kx_1\rceil:m,1},\ldots,\frac{m}{k}\tilde{U}_{\lceil kx_1\rceil:m,1}\right),\\
V_m(\mathbf{x}) &= \frac{m}{k}\mathbb{P}\left(U_1 \le \frac{kx_1}{m} \text{ or } \ldots \text{ or } U_d \le \frac{kx_d}{m}\right)\\
&= \theta\frac{n}{k}\mathbb{P}\left(U_1 \le \frac{kx_1}{m} \text{ or } \ldots \text{ or } U_d \le \frac{kx_d}{m}\right),\\
T_n(\mathbf{x}) &= \frac{m}{k}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left\{U_{i1} \le \frac{kx_1}{m} \text{ or } \ldots \text{ or } U_{id} \le \frac{kx_d}{m}\right\}\\
&= \theta\frac{1}{k}\sum_{i=1}^{n}\mathbb{1}\left\{U_{i1} \le \frac{kx_1}{m} \text{ or } \ldots \text{ or } U_{id} \le \frac{kx_d}{m}\right\},
\end{aligned}$$

with $\mathbf{x} = (x_1,\ldots,x_d) \in [0,1]^d$. Note that $\tilde{S}_m(\mathbf{x})$ contains the approximate high quantile levels based on the ranks of $\tilde{X}_{1j},\ldots,\tilde{X}_{mj}$, $j = 1,\ldots,d$, and that $\overline{\ell}_{n,k}(\mathbf{x}) = T_n(\tilde{S}_m(\mathbf{x}))$. Now, similarly to the original empirical STDF, the following decomposition can be made,

$$\begin{aligned}
\sqrt{k}\left(\overline{\ell}_{n,k}(\mathbf{x}) - ||\mathbf{x}||_1\right) &= \sqrt{k}\left(T_n(\tilde{S}_m(\mathbf{x})) - V_m(\tilde{S}_m(\mathbf{x}))\right), && (=: D_1(\mathbf{x}))\\
&+ \sqrt{k}\left(V_m(\tilde{S}_m(\mathbf{x})) - ||\tilde{S}_m(\mathbf{x})||_1\right), && (=: D_2(\mathbf{x}))\\
&+ \sqrt{k}\left(||\tilde{S}_m(\mathbf{x})||_1 - ||\mathbf{x}||_1\right). && (=: D_3(\mathbf{x}))
\end{aligned}$$

Here, it is used that under the null hypothesis of tail independence, the STDF satisfies $\ell(\mathbf{x}) = x_1 + \cdots + x_d = ||\mathbf{x}||_1$. As in the proof of Theorem 5.1, the D_1 term represents the part of the distribution introduced by the approximation of the multivariate probability (i.e., the dependence part), D_2 represents the part introduced by the approximation of the limiting value $t \downarrow 0$ by the finite value k/n for the probability and by the finite value k/m for the high thresholds, and D_3 represents the part introduced by the approximation of the marginal distributions with the ranked data (i.e., the marginal part). From the proof of Theorem 5.1, we know that for $T > 0$,

$$\sup_{\mathbf{x}\in[0,2T]^d}\left|\sqrt{k}\left(\frac{1}{\theta}T_n(\mathbf{x}) - \frac{1}{\theta}V_m(\mathbf{x})\right) - W_\ell(\mathbf{x})\right| \to 0,$$

which can be used to show that

$$\sup_{\mathbf{x}\in[0,T]^d}|D_1(\mathbf{x}) - \theta W_\ell(\mathbf{x})| \to 0.$$

Under the assumption of tail independence, this is equivalent to

$$\sup_{\mathbf{x}\in[0,T]^d}\left|D_1(\mathbf{x}) - \theta\sum_{j=1}^{d}W_{\ell,j}(x_j)\right| \to 0,$$

with $W_{\ell,j}$, $j = 1, \ldots, d$, the marginal processes defined in Equation 5.1.4. The D_2-term represents the error made from the approximation of the limiting value t with a finite value. Because of assumption 2 and 3 in Theorem 5.1, a similar reasoning can be used to show that

$$\sup_{\mathbf{x} \in [0,T]^d} |D_2(\mathbf{x})| \to 0.$$

For the final D_3-term, we use the proof of Theorem 5.1 to find that

$$\sup_{\mathbf{x} \in [0,T]^d} \left| D_3(\mathbf{x}) + \sum_{j=1}^{d} \ell_j(\mathbf{x}) \tilde{W}_\ell(x_j \mathbf{e}_j) \right| \to 0,$$

where $\tilde{W}_\ell$ denotes a Brownian motion based on the data $(\tilde{X}_{11}, \ldots, \tilde{X}_{1d}), \ldots, (\tilde{X}_{m1}, \ldots, \tilde{X}_{md})$ independent of the Brownian motion W_ℓ introduced above. Note that the processes $\tilde{W}_\ell(x_j \mathbf{e}_j)$ are equal to its marginal processes, $\tilde{W}_{\ell,j}$, defined in Equation 5.1.4, for $j = 1, \ldots, d$. Moreover, since the first partial derivatives of the ℓ-function are all equal to one under the null hypothesis of tail independence, the above expression can be simplified to

$$\sup_{\mathbf{x} \in [0,T]^d} \left| D_3(\mathbf{x}) + \sum_{j=1}^{d} \tilde{W}_{\ell,j}(x_j) \right| \to 0.$$

Putting everything together, we find that,

$$\sqrt{k} \left(\overline{\ell}_{n,k}(\mathbf{x}) - ||\mathbf{x}||_1 \right) = D_1(\mathbf{x}) + D_2(\mathbf{x}) + D_3(\mathbf{x}) \to \theta \sum_{j=1}^{d} W_{\ell,j}(x_j) - \sum_{j=1}^{d} \tilde{W}_{\ell,j}(x_j).$$

Since the Brownian motions W_ℓ and $\tilde{W}_\ell$ are based on independent datasets, they are independent. Hence, variances of the Brownian motions can be added to retrieve the final result:

$$\sqrt{k} \left(\overline{\ell}_{n,k}(\mathbf{x}) - ||\mathbf{x}||_1 \right) \to \sum_{j=1}^{d} \overline{W}_j((1+\theta)x_j),$$

where $\overline{W}_j$, $j = 1, \ldots, d$, are d independent Brownian motions. ■

B.4.3 Asymptotics of the bias-corrected adjusted empirical STDF

Proof. According to Theorem 5.6,

$$\overline{\ell}_k(\mathbf{x}) = \ell(\mathbf{x}) + \alpha\left(\frac{n}{k}\right) M(\mathbf{x}) + \frac{1}{\sqrt{k}}\overline{B}_\ell(\mathbf{x}) + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).$$

Therefore, the following decomposition holds,

$$\begin{aligned}
\sqrt{k}\left(\overline{\ell}^{BC}_{k,\overline{k}}(\mathbf{x}) - \ell(\mathbf{x})\right) &= \sqrt{k}\left(\frac{\tilde{\ell}_k(\mathbf{x}) - \left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \tilde{\alpha}_{\overline{k}}(\mathbf{x}) \frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} - \ell(\mathbf{x})\right) \\
&= \sqrt{k}\frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-1}\overline{\ell}_k(a_j\mathbf{x})}{\frac{1}{k}\sum_{j=1}^k K(a_j)} \\
&\quad - \sqrt{k}\left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \tilde{\alpha}_{\overline{k}}(\mathbf{x}) \frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} - \ell(\mathbf{x}) \\
&= \sqrt{k}\frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-1}\left(\ell(a_j\mathbf{x}) + \alpha\left(\frac{n}{k}\right) M(a_j\mathbf{x}) + \frac{1}{\sqrt{k}}\overline{B}_\ell(a_j\mathbf{x}) + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right)\right)}{\frac{1}{k}\sum_{j=1}^k K(a_j)} \\
&\quad - \sqrt{k}\left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \tilde{\alpha}_{\overline{k}}(\mathbf{x}) \frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} - \ell(\mathbf{x}) \\
&= \sqrt{k}\frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-1}\left(a_j\ell(\mathbf{x}) + a_j^{1-\rho}\alpha\left(\frac{n}{k}\right) M(\mathbf{x}) + \frac{1}{\sqrt{k}}\sqrt{a_j}\overline{B}_\ell(\mathbf{x})\right)}{\frac{1}{k}\sum_{j=1}^k K(a_j)} + \mathcal{O}(1) \\
&\quad - \sqrt{k}\left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \tilde{\alpha}_{\overline{k}}(\mathbf{x}) \frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} - \ell(\mathbf{x}) \\
&= \sqrt{k}\left(\alpha\left(\frac{n}{k}\right) M(\mathbf{x}) - \left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \tilde{\alpha}_{\overline{k}}(\mathbf{x})\right) \frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} \\
&\quad + \sqrt{k}\left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \tilde{\alpha}_{\overline{k}}(\mathbf{x}) \frac{\frac{1}{k}\sum_{j=1}^k K(a_j)\left(a_j^{-\rho} - a_j^{-\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)}\right)}{\frac{1}{k}\sum_{j=1}^k K(a_j)} \\
&\quad + \left(\frac{\int_0^1 K(u)u^{-1/2}\mathrm{d}u}{\frac{1}{k}\sum_{j=1}^k K(a_j)}\right)\overline{B}_\ell(\mathbf{x}) + \mathcal{O}(1).
\end{aligned}$$

Now, by Proposition B.3 (see below) and under the assumptions of Theorem 5.2, it follows that

$$\begin{aligned}
\sqrt{k}\left(\overline{\ell}^{BC}_{k,\overline{k}}(\mathbf{x}) - \ell(\mathbf{x})\right) &= \sqrt{k}\left(\alpha\left(\frac{n}{k}\right) M(\mathbf{x}) - \left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\right) \frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} \\
&\quad + \left(\frac{\int_0^1 K(u)u^{-1/2}\mathrm{d}u}{\frac{1}{k}\sum_{j=1}^k K(a_j)}\right)\overline{B}_\ell(\mathbf{x}) + \mathcal{O}(1).
\end{aligned}$$

Since the function α is assumed to be a regularly varying function with index $\rho < 0$, we can

write $\alpha(t) = t^\rho \mathscr{L}_\alpha(t)$ where $\mathscr{L}_\alpha$ is slowly varying at infinity. Hence,

$$\begin{aligned}
\sqrt{k}\left(\overline{\ell}^{BC}_{k,\overline{k}}(\mathbf{x}) - \ell(\mathbf{x})\right) &= \sqrt{\frac{k}{\overline{k}}}\sqrt{\overline{k}}\alpha\left(\frac{n}{\overline{k}}\right) M(\mathbf{x}) \left[\left(\frac{\overline{k}}{k}\right)^{\rho} - \left(\frac{\overline{k}}{k}\right)^{\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)}\right] \frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} \\
&\quad + \sqrt{\frac{k}{\overline{k}}}\sqrt{\overline{k}}\alpha\left(\frac{n}{\overline{k}}\right)\left(\frac{\overline{k}}{k}\right)^{\rho} M(\mathbf{x}) \left(\frac{\mathscr{L}_\alpha\left(\frac{n}{k}\right)}{\mathscr{L}_\alpha\left(\frac{n}{\overline{k}}\right)} - 1\right) \frac{\frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^k K(a_j)} \\
&\quad + \left(\frac{\int_0^1 K(u) u^{-1/2}\mathrm{d}u}{\frac{1}{k}\sum_{j=1}^k K(a_j)}\right) \overline{B}_\ell(\mathbf{x}) + \mathcal{O}(1) \\
&= \left(\frac{\int_0^1 K(u) u^{-1/2}\mathrm{d}u}{\frac{1}{k}\sum_{j=1}^k K(a_j)}\right) \overline{B}_\ell(\mathbf{x}) + \mathcal{O}(1) \\
&\quad + \mathcal{O}\left(\left(\frac{k}{\overline{k}}\right)^{1/2-\tilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \log\left(\frac{k}{\overline{k}}\right) \sqrt{\overline{k}}\alpha\left(\frac{n}{\overline{k}}\right)\left(\tilde{\rho}_{\overline{k}}(\mathbf{x}^*) - \rho\right)\right) \\
&\quad + \left(\frac{k}{\overline{k}}\right)^{1/2-\rho} \sqrt{\overline{k}}\alpha\left(\frac{n}{\overline{k}}\right) \beta\left(\frac{n}{\overline{k}}\right) \frac{1}{\beta\left(\frac{n}{\overline{k}}\right)} \left(\frac{\mathscr{L}_\alpha\left(\frac{n}{k}\right)}{\mathscr{L}_\alpha\left(\frac{n}{\overline{k}}\right)} - 1\right) \mathcal{O}(1)
\end{aligned}$$

where we used the mean value theorem. The theorem now follows under the assumptions since

$$\lim_{t\to\infty} \frac{1}{\beta(t)}\left(\frac{\mathscr{L}_\alpha(tc)}{\mathscr{L}_\alpha(t)} - 1\right) = \mathcal{O}(1).$$

To see that this is in fact the case, we refer to the proof presented in Beirlant et al. (2016), since this part of the proof only depends on the properties and assumptions of the true stable tail dependence function ℓ. ■

Proposition B.2. *For any fixed d-vector* $\mathbf{x}^*$*, under the assumptions of Theorem 5.2,*

$$\sqrt{k}\alpha\left(\frac{n}{k}\right)\left(\tilde{\rho}_k(\mathbf{x}^*) - \rho\right) \to \frac{1 - r^{\rho-1/2}}{\log r} \frac{\overline{B}(\mathbf{x}^*)}{M(\mathbf{x}^*)} \frac{\int_0^1 K(u) u^{-1/2}\mathrm{d}u}{\int_0^1 K(u) u^{-\rho}\mathrm{d}u} \frac{a^{-1/2} - 1}{a^{-\rho} - 1},$$

where

$$\tilde{\rho}_k(\mathbf{x}^*) = \left(1 - \frac{1}{\log r} \log\left|\frac{\Delta_{k,a}(r\mathbf{x}^*)}{\Delta_{k,a}(\mathbf{x}^*)}\right|\right) \wedge 0$$

and

$$\Delta_{k,a}(\mathbf{x}) = a^{-1}\tilde{\ell}_k(a\mathbf{x}) - \tilde{\ell}_k(\mathbf{x}), \quad \tilde{\ell}_k(\mathbf{x}) = \frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-1} \overline{\ell}_k(a_j\mathbf{x})$$

Proof. Using Theorem 5.6, the homogeneity properties of ℓ and M, and the fact that $\overline{B}_\ell(a\mathbf{x}) \sim \sqrt{a}\overline{B}(\mathbf{x})$, it follows that

$$\begin{aligned}
\tilde{\ell}_k(\mathbf{x}^*) &= \frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-1} \overline{\ell}_k(a_j\mathbf{x}^*) \\
&= \frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-1}\left(\ell(a_j\mathbf{x}^*) + \alpha\left(\frac{n}{k}\right) M(a_j\mathbf{x}^*)\right) \\
&\qquad + \frac{1}{\sqrt{k}}\left(\int_0^1 K(u) u^{-1/2}\mathrm{d}u\right) \overline{B}_\ell(\mathbf{x}^*) + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) \\
&= \frac{1}{k}\sum_{j=1}^k K(a_j)\ell(\mathbf{x}^*) + \alpha\left(\frac{n}{k}\right) M(\mathbf{x}^*) \frac{1}{k}\sum_{j=1}^k K(a_j) a_j^{-\rho} \\
&\qquad + \frac{1}{\sqrt{k}}\left(\int_0^1 K(u) u^{-1/2}\mathrm{d}u\right) \overline{B}_\ell(\mathbf{x}^*) + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}\Delta_{k,a}(\mathbf{x}^*) &= a^{-1}\tilde{\ell}_k(a\mathbf{x}^*) - \tilde{\ell}_k(\mathbf{x}^*) \\ &= \alpha\left(\frac{n}{k}\right) M(\mathbf{x}^*)\frac{1}{k}\sum_{j=1}^{k} K(a_j)a_j^{-\rho}\left(a^{-rho}-1\right) \\ &\qquad + \frac{1}{\sqrt{k}}\left(\int_0^1 K(u)u^{-1/2}\mathrm{d}u\right)\left(a^{-1/2}-1\right)\overline{B}_\ell(\mathbf{x}^*) + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).\end{aligned}$$

A similar expression can be obtained for $\Delta_{k,a}(r\mathbf{x}^*)$ from which it can be deduced that

$$\tilde{\rho}_k(\mathbf{x}^*) = \rho + \frac{1-r^{\rho-1/2}}{\log r}\frac{1}{\sqrt{k}}\frac{\overline{B}_\ell(\mathbf{x}^*)}{\alpha\left(\frac{n}{k}\right)M(\mathbf{x}^*)}\frac{\int_0^1 K(u)u^{-1/2}\mathrm{d}u}{\frac{1}{k}\sum_{j=1}^{k}K(a_j)a_j^{-\rho}}\frac{1^{-1/2}-1}{a^{-\rho}-1} + \mathcal{O}\left(\frac{1}{\sqrt{k}\alpha\left(\frac{n}{k}\right)}\right).$$

The proposition follows. ■

Proposition B.3. *For any fixed d-vector* $\mathbf{x}^*$ *, under the assumptions of Theorem 5.2,*

$$\begin{aligned}\sqrt{k}\left(\tilde{\alpha}_k(\mathbf{x}) - \alpha\left(\frac{n}{k}\right)M(\mathbf{x})\right) \to &\frac{1-r^{\rho-1/2}}{\log r}\frac{M(\mathbf{x})}{M(\mathbf{x}^*)}\frac{\int_0^1 K(u)u^{-1/2}\mathrm{d}u}{\int_0^1 K(u)u^{-\rho}\mathrm{d}u}\frac{a^{-1/2}-1}{a^{-\rho}-1}\frac{\rho^2+\rho-1}{\rho(1-\rho)(1-2\rho)}\overline{B}(\mathbf{x}^*) \\ &+ \frac{2(1-\rho)}{\rho}\overline{B}(\mathbf{x}),\end{aligned}$$

in $D([0,T]^d)$ *for every* $T>0$*. Here,*

$$\tilde{\alpha}_k(\mathbf{x}) = \frac{\sum_{j=1}^{k}\sum_{l=1}^{k}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)} - a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)\hat{\ell}_{k,a_j}(\mathbf{x})}{\sum_{j=1}^{k}\sum_{l=1}^{k}a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)} - a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)},$$

and $\hat{\rho}_k(\mathbf{x}^*)$ *is defined as in the proposition above.*

Proof. To prove the proposition, we show that the following difference,

$$\begin{aligned}D = \Bigg|&\sqrt{k}\left(\tilde{\alpha}_k(\mathbf{x}) - \alpha\left(\frac{n}{k}\right)M(\mathbf{x})\right) \\ &- \frac{1-r^{\rho-1/2}}{\log r}\frac{M(\mathbf{x})}{M(\mathbf{x}^*)}\frac{\int_0^1 K(u)u^{-1/2}\mathrm{d}u}{\int_0^1 K(u)u^{-\rho}\mathrm{d}u}\frac{a^{-1/2}-1}{a^{-\rho}-1}\frac{\rho^2+\rho-1}{\rho(1-\rho)(1-2\rho)}\overline{B}_\ell(\mathbf{x}^*) \\ &- \frac{2(1-\rho)}{\rho}\overline{B}_\ell(\mathbf{x})\Bigg|,\end{aligned}$$

tends uniformly to zero as $n\to\infty$ almost surely.

$$\begin{aligned}D \le \Bigg|&\sqrt{k}\alpha\left(\frac{n}{k}\right)M(\mathbf{x})\frac{\sum_{j=1}^{k}\sum_{l=1}^{k}\left(a_j^{-\rho}-a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}-a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)}{\sum_{j=1}^{k}\sum_{l=1}^{k}a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}-a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)} \\ &\frac{1-r^{\rho-1/2}}{\log r}\frac{M(\mathbf{x})}{M(\mathbf{x}^*)}\frac{\int_0^1 K(u)u^{-1/2}\mathrm{d}u}{\int_0^1 K(u)u^{-\rho}\mathrm{d}u}\frac{a^{-1/2}-1}{a^{-\rho}-1}\frac{\rho^2+\rho-1}{\rho(1-\rho)(1-2\rho)}\overline{B}_\ell(\mathbf{x}^*)\Bigg| \\ &+\left|\left(\frac{\sum_{j=1}^{k}\sum_{l=1}^{k}a_j^{-1/2}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}-a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)}{\sum_{j=1}^{k}\sum_{l=1}^{k}a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}-a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)} - \frac{2(1-\rho)}{\rho}\right)\overline{B}_\ell(\mathbf{x})\right| + \mathcal{O}(1).\end{aligned}$$

The main idea now is to replace the terms with $\tilde{\rho}_k(\mathbf{x}^*)$ by the same terms with ρ and to study the difference by the mean value theorem. For instance, if we look at the term

$$\frac{1}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{k}a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}-a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right),$$

appearing several times as a denominator in the bound of D, we can rewrite as follows:

$$\begin{aligned}
&\frac{1}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{k} a_j^{-\rho}\left(a_j^{-\rho}-a_l^{-\rho}\right)+\frac{1}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{k}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}\left(a_j^{-\tilde{\rho}_k(\mathbf{x}^*)}-a_l^{-\tilde{\rho}_k(\mathbf{x}^*)}\right)-a_j^{-\rho}\left(a_j^{-\rho}-a_l^{-\rho}\right)\right)\\
&\quad=\frac{\rho^2}{(1-\rho)^2(1-2\rho)}+\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)-2\left(\tilde{\rho}_k(\mathbf{x}^*)-\rho\right)\left(\frac{1}{k}\sum_{j=1}^{k} a_j^{-2\breve{\rho}_k(\mathbf{x}^*)}\log a_j\right)\\
&\quad+2\left(\tilde{\rho}_k(\mathbf{x}^*)-\rho\right)\left(\frac{1}{k}\sum_{j=1}^{k} a_j^{-\breve{\rho}_k(\mathbf{x}^*)}\right)\left(\frac{1}{k}\sum_{j=1}^{k} a_j^{-\breve{\rho}_k(\mathbf{x}^*)}\log a_j\right)
\end{aligned}$$

by the mean value theorem where $\breve{\rho}_k(\mathbf{x}^*)$ is an intermediate value between $\tilde{\rho}_k(\mathbf{x}^*)$ and ρ. Using the same appraoch for each term in combination with Proposition B.2, the desired result is retrieved. ■

Appendix C

Additional simulation results

C.1 Evaluating the global STDF estimator performance

It is claimed by Segers et al. (2017) that the following equalities hold,

$$\text{Integrated squared bias}: \int_{[0,1]^d} \left(\mathbb{E}\left[\hat{\ell}_n(\mathbf{x})\right] - \ell(\mathbf{x})\right)^2 \mathrm{d}\mathbf{x} = \mathbb{E}\left[\left(\hat{\ell}_n^{(1)}(\mathbf{U}) - \ell(\mathbf{U})\right)\left(\hat{\ell}_n^{(2)}(\mathbf{U}) - \ell(\mathbf{U})\right)\right]$$

$$\text{Integrated variance}: \int_{[0,1]^d} \mathbb{E}\left[\left(\hat{\ell}_n(\mathbf{x}) - \mathbb{E}\left[\hat{\ell}_n(\mathbf{x})\right]\right)^2\right] \mathrm{d}\mathbf{x} = \mathbb{E}\left[\frac{1}{2}\left(\hat{\ell}_n^{(1)}(\mathbf{U}) - \hat{\ell}_n^{(2)}(\mathbf{U})\right)^2\right]$$

$$\text{Integrated MSE}: \int_{[0,1]^d} \mathbb{E}\left[\left(\hat{\ell}_n(\mathbf{x}) - \ell(\mathbf{x})\right)^2\right] \mathrm{d}\mathbf{x} = \mathbb{E}\left[\frac{1}{2}\left(\left(\hat{\ell}_n^{(1)}(\mathbf{U}) - \ell(\mathbf{U})\right)^2 + \left(\hat{\ell}_n^{(2)}(\mathbf{U}) - \ell(\mathbf{U})\right)^2\right)\right].$$

By conditionalizing on the random vector $\mathbf{U}$ these equalities can be shown as follows. Here we use the tower property of conditional expectation and the fact that $\hat{\ell}_n^{(1)}$ and $\hat{\ell}_n^{(2)}$ are estimators based on two independent random samples drawn from the same distribution. For the integrated squared bias it follows that,

$$\begin{aligned}
&\mathbb{E}\left[\left(\hat{\ell}_{n,k}^{(1)}(\mathbf{U}) - \ell(\mathbf{U})\right)\left(\hat{\ell}_n^{(2)}(\mathbf{U}) - \ell(\mathbf{U})\right)\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\mathbb{E}\left[\left(\hat{\ell}_{n,k}^{(1)}(\mathbf{U}) - \ell(\mathbf{U})\right)\left(\hat{\ell}_{n,k}^{(2)}(\mathbf{U}) - \ell(\mathbf{U})\right)\Big|\mathbf{U} = \mathbf{u}\right]\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\left(\mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{u})\right]\right)^2 - 2\ell(\mathbf{u})\mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{u})\right] + (\ell(\mathbf{u}))^2\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\left(\mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{u})\right] - \ell(\mathbf{u})\right)^2\right] = \int_{[0,1]^d}\left(\mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{u}) - \ell(\mathbf{u})\right]\right)^2 \mathrm{d}\mathbf{u}.
\end{aligned}$$

Similarly, it follows for the integrated variance that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{2}\left(\hat{\ell}_{n,k}^{(1)}(\mathbf{U}) - \hat{\ell}_{n,k}^{(2)}(\mathbf{U})\right)^2\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\mathbb{E}\left[\frac{1}{2}\left(\hat{\ell}_{n,k}^{(1)}(\mathbf{U}) - \hat{\ell}_{n,k}^{(2)}(\mathbf{U})\right)^2\Big|\mathbf{U} = \mathbf{u}\right]\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\frac{1}{2}\mathbb{E}\left[\left(\hat{\ell}_{n,k}^{(1)}(\mathbf{u})\right)^2 - 2\hat{\ell}_n^{(1)}(\mathbf{u})\hat{\ell}_{n,k}^{(2)}(\mathbf{u}) + \left(\hat{\ell}_{n,k}^{(2)}(\mathbf{u})\right)^2\right]\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{u})\right)^2\right] - \left(\mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{u})\right]\right)^2\right] = \int_{[0,1]^d}\mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{u}) - \mathbb{E}\left[\hat{\ell}_{n,k}(\mathbf{u})\right]\right)^2\right]\mathrm{d}\mathbf{u}.
\end{aligned}$$

Lastly, it follows for the integrated MSE that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{2}\left(\left(\hat{\ell}_{n,k}^{(1)}(\mathbf{U}) - \ell(\mathbf{U})\right)^2 + \left(\hat{\ell}_{n,k}^{(2)}(\mathbf{U}) - \ell(\mathbf{U})\right)^2\right)\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\mathbb{E}\left[\frac{1}{2}\left(\left(\hat{\ell}_{n,k}^{(1)}(\mathbf{U}) - \ell(\mathbf{U})\right)^2 + \left(\hat{\ell}_{n,k}^{(2)}(\mathbf{U}) - \ell(\mathbf{U})\right)^2\right)\Big|\mathbf{U} = \mathbf{u}\right]\right] \\
&\quad = \mathbb{E}_{\mathbf{u}}\left[\mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{u}) - \ell(\mathbf{u})\right)^2\right]\right] = \int_{[0,1]^d}\mathbb{E}\left[\left(\hat{\ell}_{n,k}(\mathbf{u}) - \ell(\mathbf{u})\right)^2\right]\mathrm{d}\mathbf{u}.
\end{aligned}$$

C.2 Initial STDF simulations

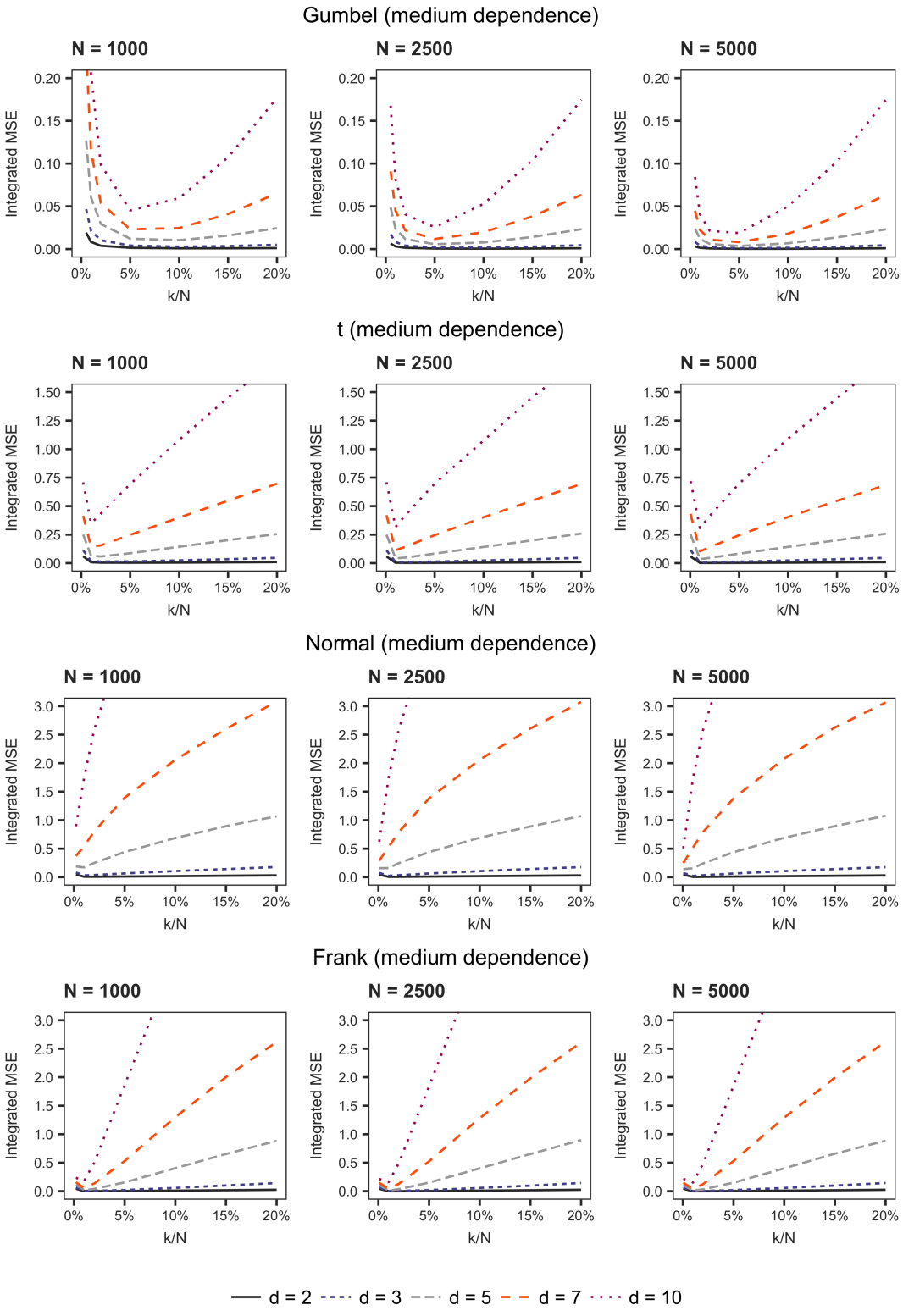

Figure C.2.1: Initial STDF simulations to determine the sample size N and the threshold value k. Shown is the integrated mean squared error (MSE) of the empirical STDF for data simulated from the Gumbel, t-, Normal, and Frank copula in dimensions $d \in \{2, 3, 5, 7, 10\}$. Model parameters are chosen such that all bivariate margins have medium dependence ($\rho = 0.5$). Results are based on $B = 20,000$ samples of k tail observations from a sample of size N.

C.3 STDF estimator simulations

Table C.3.1: Evaluation estimators in dimension $d = 2$ for tail dependent data with $k = 1\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 2, k = 25		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Gumbel							
Strong dep.	NP	0.14	1.7	1.8	0.0043	5.1	5.1
	NP-Beta	0.26	0.98	1.2	0.59	3.4	4
	NP-Kernel				0.06	5.5	5.6
	NP-BC1	0.97	5.6	6.6	0.19	21	21
	NP-BC2	1.6	11	12	0.58	27	28
	NP-BC3				0.019	5.8	5.8
Medium dep.	NP	0.18	2.9	3	0.028	7.7	7.7
	NP-Beta	0.1	2	2.1	0.089	6	6
	NP-Kernel				0.21	7.9	8.1
	NP-BC1	0.97	8.9	9.9	1	27	28
	NP-BC2	1.6	12	14	0.069	40	40
	NP-BC3				0.029	7.3	7.3
Light dep.	NP	0.24	2.2	2.5	0.018	5.3	5.3
	NP-Beta	0.012	1.7	1.7	0.0088	4.6	4.6
	NP-Kernel				0.28	6.2	6.5
	NP-BC1	0.76	5.7	6.5	0.0082	15	15
	NP-BC2	1.5	7.8	9.3	0.14	25	25
	NP-BC3				0.051	5.7	5.7
t							
Strong dep.	NP	0.17	1.9	2	0.045	6	6
	NP-Beta	0.34	1.2	1.6	1.1	4	5.1
	NP-Kernel				0.02	5.6	5.6
	NP-BC1	1	6.1	7.1	0.41	23	23
	NP-BC2	1.2	7.6	8.8	0.47	24	24
	NP-BC3				0.028	6	6
Medium dep.	NP	0.55	2.5	3.1	2	6.4	8.4
	NP-Beta	0.25	1.9	2.2	1.7	5.4	7
	NP-Kernel				3.1	6.7	9.8
	NP-BC1	0.73	6.8	7.5	0.056	22	22
	NP-BC2	2	12	14	0.9	39	40
	NP-BC3				2.5	7.5	10
Light dep.	NP	0.8	1	1.8	3.3	2.9	6.2
	NP-Beta	0.53	0.75	1.3	3.1	2.3	5.4
	NP-Kernel				4.2	2.7	6.8
	NP-BC1	1.2	1.9	3.1	2.9	7.3	10
	NP-BC2	3	3.2	6.2	1.6	6.8	8.3
	NP-BC3				3.9	3.3	7.3

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 25$ tail observations are used.

Table C.3.2: Evaluation estimators in dimension $d = 2$ for tail independent data with $k = 1\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 2, k = 25		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Normal							
Strong dep.	NP	28	2.9	31	142	7.7	149
	NP-Beta	26	1.9	28	137	5.6	143
	NP-Kernel				144	7.6	152
	NP-BC1	18	8	26	77	29	106
	NP-BC2	24	22	46	71	60	131
	NP-BC3				98	12	110
Medium dep.	NP	3.1	1.4	4.5	17	3.9	20
	NP-Beta	2.7	1	3.7	16	3.2	20
	NP-Kernel				18	3.8	22
	NP-BC1	2.3	2.5	4.8	6.3	7.8	14
	NP-BC2	6.3	7.9	14	7.1	13	20
	NP-BC3				9.6	4.5	14
Light dep.	NP	0.41	0.37	0.79	1.1	1.3	2.4
	NP-Beta	0.15	0.27	0.42	1.2	1.1	2.2
	NP-Kernel				1.9	1.2	3.1
	NP-BC1	0.92	0.48	1.4	0.78	2.9	3.6
	NP-BC2	3.1	1.2	4.3	0.51	3.3	3.8
	NP-BC3				0.9	1.1	2
Frank							
Strong dep.	NP	0.86	0.7	1.6	4.9	2.4	7.2
	NP-Beta	0.59	0.47	1.1	5.1	1.8	6.9
	NP-Kernel				6.2	2.6	8.7
	NP-BC1	0.86	0.68	1.5	0.63	2.5	3.1
	NP-BC2	11	8.7	20	108	157	265
	NP-BC3				5.5	2.1	7.6
Medium dep.	NP	0.38	0.33	0.71	0.96	1.3	2.2
	NP-Beta	0.13	0.24	0.37	1.1	0.92	2
	NP-Kernel				2	1.2	3.3
	NP-BC1	0.81	0.34	1.1	0.48	2	2.5
	NP-BC2	12	13	25	0.82	5.7	6.5
	NP-BC3				0.0015	0.086	0.087
Light dep.	NP	0.29	0.17	0.46	0.27	0.7	0.97
	NP-Beta	0.034	0.12	0.15	0.26	0.48	0.74
	NP-Kernel				0.89	0.61	1.5
	NP-BC1	0.81	0.18	0.99	0.36	1.7	2.1
	NP-BC2	3.5	0.74	4.2	0.14	1.8	1.9
	NP-BC3				0.1	0.38	0.48

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 25$ tail observations are used.

Table C.3.3: Evaluation estimators in dimension $d = 2$ for tail dependent data with $k = 5\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 2, k = 125		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Gumbel							
Strong dep.	NP	0.012	0.31	0.33	0.053	1.1	1.2
	NP-Beta	0.0018	0.25	0.25	0.032	0.78	0.81
	NP-Kernel				0.08	0.94	1
	NP-BC1	0.092	2.4	2.5	0.84	14	15
	NP-BC2	0.21	4	4.2	0.0029	6.7	6.7
	NP-BC3				0.00084	1.1	1.1
Medium dep.	NP	0.052	0.55	0.6	0.4	1.6	1.9
	NP-Beta	0.034	0.49	0.52	0.47	1.3	1.7
	NP-Kernel				0.36	1.5	1.9
	NP-BC1	0.14	3.1	3.2	0.13	10	10
	NP-BC2	0.12	4.1	4.2	0.015	7.9	7.9
	NP-BC3				0.028	1.7	1.7
Light dep.	NP	0.15	0.47	0.62	1.6	1.2	2.8
	NP-Beta	0.14	0.43	0.57	1.3	1.2	2.5
	NP-Kernel				1.1	1.1	1.2
	NP-BC1	0.086	1.8	1.9	0.083	6	6.1
	NP-BC2	0.095	2.5	2.6	0.023	5.6	5.6
	NP-BC3				0.025	1.6	1.6
t							
Strong dep.	NP	0.0059	0.38	0.38	0.0012	1.2	1.2
	NP-Beta	0.016	0.3	0.32	0.034	0.92	0.95
	NP-Kernel				0.00057	1.1	1.1
	NP-BC1	0.16	2.5	2.6	0.73	13	14
	NP-BC2	0.037	2.1	2.2	0.0052	4.7	4.7
	NP-BC3				0.0015	1.1	1.1
Medium dep.	NP	1.9	0.54	2.5	13	1.2	14
	NP-Beta	1.9	0.47	2.4	13	1.2	14
	NP-Kernel				11	1.5	13
	NP-BC1	0.18	2.4	2.6	1.2	8	9.2
	NP-BC2	0.32	4.6	4.9	1.3	9.4	11
	NP-BC3				3.2	2.5	5.7
Light dep.	NP	2.4	0.3	2.7	15	0.83	16
	NP-Beta	2.3	0.27	2.6	15	0.82	15
	NP-Kernel				13	0.85	14
	NP-BC1	0.54	0.75	1.3	2.5	2.6	5.1
	NP-BC2	0.74	1.3	2	2.7	3.2	5.9
	NP-BC3				5.3	1.5	6.7

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 125$ tail observations are used.

Table C.3.4: Evaluation estimators in dimension $d = 2$ for tail independent data with $k = 5\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 2, k = 125		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Normal							
Strong dep.	NP	49	0.53	49	244	1.2	245
	NP-Beta	47	0.44	48	243	1.3	244
	NP-Kernel				231	1.3	232
	NP-BC1	28	3.2	31	129	15	144
	NP-BC2	27	14	41	108	18	126
	NP-BC3				129	12	141
Medium dep.	NP	10	0.44	10	60	1.1	62
	NP-Beta	9.8	0.38	10	59	1.1	60
	NP-Kernel				52	1.2	54
	NP-BC1	2.7	1.4	4.1	15	5	20
	NP-BC2	2.5	4	6.5	12	7.5	19
	NP-BC3				16	4	20
Light dep.	NP	1.5	0.21	1.7	11	0.73	12
	NP-Beta	1.4	0.18	1.6	11	0.64	12
	NP-Kernel				9.4	0.61	10
	NP-BC1	0.21	0.28	0.49	1.1	1.2	2.3
	NP-BC2	0.32	0.56	0.88	0.78	1.5	2.2
	NP-BC3				1.2	0.99	2.2
Frank							
Strong dep.	NP	10	0.41	11	81	1.2	82
	NP-Beta	10	0.35	10	80	1.1	81
	NP-Kernel				65	1.2	66
	NP-BC1	0.57	0.64	1.2	3.7	3.2	6.8
	NP-BC2	49	9.4	58	323	225	548
	NP-BC3				52	1.5	53
Medium dep.	NP	2.8	0.26	3.1	23	0.78	24
	NP-Beta	2.7	0.22	2.9	23	0.72	24
	NP-Kernel				18	0.77	19
	NP-BC1	0.17	0.27	0.44	0.69	1	1.7
	NP-BC2	4	12	16	0.063	0.78	0.84
	NP-BC3				0.006	0.23	0.24
Light dep.	NP	0.79	0.15	0.94	6.6	0.54	7.2
	NP-Beta	0.74	0.13	0.87	6.6	0.52	7.1
	NP-Kernel				5.3	0.48	5.8
	NP-BC1	0.079	0.13	0.21	0.16	0.41	0.57
	NP-BC2	0.18	0.34	0.52	0.2	0.67	0.87
	NP-BC3				0.035	0.23	0.26

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 125$ tail observations are used.

Table C.3.5: Evaluation estimators in dimension $d = 3$ for tail dependent data with $k = 1\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 3, k = 25		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Gumbel							
Strong dep.	NP	0.16	4.2	4.3	0.011	3	3
	NP-Beta	0.74	2.6	3.4	0.59	2.5	3.1
	NP-Kernel				0.0014	3.2	3.2
	NP-BC1	1.3	16	17	0.66	20	21
	NP-BC2	3.6	32	36	0.46	17	18
	NP-BC3				0.0032	3.4	3.4
Medium dep.	NP	0.29	7.9	8.2	0.00011	5.3	5.3
	NP-Beta	0.33	6.1	6.4	0.2	4.8	5
	NP-Kernel				0.066	5.7	5.7
	NP-BC1	1.5	29	30	0.49	20	20
	NP-BC2	3	42	45	0.84	31	32
	NP-BC3				0.0058	5.8	5.8
Light dep.	NP	0.26	7.1	7.3	0.017	4.3	4.3
	NP-Beta	0.1	5.7	5.8	0.0017	4	4
	NP-Kernel				0.14	5.1	5.2
	NP-BC1	1.1	21	22	0.34	15	15
	NP-BC2	3.4	31	34	0.58	24	25
	NP-BC3				0.025	5.6	5.7
t							
Strong dep.	NP	1.5	4.1	5.7	7.3	3.1	10
	NP-Beta	0.61	2.8	3.4	5.1	2	7.1
	NP-Kernel				9.2	2.8	12
	NP-BC1	1.5	15	17	4.6	13	17
	NP-BC2	1.9	18	20	5.6	14	20
	NP-BC3				8.4	3	11
Medium dep.	NP	3.2	6.6	9.8	5.4	4.4	9.8
	NP-Beta	2.2	5.2	7.4	4.8	3.8	8.5
	NP-Kernel				5.5	4.4	9.9
	NP-BC1	1.4	22	23	0.83	17	18
	NP-BC2	4.3	44	49	0.42	34	34
	NP-BC3				1.8	4.8	6.6
Light dep.	NP	4.2	2.9	7.1	6.8	1.9	8.7
	NP-Beta	3.7	2.3	5.9	6.8	1.6	8.4
	NP-Kernel				7.1	2	9.1
	NP-BC1	3.1	6	9.1	2.2	5	7.2
	NP-BC2	7.3	11	19	2.3	9.6	12
	NP-BC3				1.4	2.2	3.7

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 25$ tail observations are used.

Table C.3.6: Evaluation estimators in dimension $d = 3$ for tail independent data with $k = 1\%$.

$\mathbf{d = 3, k = 25}$		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
		Int. bias^2 $\times 1000$	Int. var. $\times 1000$	Int. MSE $\times 1000$	Bias^2 $\times 1000$	Var. $\times 1000$	MSE $\times 1000$
Normal							
Strong dep.	NP	153	6.9	160	200	3.9	204
	NP-Beta	149	5.2	154	198	3.4	201
	NP-Kernel				198	5	203
	NP-BC1	96	25	121	125	16	141
	NP-BC2	108	94	202	94	52	146
	NP-BC3				116	11	128
Medium dep.	NP	19	3.9	23	30	2.9	33
	NP-Beta	18	3.1	22	30	2.3	32
	NP-Kernel				30	2.9	33
	NP-BC1	8.6	8	17	8.5	7.3	16
	NP-BC2	20	29	50	8	16	24
	NP-BC3				5.1	3.7	8.8
Light dep.	NP	1.5	1.1	2.6	2.4	0.97	3.3
	NP-Beta	1.1	0.81	1.9	2.2	0.72	3
	NP-Kernel				2.9	0.79	3.7
	NP-BC1	1.8	1.3	3.1	0.6	1.2	1.8
	NP-BC2	7.2	3.6	11	0.85	2.8	3.6
	NP-BC3				0.04	0.22	0.26
Frank							
Strong dep.	NP	4.8	1.9	6.7	11	1.7	13
	NP-Beta	4.4	1.4	5.8	11	1.4	12
	NP-Kernel				9.4	1.6	11
	NP-BC1	1.7	1.9	3.6	0.56	1.6	2.2
	NP-BC2	35	20	54	203	162	365
	NP-BC3				9.1	1.4	10
Medium dep.	NP	1.4	0.97	2.3	2.7	0.95	3.6
	NP-Beta	0.99	0.7	1.7	2.7	0.65	3.3
	NP-Kernel				2.8	0.73	3.5
	NP-BC1	1.4	0.96	2.4	0.29	0.79	1.1
	NP-BC2	53	80	134	0.92	6.4	7.3
	NP-BC3				0.17	0.68	0.85
Light dep.	NP	0.61	0.5	1.1	0.64	0.48	1.1
	NP-Beta	0.26	0.37	0.62	0.67	0.34	1
	NP-Kernel				0.93	0.4	1.3
	NP-BC1	1.4	0.54	2	0.13	0.54	0.67
	NP-BC2	8.6	2.1	11	0.22	1.4	1.6
	NP-BC3				0.000033	0.0032	0.0033

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 25$ tail observations are used.

Table C.3.7: Evaluation estimators in dimension $d = 3$ for tail dependent data with $k = 5\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 3, k = 125		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Gumbel							
Strong dep.	NP	0.027	0.77	0.8	0.089	0.65	0.73
	NP-Beta	0.0086	0.64	0.65	0.037	0.53	0.57
	NP-Kernel				0.071	0.59	0.66
	NP-BC1	0.26	8.4	8.6	0.3	8.6	8.9
	NP-BC2	0.23	11	11	0.0057	3.4	3.4
	NP-BC3				0.0056	0.79	0.8
Medium dep.	NP	0.23	1.6	1.8	0.7	1	1.7
	NP-Beta	0.19	1.4	1.6	0.53	0.99	1.5
	NP-Kernel				0.53	0.95	1.5
	NP-BC1	0.39	11	11	0.51	10	11
	NP-BC2	0.15	11	11	0.087	6.3	6.3
	NP-BC3				0.0031	1.7	1.7
Light dep.	NP	0.91	1.4	2.4	2.4	0.85	3.3
	NP-Beta	0.88	1.3	2.2	2.2	0.83	3.1
	NP-Kernel				1.9	0.83	2.7
	NP-BC1	0.35	6.6	7	0.35	6.1	6.4
	NP-BC2	0.2	8.7	8.8	0.39	7.2	7.6
	NP-BC3				0.93	2.3	3.2
t							
Strong dep.	NP	2	0.8	2.8	9.6	0.54	10
	NP-Beta	1.7	0.66	2.3	8.6	0.44	9
	NP-Kernel				9.8	0.5	10
	NP-BC1	1	8.4	9.4	5.9	7.3	13
	NP-BC2	1.7	4.4	6	7.7	2.8	10
	NP-BC3				8.9	0.54	9.4
Medium dep.	NP	12	1.4	13	21	0.8	22
	NP-Beta	12	1.2	13	21	0.73	22
	NP-Kernel				20	0.74	21
	NP-BC1	1.6	8.6	10	2.7	9.1	12
	NP-BC2	1.4	15	16	1	7.7	8.7
	NP-BC3				2.1	2.7	4.8
Light dep.	NP	16	0.82	16	28	0.52	28
	NP-Beta	16	0.74	16	28	0.49	29
	NP-Kernel				24	0.53	24
	NP-BC1	3.1	2.8	6	4.5	2.8	7.3
	NP-BC2	2.4	4.3	6.6	1.3	3.5	4.8
	NP-BC3				0.79	1.7	2.5

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 125$ tail observations are used.

Table C.3.8: Evaluation estimators in dimension $d = 3$ for tail independent data with $k = 5\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 3, k = 125		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Normal							
Strong dep.	NP	248	1.2	250	330	0.66	331
	NP-Beta	250	1.1	251	326	0.72	326
	NP-Kernel				312	0.75	312
	NP-BC1	152	10	162	196	12	208
	NP-BC2	119	63	182	138	18	156
	NP-BC3				148	14	161
Medium dep.	NP	62	1.1	63	99	0.7	100
	NP-Beta	63	1	64	99	0.62	99
	NP-Kernel				88	0.73	89
	NP-BC1	17	5.2	23	25	5.9	31
	NP-BC2	6.7	13	20	7	6.5	14
	NP-BC3				7.5	4.6	12
Light dep.	NP	10	0.57	11	22	0.42	22
	NP-Beta	11	0.51	11	22	0.37	22
	NP-Kernel				18	0.41	18
	NP-BC1	1	0.99	2	1.8	1.3	3
	NP-BC2	0.58	1.4	1.9	0.15	0.74	0.89
	NP-BC3				0.0092	0.11	0.12
Frank							
Strong dep.	NP	64	1.1	65	130	0.72	131
	NP-Beta	64	0.91	65	131	0.62	132
	NP-Kernel				106	0.65	107
	NP-BC1	3	2.6	5.7	5.2	4.7	9.9
	NP-BC2	246	20	266	598	167	765
	NP-BC3				93	0.79	94
Medium dep.	NP	19	0.71	20	43	0.55	44
	NP-Beta	19	0.6	19	43	0.58	44
	NP-Kernel				34	0.51	35
	NP-BC1	0.75	0.9	1.6	1.6	1.4	3
	NP-BC2	53	97	150	0	0	0
	NP-BC3				1.5	4.2	5.7
Light dep.	NP	5.5	0.42	5.9	14	0.36	14
	NP-Beta	5.5	0.38	5.8	14	0.32	14
	NP-Kernel				10	0.36	11
	NP-BC1	0.24	0.39	0.64	0.38	0.47	0.85
	NP-BC2	0.21	0.6	0.81	0.005	0.1	0.11
	NP-BC3				0.0000062	0.0015	0.0015

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 125$ tail observations are used.

Table C.3.9: Evaluation estimators in dimension $d = 5$ for tail dependent data with $k = 1\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 5, k = 25		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Gumbel							
Strong dep.	NP	0.28	10	11	0.00012	2.1	2.1
	NP-Beta	1.8	7.5	9.2	0.25	1.6	1.8
	NP-Kernel				0.00093	2	2
	NP-BC1	3	42	45	0.22	7.4	7.6
	NP-BC2	6.8	83	89	0.26	11	11
	NP-BC3				0.0044	1.9	1.9
Medium dep.	NP	0.51	24	24	0.021	3.4	3.4
	NP-Beta	0.69	20	20	0.02	3.3	3.3
	NP-Kernel				0.064	4.5	4.5
	NP-BC1	4.4	93	97	0.61	18	19
	NP-BC2	5.5	150	155	0.51	40	40
	NP-BC3				0.12	4.5	4.6
Light dep.	NP	0.63	25	25	0.21	3.8	4
	NP-Beta	0.13	21	21	0.056	3.1	3.1
	NP-Kernel				0.31	4.6	4.9
	NP-BC1	3.1	80	83	0.41	14	15
	NP-BC2	7.6	152	160	0.57	86	86
	NP-BC3				27	6.3	33
t							
Strong dep.	NP	22	8.2	30	27	1.3	28
	NP-Beta	13	5.7	19	21	0.98	22
	NP-Kernel				27	1.4	28
	NP-BC1	15	34	48	22	6.1	28
	NP-BC2	16	37	53	22	7.3	29
	NP-BC3				27	1.3	28
Medium dep.	NP	24	17	42	11	2.8	14
	NP-Beta	20	14	34	10	2.4	13
	NP-Kernel				11	2.9	14
	NP-BC1	5.3	69	74	1.7	12	14
	NP-BC2	10	182	193	0.42	57	58
	NP-BC3				0.32	6.5	6.8
Light dep.	NP	33	8.8	42	14	1.4	16
	NP-Beta	30	6.9	37	14	1.1	15
	NP-Kernel				15	1.5	16
	NP-BC1	12	22	34	3.8	4.3	8.1
	NP-BC2	20	53	74	2.2	31	33
	NP-BC3				2.1	0.0058	2.1

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 25$ tail observations are used.

Table C.3.10: Evaluation estimators in dimension $d = 5$ for tail independent data with $k = 1\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 5, k = 25		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Normal							
Strong dep.	NP	911	17	928	296	2.5	298
	NP-Beta	868	14	881	290	2	292
	NP-Kernel				291	2.4	293
	NP-BC1	593	76	670	195	13	209
	NP-BC2	502	413	914	99	63	162
	NP-BC3				139	19	158
Medium dep.	NP	148	12	159	59	1.8	61
	NP-Beta	140	9.2	149	59	1.7	61
	NP-Kernel				58	2.2	60
	NP-BC1	51	30	81	21	6.3	27
	NP-BC2	60	133	193	4.4	27	32
	NP-BC3				0.26	2.7	3
Light dep.	NP	11	3.5	15	6.5	0.7	7.2
	NP-Beta	10	2.6	13	6	0.58	6.6
	NP-Kernel				6.2	0.76	7
	NP-BC1	4.3	4.4	8.8	0.76	1.1	1.8
	NP-BC2	20	20	39	9	70	78
	NP-BC3				0.035	0.41	0.44
Frank							
Strong dep.	NP	38	5.9	44	26	1.1	27
	NP-Beta	38	4.3	42	26	0.9	27
	NP-Kernel				21	1.1	22
	NP-BC1	4.3	5.8	10	1	1.6	2.6
	NP-BC2	162	42	203	429	86	514
	NP-BC3				20	1	21
Medium dep.	NP	9.4	3.1	12	6.7	0.68	7.3
	NP-Beta	8.9	2.3	11	6.5	0.64	7.1
	NP-Kernel				6	0.6	6.5
	NP-BC1	2.8	2.8	5.6	0.23	0.63	0.86
	NP-BC2	481	460	941	86	182	267
	NP-BC3				5.3	0.67	6
Light dep.	NP	2.9	1.6	4.5	1.7	0.36	2
	NP-Beta	2.3	1.2	3.5	1.8	0.33	2.1
	NP-Kernel				2	0.33	2.3
	NP-BC1	2.4	1.5	3.9	0.064	0.2	0.26
	NP-BC2	30	17	47	40	137	176
	NP-BC3				0.32	0.51	0.82

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 25$ tail observations are used.

Table C.3.11: Evaluation estimators in dimension $d = 5$ for tail dependent data with $k = 1\%$.

		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
d = 5, k = 125		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Gumbel							
Strong dep.	NP	0.14	1.9	2	0.12	0.34	0.46
	NP-Beta	0.026	1.6	1.6	0.055	0.29	0.35
	NP-Kernel				0.064	0.32	0.38
	NP-BC1	0.62	23	23	0.24	6	6.2
	NP-BC2	0.56	23	23	0.0023	2.2	2.2
	NP-BC3				0.00078	0.5	0.5
Medium dep.	NP	1.4	4.3	5.7	1.1	0.67	1.7
	NP-Beta	1	4	5	0.9	0.65	1.5
	NP-Kernel				0.75	0.74	1.5
	NP-BC1	2.3	41	43	0.52	12	12
	NP-BC2	0.49	33	33	0.16	7.4	7.5
	NP-BC3				1	2	3
Light dep.	NP	7	4.5	11	5.4	0.64	6
	NP-Beta	6.5	4.3	11	5	0.6	5.6
	NP-Kernel				3.8	0.8	4.6
	NP-BC1	2	28	30	0.19	6.7	6.9
	NP-BC2	2.7	41	44	18	17	34
	NP-BC3				51	2	53
t							
Strong dep.	NP	26	1.6	28	29	0.23	29
	NP-Beta	24	1.3	25	27	0.23	28
	NP-Kernel				29	0.24	29
	NP-BC1	18	20	39	25	3	28
	NP-BC2	24	7.7	32	27	1.5	29
	NP-BC3				27	0.28	28
Medium dep.	NP	80	3.2	83	36	0.4	37
	NP-Beta	78	2.9	81	35	0.46	36
	NP-Kernel				32	0.5	32
	NP-BC1	13	38	52	5.9	13	18
	NP-BC2	4.2	52	56	0.32	16	16
	NP-BC3				1.6	7	8.6
Light dep.	NP	121	2.3	123	57	0.35	57
	NP-Beta	120	2.1	122	56	0.28	57
	NP-Kernel				50	0.39	50
	NP-BC1	22	15	37	9	4.7	14
	NP-BC2	4.5	18	22	1.7	0.73	2.4
	NP-BC3				2.1	0.00096	2.1

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 125$ tail observations are used.

Table C.3.12: Evaluation estimators in dimension $d = 5$ for tail independent data with $k = 5\%$.

d = 5, k = 125		STDF results - $\hat{\ell}_n$			TDC results - $\hat{\Lambda}_n$		
		Int. bias^2 ×1000	Int. var. ×1000	Int. MSE ×1000	Bias^2 ×1000	Var. ×1000	MSE ×1000
Normal							
Strong dep.	NP	1370	2.7	1370	444	0.32	444
	NP-Beta	1350	2.4	1350	441	0.29	442
	NP-Kernel				427	0.36	427
	NP-BC1	898	42	940	283	14	297
	NP-BC2	538	299	836	137	27	165
	NP-BC3				164	26	190
Medium dep.	NP	436	2.9	439	174	0.35	175
	NP-Beta	429	2.6	431	174	0.31	174
	NP-Kernel				158	0.42	158
	NP-BC1	124	29	154	43	12	55
	NP-BC2	11	43	54	0.085	1.3	1.4
	NP-BC3				0.54	6.2	6.7
Light dep.	NP	90	1.6	92	49	0.3	50
	NP-Beta	88	1.4	90	50	0.24	50
	NP-Kernel				41	0.26	41
	NP-BC1	7.3	5	12	2.8	2	4.8
	NP-BC2	0.44	2.5	2.9	2.3	46	48
	NP-BC3				0.19	2	2.2
Frank							
Strong dep.	NP	443	2.6	446	222	0.4	222
	NP-Beta	444	2.3	446	220	0.35	220
	NP-Kernel				186	0.36	186
	NP-BC1	19	15	34	7.1	7	14
	NP-BC2	1360	39	1400	923	35	958
	NP-BC3				171	0.42	172
Medium dep.	NP	149	1.9	151	88	0.32	88
	NP-Beta	148	1.7	150	88	0.28	88
	NP-Kernel				71	0.33	71
	NP-BC1	4.9	4.5	9.4	1.5	2	3.5
	NP-BC2	1330	547	1880	122	228	350
	NP-BC3				56	0.69	57
Light dep.	NP	47	1.3	49	32	0.26	33
	NP-Beta	48	1.1	49	32	0.22	32
	NP-Kernel				25	0.23	25
	NP-BC1	1.4	1.6	3	0.48	0.64	1.1
	NP-BC2	0.089	1.1	1.2	45	167	212
	NP-BC3				3.2	3.6	6.8

Considered are the empirical nonparametric estimator (NP), the beta-smoothed estimator of Kiriliouk et al. (2018) (NP-Beta), the kernel-smoothed estimator (NP-Kernel), the bias-corrected estimators of Fougères et al. (2015) (NP-BC1, NP-BC2) and the bias-corrected estimator of Beirlant et al. (2016) (NP-BC3). Data are simulated from distributions with different dependence strengths. Strong, medium and weak dependence correspond to bivariate correlations of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. STDF results are based on $B = 20,000$ simulation samples, while the TDC results are based on $B = 500$ simulation samples. Each sample consists of $N = 2500$ observations, of which $k = 125$ tail observations are used.

C.4 Testing simulations

Table C.4.1: Evaluation test statistics in dimension $d = 2$ for tail dependent data.

		STDF test rejection rates				TDC test rejection rates	
		I_n		S_n		T_n	
d = 2, k = 25		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Gumbel							
Strong dep.	NP	0.098	0.012	0.11	0.012	0.36	0.072
	NP-Kernel			0.058	0.004	0.29	0.034
	NP-BC					0	0
Medium dep.	NP	0.076	0.014	0.07	0.004	0.19	0.028
	NP-Kernel			0.03	0	0.14	0.014
	NP-BC					0	0
Light dep.	NP	0.054	0.012	0.038	0.004	0.084	0.01
	NP-Kernel			0.022	0.004	0.06	0
	NP-BC					0	0
t							
Strong dep.	NP	0.098	0	0.11	0.004	0.33	0.068
	NP-Kernel			0.078	0	0.28	0.04
	NP-BC					0	0
Medium dep.	NP	0.044	0.014	0.024	0.004	0.1	0.006
	NP-Kernel			0.038	0.004	0.096	0.012
	NP-BC					0	0
Light dep.	NP	0.054	0.012	0.03	0.01	0.048	0.006
	NP-Kernel			0.03	0.01	0.038	0.002
	NP-BC					0	0

Tests are conducted to evaluate the null hypothesis of tail independence. Data are simulated from distributions with different dependence strengths. Strong, medium and weak tail dependence correspond to a bivariate correlation of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. Results are based on $B = 500$ samples with sample size $N = 2500$. The sample is split into two equally sized subsamples ($\theta = 1$) and $k = 1\%$ of the remaining 1250 observations are used as tail observations. All implementations of the test statistics are our own (see Appendix E.2).

Table C.4.2: Evaluation test statistics in dimension $d = 2$ for tail independent data.

		STDF test rejection rates				TDC test rejection rates	
		I_n		S_n		T_n	
d = 2, k = 25		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Normal							
Strong dep.	NP	0.05	0.01	0.036	0.006	0.17	0.014
	NP-Kernel			0.036	0.004	0.14	0.012
	NP-BC					0	0
Medium dep.	NP	0.046	0.01	0.032	0.006	0.076	0.002
	NP-Kernel			0.028	0.006	0.068	0.002
	NP-BC					0	0
Light dep.	NP	0.048	0.01	0.034	0.01	0.05	0
	NP-Kernel			0.032	0.006	0.036	0
	NP-BC					0	0
Frank							
Strong dep.	NP	0.028	0.008	0.044	0.008	0.04	0.006
	NP-Kernel			0.02	0.004	0.028	0.004
	NP-BC					0	0
Medium dep.	NP	0.022	0.002	0.028	0.008	0.03	0
	NP-Kernel			0.018	0.002	0.02	0
	NP-BC					0	0
Light dep.	NP	0.032	0.004	0.034	0.008	0.024	0
	NP-Kernel			0.024	0.006	0.014	0
	NP-BC					0	0

Tests are conducted to evaluate the null hypothesis of tail independence. Data are simulated from distributions with different dependence strengths. Strong, medium and weak tail dependence correspond to a bivariate correlation of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. Results are based on $B = 500$ samples with sample size $N = 2500$. The sample is split into two equally sized subsamples ($\theta = 1$) and $k = 1\%$ of the remaining 1250 observations are used as tail observations. All implementations of the test statistics are our own (see Appendix E.2).

Table C.4.3: Evaluation test statistics in dimension $d = 3$ for tail dependent data.

		STDF test rejection rates				TDC test rejection rates	
		I_n		S_n		T_n	
d = 3, k = 25		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Gumbel							
Pairwise testing							
Strong dep.	NP	0.038	0.004	0.064	0.004	0.25	0.032
	NP-Kernel			0.034	0	0.2	0.022
Medium dep.	NP	0.036	0.006	0.036	0.004	0.098	0.008
	NP-Kernel			0.02	0.002	0.084	0.008
Light dep.	NP	0.042	0.008	0.03	0.004	0.03	0
	NP-Kernel			0.024	0.002	0.026	0.002
Multivariate testing							
Strong dep.	NP	0.32	0.072	0.41	0.15	0.75	0.41
	NP-Kernel			0.29	0.062	0.66	0.31
	NP-BC					0.8	0.004
Medium dep.	NP	0.13	0.014	0.15	0.032	0.39	0.13
	NP-Kernel			0.098	0.012	0.32	0.078
	NP-BC					0.09	0
Light dep.	NP	0.052	0	0.032	0	0.15	0.018
	NP-Kernel			0.022	0.002	0.14	0.014
	NP-BC					0	0
t							
Pairwise testing							
Strong dep.	NP	0.056	0.006	0.078	0.004	0.23	0.054
	NP-Kernel			0.034	0.004	0.19	0.022
Medium dep.	NP	0.04	0.014	0.028	0.008	0.038	0
	NP-Kernel			0.028	0.012	0.022	0.002
Light dep.	NP	0.034	0.022	0.038	0.018	0.032	0
	NP-Kernel			0.024	0.018	0.02	0
Multivariate testing							
Strong dep.	NP	0.32	0.082	0.42	0.15	0.75	0.41
	NP-Kernel			0.32	0.086	0.66	0.32
	NP-BC					0.81	0.002
Medium dep.	NP	0.058	0.014	0.058	0.004	0.2	0.034
	NP-Kernel			0.042	0.006	0.17	0.02
	NP-BC					0.014	0
Light dep.	NP	0.052	0.016	0.04	0.01	0.086	0.004
	NP-Kernel			0.024	0.012	0.078	0.008
	NP-BC					0	0

Tests are conducted to evaluate the null hypothesis of tail independence with either a multivariate testing approach or a pairwise testing approach with a Hochberg correction for the p-values. Data are simulated from distributions with different dependence strengths. Strong, medium and weak tail dependence correspond to a bivariate correlation of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. Results are based on $B = 500$ samples with sample size $N = 2500$. The sample is split into two equally sized subsamples ($\theta = 1$) and $k = 1\%$ of the remaining 1250 observations are used as tail observations. All implementations of the test statistics are our own (see Appendix E.2).

Table C.4.4: Evaluation test statistics in dimension $d = 5$ for tail independent data.

		STDF test rejection rates				TDC test rejection rates	
		I_n		S_n		T_n	
d = 3, k = 25		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Normal							
Pairwise testing							
Strong dep.	NP	0.034	0.006	0.02	0	0.084	0.008
	NP-Kernel			0.016	0.002	0.098	0.004
Medium dep.	NP	0.04	0.018	0.024	0.008	0.02	0
	NP-Kernel			0.026	0.01	0.024	0
Light dep.	NP	0.044	0.014	0.036	0.008	0.012	0
	NP-Kernel			0.026	0.006	0.012	0
Multivariate testing							
Strong dep.	NP	0.13	0.022	0.12	0.02	0.37	0.13
	NP-Kernel			0.098	0.01	0.3	0.09
	NP-BC					0.27	0
Medium dep.	NP	0.046	0.008	0.04	0.006	0.12	0.012
	NP-Kernel			0.038	0.002	0.086	0.016
	NP-BC					0.002	0
Light dep.	NP	0.044	0.016	0.032	0.006	0.044	0.006
	NP-Kernel			0.028	0.004	0.036	0.002
	NP-BC					0	0
Frank							
Pairwise testing							
Strong dep.	NP	0.032	0.01	0.026	0.008	0.018	0.002
	NP-Kernel			0.02	0.006	0.028	0.002
Medium dep.	NP	0.044	0.02	0.054	0.014	0.006	0
	NP-Kernel			0.028	0.016	0.01	0
Light dep.	NP	0.046	0.02	0.046	0.02	0.016	0
	NP-Kernel			0.032	0.014	0.02	0
Multivariate testing							
Strong dep.	NP	0.032	0.01	0.036	0.004	0.068	0.008
	NP-Kernel			0.02	0.002	0.062	0.006
	NP-BC					0.14	0
Medium dep.	NP	0.05	0.014	0.038	0.01	0.034	0.002
	NP-Kernel			0.026	0.01	0.034	0.002
	NP-BC					0	0
Light dep.	NP	0.044	0.01	0.044	0.014	0.054	0.008
	NP-Kernel			0.038	0.01	0.054	0.008
	NP-BC					0	0

Tests are conducted to evaluate the null hypothesis of tail independence with either a multivariate testing approach or a pairwise testing approach with a Hochberg correction for the p-values. Data are simulated from distributions with different dependence strengths. Strong, medium and weak tail dependence correspond to a bivariate correlation of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. Results are based on $B = 500$ samples with sample size $N = 2500$. The sample is split into two equally sized subsamples ($\theta = 1$) and $k = 1\%$ of the remaining 1250 observations are used as tail observations. All implementations of the test statistics are our own (see Appendix E.2).

Table C.4.5: Evaluation test statistics in dimension $d = 3$ for mixed dependencies.

		STDF test rejection rates				TDC test rejection rates	
		I_n		S_n		T_n	
d = 3, k = 25		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Asymmetric tail dependent distributions							
Pairwise testing							
t	NP	0.036	0.008	0.026	0.002	0.052	0.004
	NP-Kernel			0.022	0.002	0.048	0.002
D-Vine 1	NP	0.03	0.016	0.032	0.008	0.072	0.006
	NP-Kernel			0.018	0.008	0.07	0
D-Vine 2	NP	0.034	0.018	0.024	0.012	0.036	0
	NP-Kernel			0.024	0.008	0.04	0
Multivariate testing							
t	NP	0.076	0.006	0.072	0.01	0.28	0.06
	NP-Kernel			0.034	0.004	0.24	0.032
	NP-BC					0.02	0
D-Vine 1	NP	0.08	0.014	0.078	0.016	0.27	0.052
	NP-Kernel			0.058	0.004	0.22	0.042
	NP-BC					0.028	0
D-Vine 2	NP	0.058	0.012	0.054	0.008	0.15	0.024
	NP-Kernel			0.03	0.006	0.12	0.016
	NP-BC					0.016	0
Asymmetric tail independent distributions							
Pairwise testing							
Normal	NP	0.036	0.01	0.028	0.006	0.058	0.002
	NP-Kernel			0.026	0.008	0.032	0
D-Vine 3	NP	0.04	0.02	0.036	0.012	0.036	0
	NP-Kernel			0.026	0.016	0.032	0
D-Vine 4	NP	0.046	0.02	0.04	0.014	0.018	0
	NP-Kernel			0.026	0.014	0.024	0
Multivariate testing							
Normal	NP	0.056	0.004	0.048	0.002	0.18	0.038
	NP-Kernel			0.028	0.004	0.13	0.018
	NP-BC					0.004	0
D-Vine 3	NP	0.056	0.01	0.046	0.006	0.164	0.03
	NP-Kernel			0.03	0.004	0.112	0.014
	NP-BC					0.022	0
D-Vine 4	NP	0.06	0.018	0.042	0.008	0.04	0.002
	NP-Kernel			0.034	0.006	0.032	0.006
	NP-BC					0	0

Tests are conducted to evaluate the null hypothesis of tail independence with either a multivariate testing approach or a pairwise testing approach with a Hochberg correction for the p-values. Data are simulated from distributions with different dependence strengths. Strong, medium and weak tail dependence correspond to a bivariate correlation of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. Results are based on $B = 500$ samples with sample size $N = 2500$. The sample is split into two equally sized subsamples ($\theta = 1$) and $k = 1\%$ of the remaining 1250 observations are used as tail observations. All implementations of the test statistics are our own (see Appendix E.2).

Table C.4.6: Evaluation test statistics in dimension $d = 5$ for tail dependent data.

		STDF test rejection rates				TDC test rejection rates	
		I_n^*		S_n		T_n	
d = 5, k = 25		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Gumbel							
Pairwise testing							
Strong dep.	NP			0.04	0	0.16	0.024
	NP-Kernel					0.13	0.012
Medium dep.	NP			0.014	0.008	0.044	0.004
	NP-Kernel					0.036	0
Light dep.	NP			0.02	0.008	0.016	0.002
	NP-Kernel					0.014	0
Multivariate testing							
Strong dep.	NP			0.94	0.81	0.98	0.94
	NP-Kernel					0.97	0.9
	NP-BC					0.99	0.96
Medium dep.	NP			0.59	0.3	0.85	0.57
	NP-Kernel					0.77	0.47
	NP-BC					0.74	0.25
Light dep.	NP			0.14	0.024	0.35	0.076
	NP-Kernel					0.29	0.072
	NP-BC					0	0
t							
Pairwise testing							
Strong dep.	NP			0.036	0.002	0.16	0.026
	NP-Kernel					0.13	0.012
Medium dep.	NP			0.024	0.016	0.018	0
	NP-Kernel					0.008	0
Light dep.	NP			0.052	0.024	0.01	0
	NP-Kernel					0.008	0
Multivariate testing							
Strong dep.	NP			0.95	0.85	1	0.96
	NP-Kernel					0.98	0.91
	NP-BC					1	0.98
Medium dep.	NP			0.31	0.08	0.62	0.28
	NP-Kernel					0.52	0.22
	NP-BC					0.29	0.03
Light dep.	NP			0.078	0.012	0.18	0.038
	NP-Kernel					0.16	0.036
	NP-BC					0.002	0

* Since the integral-based tests did not offer significant improvements relative to the supremum-based test or the multivariate TDC-based test and because of relatively long running times for the integral-based tests, it was decided to omit these simulations for dimension $d = 5$.

Tests are conducted to evaluate the null hypothesis of tail independence with either a multivariate testing approach or a pairwise testing approach with a Hochberg correction for the p-values. Data are simulated from distributions with different dependence strengths. Strong, medium and weak tail dependence correspond to a bivariate correlation of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. Results are based on $B = 500$ samples with sample size $N = 2500$. The sample is split into two equally sized subsamples ($\theta = 1$) and $k = 1\%$ of the remaining 1250 observations are used as tail observations. All implementations of the test statistics are our own (see Appendix E.2).

Table C.4.7: Evaluation test statistics in dimension $d = 5$ for tail independent data.

		STDF test rejection rates				TDC test rejection rates	
		I_n^*		S_n		T_n	
d = 5, k = 25		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Normal							
Pairwise testing							
Strong dep.	NP			0.01	0	0.026	0.002
	NP-Kernel					0.02	0
Medium dep.	NP			0.024	0.012	0.004	0
	NP-Kernel					0.006	0
Light dep.	NP			0.044	0.014	0.006	0
	NP-Kernel					0.006	0
Multivariate testing							
Strong dep.	NP			0.56	0.23	0.86	0.55
	NP-Kernel					0.78	0.43
	NP-BC					0.74	0.55
Medium dep.	NP			0.074	0.01	0.28	0.046
	NP-Kernel					0.18	0.042
	NP-BC					0.27	0
Light dep.	NP			0.018	0.004	0.06	0.004
	NP-Kernel					0.042	0.004
	NP-BC					0	0
Frank							
Pairwise testing							
Strong dep.	NP			0.044	0.022	0.004	0
	NP-Kernel					0	0
Medium dep.	NP			0.04	0.008	0.004	0
	NP-Kernel					0.006	0
Light dep.	NP			0.066	0.024	0.008	0
	NP-Kernel					0.006	0
Multivariate testing							
Strong dep.	NP			0.044	0.01	0.16	0.026
	NP-Kernel					0.092	0.028
	NP-BC					0.99	0.99
Medium dep.	NP			0.036	0.002	0.082	0.016
	NP-Kernel					0.068	0.008
	NP-BC					0.83	0
Light dep.	NP			0.032	0.01	0.052	0.006
	NP-Kernel					0.052	0.002
	NP-BC					0	0

* Since the integral-based tests did not offer significant improvements relative to the supremum-based test or the multivariate TDC-based test and because of relatively long running times for the integral-based tests, it was decided to omit these simulations for dimension $d = 5$.

Tests are conducted to evaluate the null hypothesis of tail independence with either a multivariate testing approach or a pairwise testing approach with a Hochberg correction for the p-values. Data are simulated from distributions with different dependence strengths. Strong, medium and weak tail dependence correspond to a bivariate correlation of $\rho = 0.8$, $\rho = 0.5$ and $\rho = 0.2$, respectively. Results are based on $B = 500$ samples with sample size $N = 2500$. The sample is split into two equally sized subsamples ($\theta = 1$) and $k = 1\%$ of the remaining 1250 observations are used as tail observations. All implementations of the test statistics are our own (see Appendix E.2).

Appendix D

Additional FX results

D.1 Timeseries filtering process

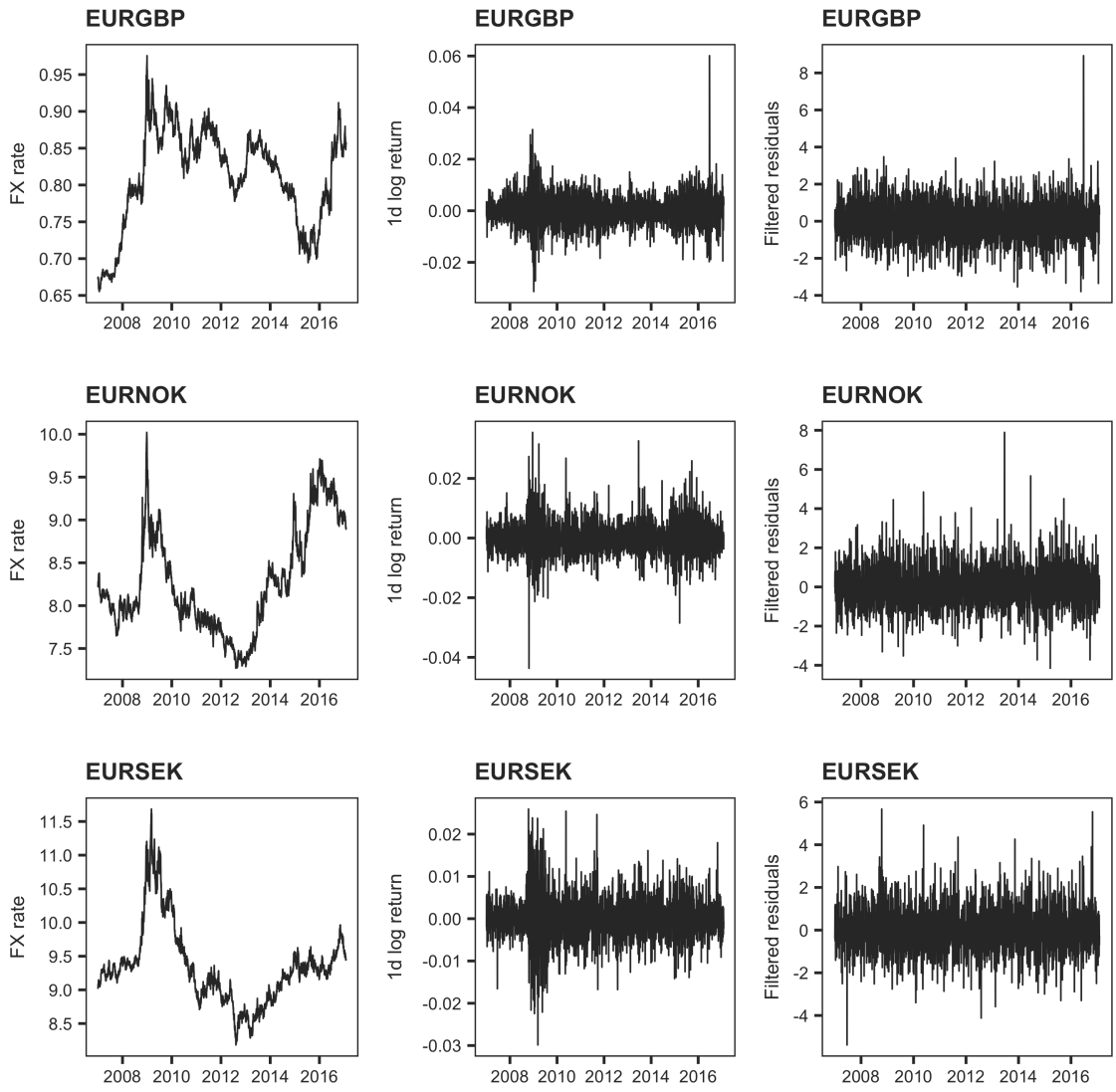

Figure D.1.1: Timeseries filtering process for three European FX rates.

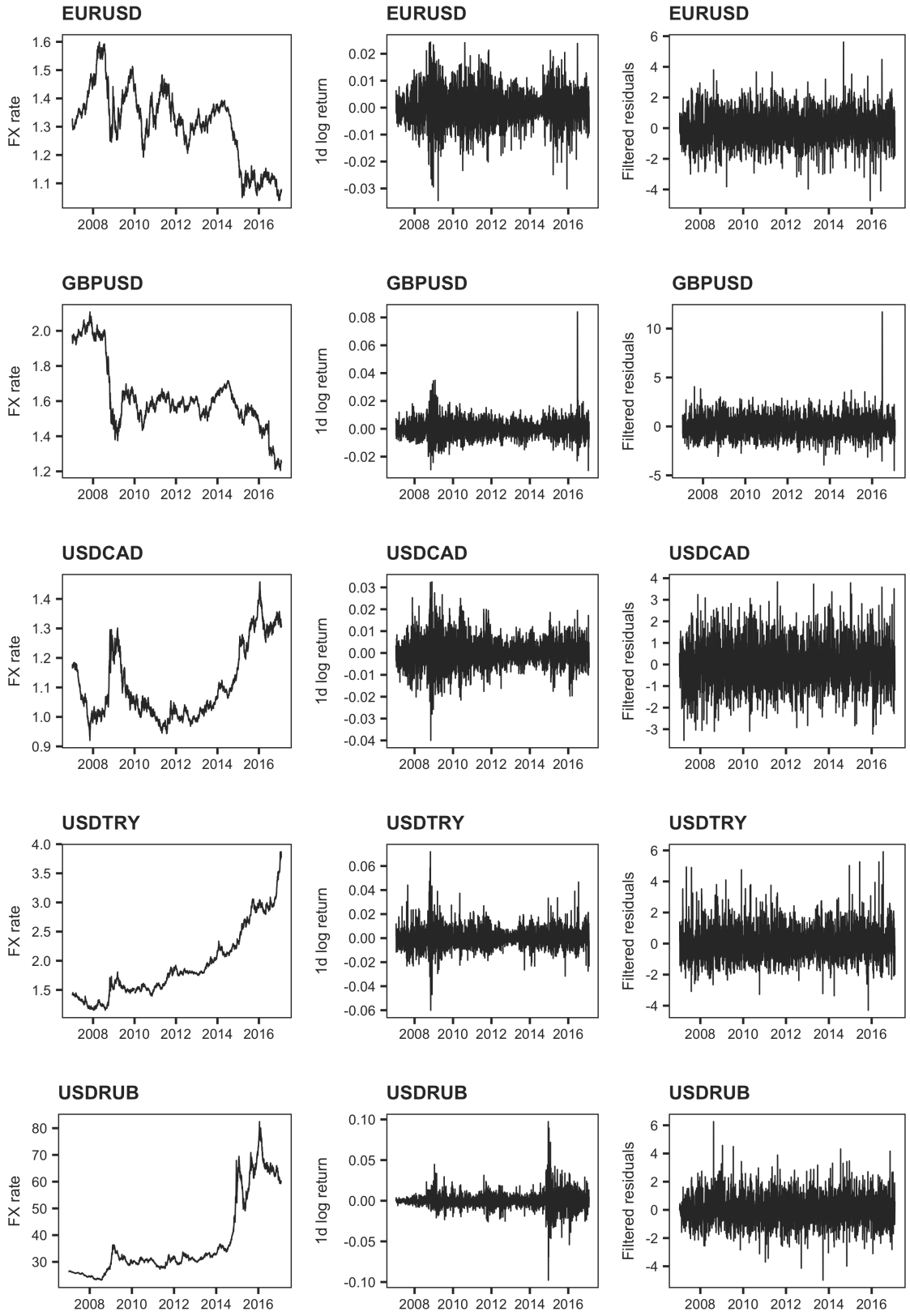

Figure D.1.2: Timeseries filtering process for five worldwide FX rates.

D.2 Test results for model assumptions

Table D.2.1: ADF-test p-values for stationarity in timeseries.

	Raw	1-day log returns	1M log returns	3M log returns	6M log returns	12M log returns
EURGBP	0.39	0.01	0.01	0.01	0.16	0.76
EURNOK	0.53	0.01	0.01	0.01	0.05	0.45
EURSEK	0.59	0.01	0.01	0.01	0.08	0.52
EURUSD	0.11	0.01	0.01	0.01	0.11	0.28
USDCAD	0.54	0.01	0.01	0.01	0.07	0.48
USDTRY	0.99	0.01	0.01	0.01	0.05	0.15
USDRUB	0.62	0.01	0.01	0.01	0.30	0.84
GBPUSD	0.67	0.01	0.01	0.03	0.33	0.61

The ADF-test (Augmented Dickey Fuller test) evaluates the alternative hypothesis of stationarity. Low p-values support the alternative hypothesis. High p-values support the null hypothesis of non-stationarity.

Table D.2.2: Goodness-of-fit tests p-values for a t-distribution of the filtered residuals.

	1-day log returns		1M log returns		3M log returns		6M log returns		12M log returns	
	KS	AD	KS	AD	KS	AD	KS	AD	KS	AD
EURGBP	0.90	0.76	0.92	0.84	0.95	0.93	0.41	0.72	0.97	0.94
EURNOK	0.16	0.06	0.73	0.56	0.85	0.87	0.44	0.73	0.94	0.89
EURSEK	0.46	0.39	0.72	0.86	0.47	0.69	0.50	0.59	0.93	0.96
EURUSD	0.16	0.23	0.88	0.83	0.62	0.80	0.58	0.48	0.37	0.41
USDCAD	0.45	0.67	0.76	0.79	0.90	0.90	0.31	0.56	0.79	0.85
USDTRY	0.13	0.02	0.03	0.01	0.95	0.96	0.96	0.95	0.98	1.00
USDRUB	0.45	0.24	0.22	0.21	0.45	0.29	0.59	0.41	0.36	0.30
GBPUSD	0.49	0.46	0.24	0.17	0.78	0.83	0.52	0.54	0.72	0.88

The Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests evaluate the null hypothesis of t-distributed residuals. The residuals are retrieved from the fitted ARMA-GRJ-GARCH model and are tested against the maximum likelihood fitted t-distribution. Low p-values support the alternative hypothesis. High p-values support the null hypothesis and indicate that residuals do follow a t-distribution.

Table D.2.3: Ljung-Box test p-values for autocorrelation in timeseries.

	1-day log returns		1M log returns		3M log returns		6M log returns		12M log returns	
	Raw	Filtered	Raw	Filtered	Raw	Filtered	Raw	Filtered	Raw	Filtered
EURGBP	0.01	0.51	0.00	0.39	0.00	0.56	0.00	0.84	0.00	0.38
EURNOK	0.51	0.25	0.00	0.91	0.00	0.43	0.00	0.27	0.00	0.66
EURSEK	0.03	0.66	0.00	0.88	0.00	0.89	0.00	0.97	0.00	0.94
EURUSD	0.54	0.51	0.00	0.83	0.00	0.89	0.00	0.95	0.00	0.42
USDCAD	0.07	0.77	0.00	0.90	0.00	0.70	0.00	0.55	0.00	0.37
USDTRY	0.16	0.57	0.00	0.36	0.00	0.55	0.00	0.91	0.00	0.86
USDRUB	0.00	0.15	0.00	0.64	0.00	0.44	0.00	0.33	0.00	0.39
GBPUSD	0.02	0.77	0.00	0.55	0.00	0.84	0.00	0.78	0.00	0.93

The Ljung-Box test evaluates the null hypothesis of temporal independence in a given timeseries. Low p-values support the alternative hypothesis and indicate temporal dependence in a timeseries. High p-values support the null hypothesis.

D.3 Transformed observations worldwide FX rates

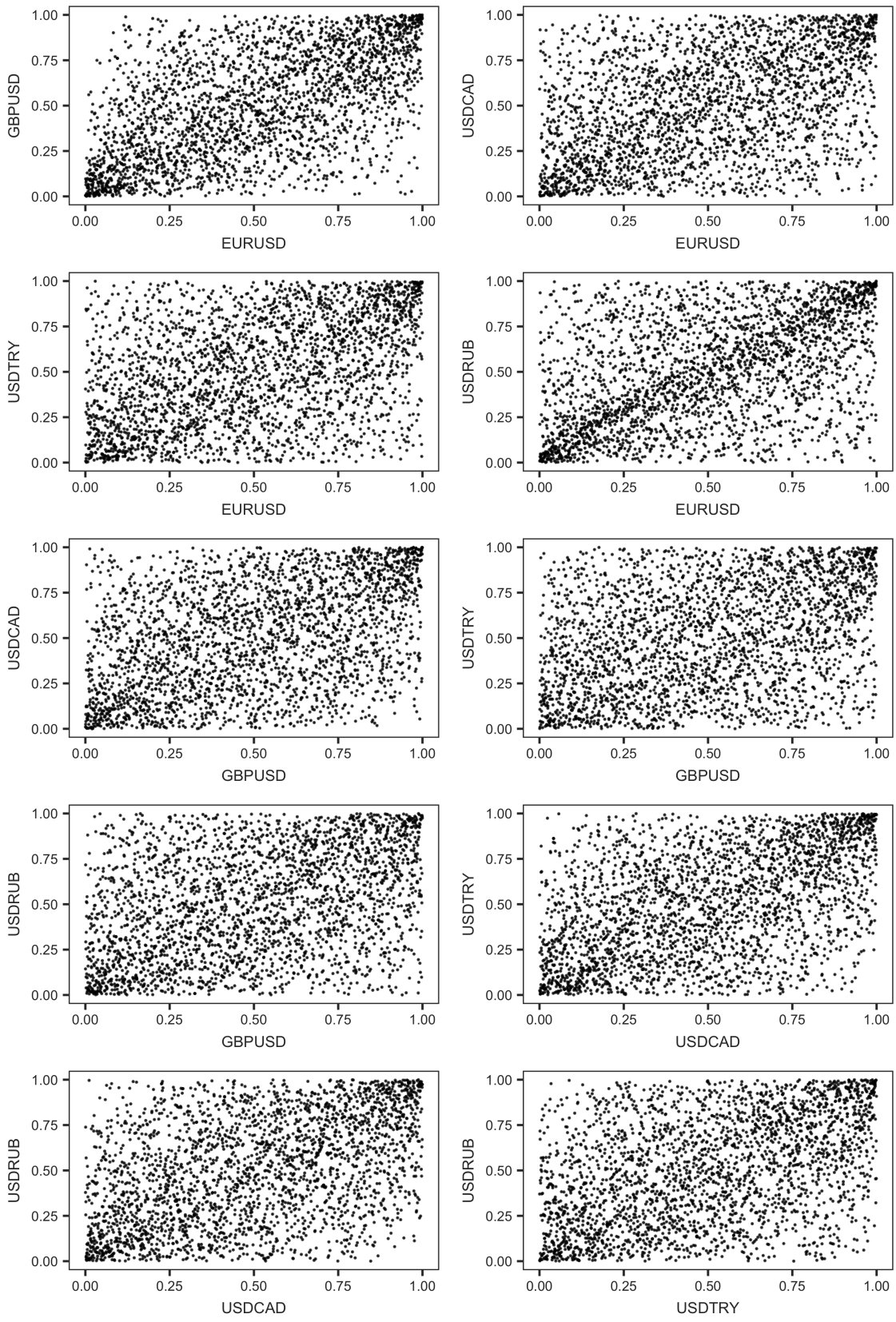

Figure D.3.1: Pairwise copula plots for five worldwide FX rates.

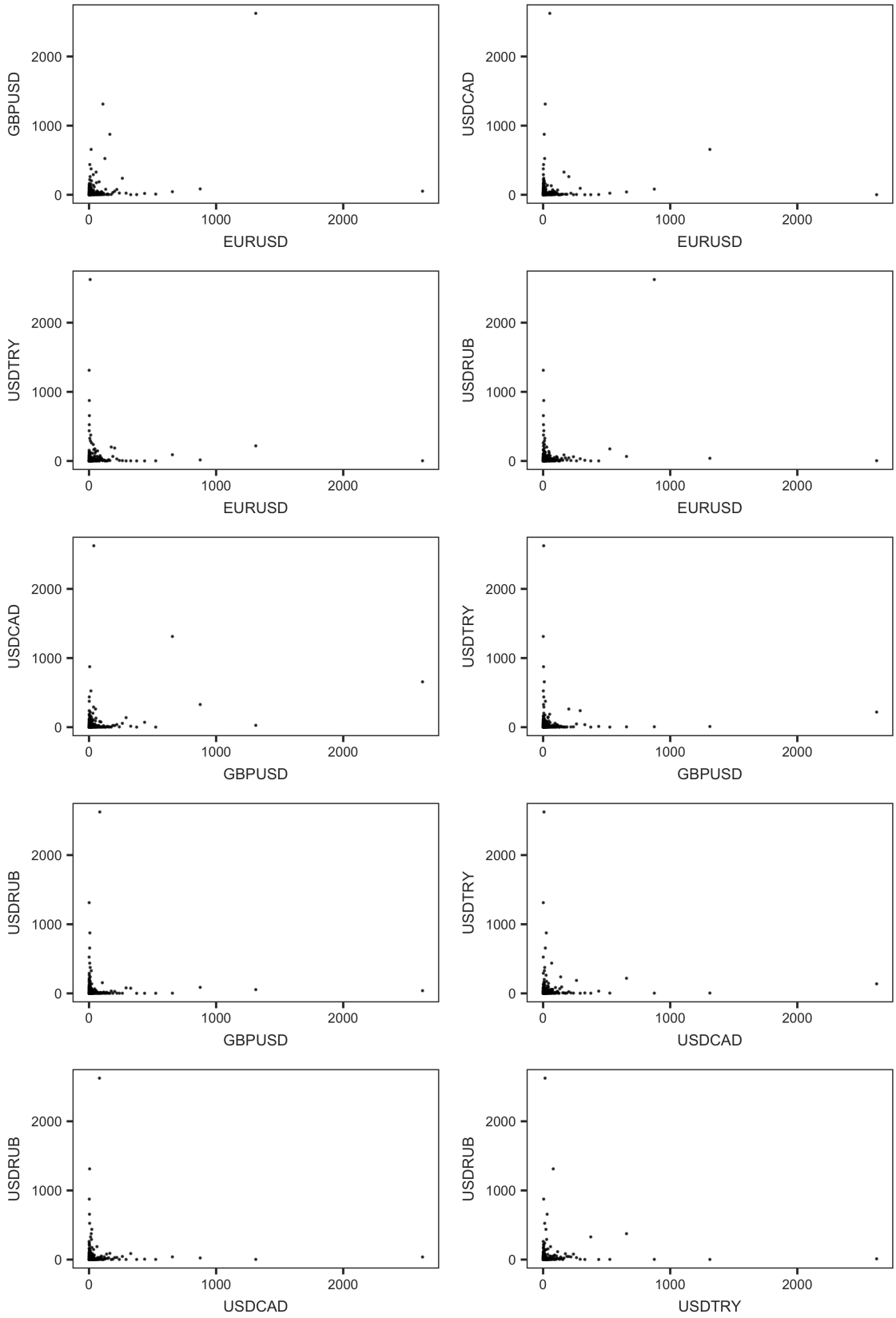

Figure D.3.2: Pairwise Pareto plots for five worldwide FX rates.

D.4 Pairwise STDF estimation worldwide FX rates

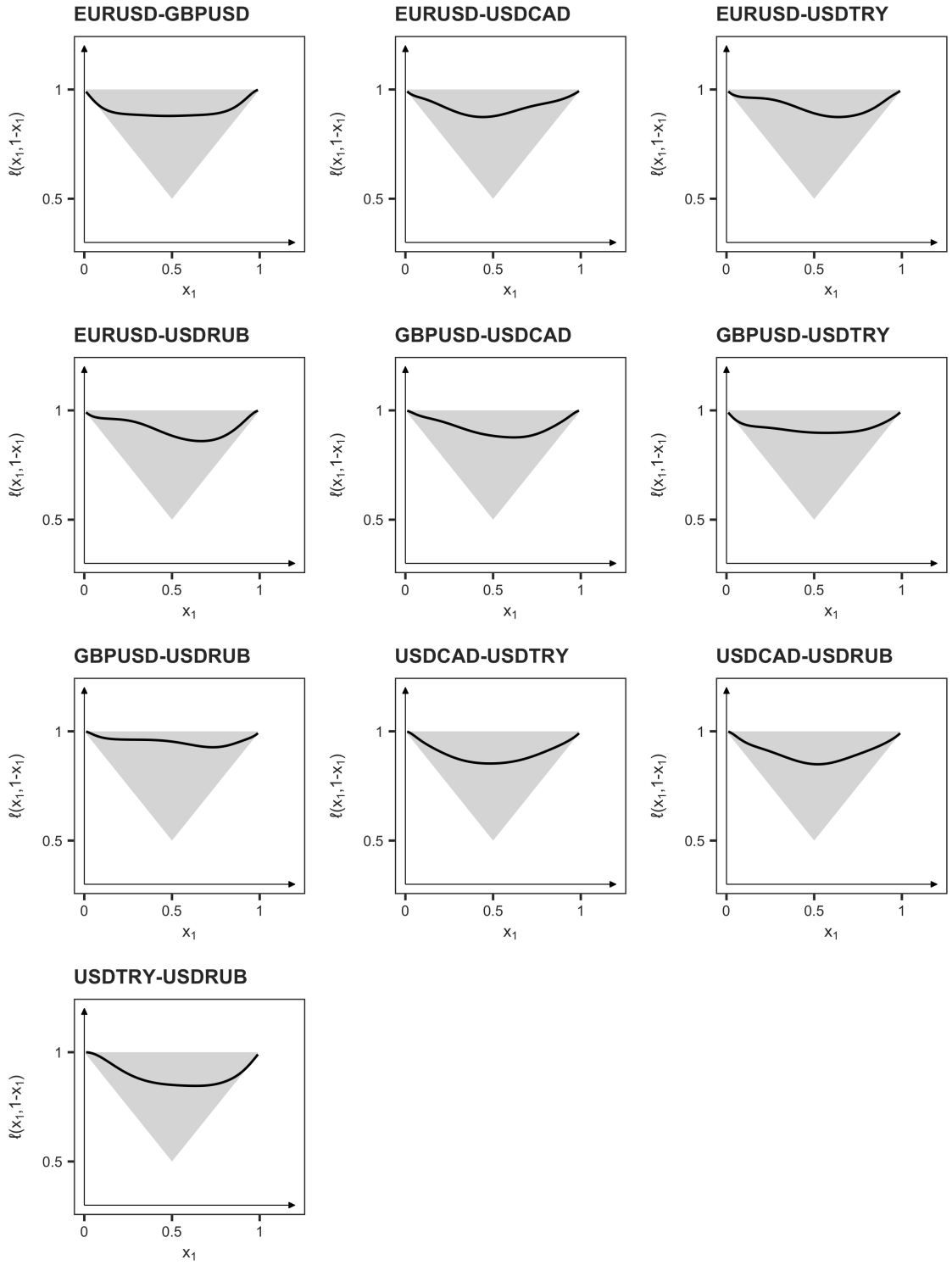

Figure D.4.1: Estimated pairwise STDFs for USD-exchange rates given a short USD position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month log returns.

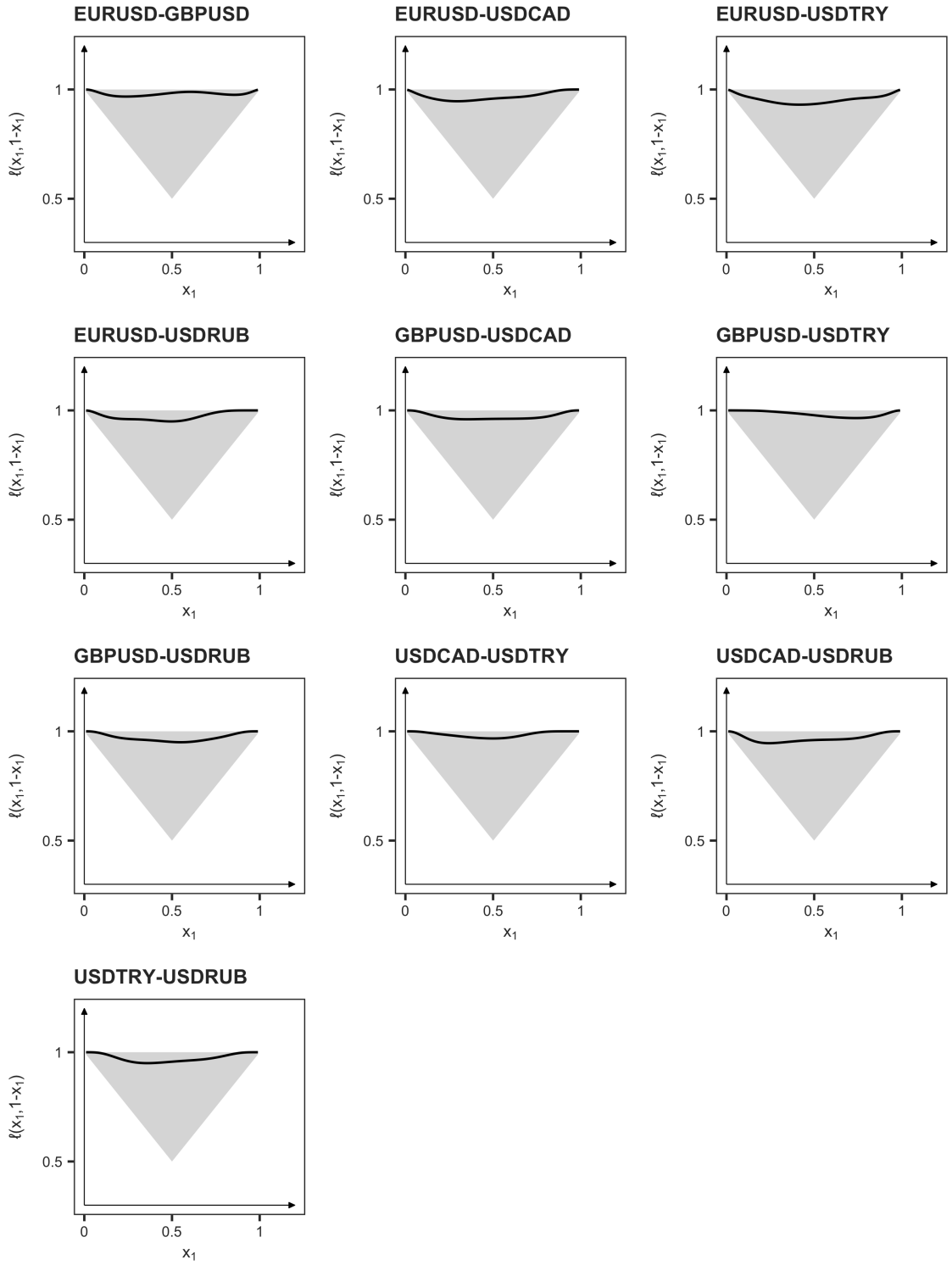

Figure D.4.2: Estimated pairwise STDFs for USD-exchange rates given a long USD position based on the beta-copula smoothed empirical STDF using $k = 1\%$ of tail observations of the standardized residuals of the filtered 1-month log returns.

Appendix E

R Codes

E.1 STDF estimators

Empirical STDF

```
ell_np_est <- function(data, x = NULL, k = 0.01){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- round(k*n)}
  df_ranked <- sapply(data, rank)
  test <- as.data.frame(matrix(NA, nrow(df_ranked), 0))
  for(i in 1:ncol(df_ranked)){
    test <- cbind(test, df_ranked[,i] > n+(1/2)-k*x[i])
  }
  ell <- sum(rowSums(test)>0)/k
  return(ell)
}
```

Beta-smoothed empirical STDF

```
ell_np_beta_est <- function(data, x = NULL, k = 0.01){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- k*n}

  ell <- (n/k)*(1-C.n(1-(k/n)*t(as.matrix(x)), as.matrix(data), smoothing = "beta",
                offset = 0, method = "C", ties.method = "average"))

  return(ell)
}
```

Kernel-smoothed empirical STDF

```
Kernel <- function(t,tau){return((tau + 1)*(t^tau))}

ell_np_kernel_est <- function(data, x = NULL, k = 0.01, tau=5){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- round(k*n)}

  a <- (1:k)/(k+1)
  s1 <- sum(sapply(a, function(aj){Kernel(aj,tau)*(1/aj)*ell_np_est(data, aj*x, k)
      }))
  s2 <- sum(sapply(a, function(aj){Kernel(aj,tau)}))
  s1/s2
}
```

Fougères et al. (2015) bias-corrected empirical STDF

```
ell_np_rescaled <- function(data, a, x, k){
  (1/a)*ell_np_est(data, x = a*x, k)
}

Delta <- function(data, a, x, k){
```

```
  ell_np_rescaled(data, a, x, k)-ell_np_est(data, x, k)
}

rho_hat <- function(data, a, r, x, k){
  res <- 1 - (1/log(r))*log(abs(Delta(data,a,r*x,k)/Delta(data,a,x,k)))
  return(ifelse(res > -0.1, -1, res)) #from Kiriliouk
  #return(min(res, 0))
}

ell_np_est_BC1 <- function(data, x = NULL, k = 0.1, k1 = (nrow(data)-10), a=0.4){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- round(k*n)}
  if(k1 < 1){k1 <- round(k1*n)}

  rhot <- rho_hat(data, a, r=0.4, x, k)
  a_temp <- (a^(-rhot)+1)^(-1/rhot)
  finalval <- ell_np_rescaled(data, a, x, k) - Delta(data, a_temp, x, k)

  #finalval
  return(ifelse(finalval < max(x), max(x), ifelse(finalval > sum(x), sum(x),
      finalval)))
}

ell_np_est_BC2 <- function(data, x = NULL, k = 0.1, k1 = (nrow(data-10)), a=0.4){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- round(k*n)}
  if(k1 < 1){k1 <- round(k1*n)}

  Delta_temp_ax <- Delta(data, a, a*x, k1)
  Delta_temp_x <- Delta(data, a, x, k1)

  finalval <- (ell_np_est(data, x, k)*Delta_temp_ax - ell_np_est(data, a*x, k)*
      Delta_temp_x)/(Delta_temp_ax-a*Delta_temp_x)

  #finalval
  return(ifelse(finalval < max(x), max(x), ifelse(finalval > sum(x), sum(x),
      finalval)))
}
```

Beirlant et al. (2016) bias-corrected empirical STDF

```
ell_np_rescaled <- function(data, a, x, k){
  (1/a)*ell_np_est(data, x = a*x, k)
}

tilde_ell <- function(data, tau, x, k){
  sum(sapply(c(1:k)/(k+1), function(i){Kernel(i,tau)*ell_np_rescaled(data, a=i, x,
      k)}))/k
}


Delta_Kernel <- function(data, a, tau, x, k){
  tilde_ell(data, tau, a*x, k)/a - tilde_ell(data, tau, x, k)
}

rho_hat_Kernel <- function(data, a, tau, r, x, k){
  res <- 1 - (1/log(r))*log(abs(Delta_Kernel(data,a,tau,r*x,k)/Delta_Kernel(data,a,
      tau,x,k)))
  return(ifelse(res > -0.1, -1, res)) #from Kiriliouk
  #return(min(res, 0))
}

alphaM_hat <- function(data, rho, x, k){
  a_seq <- (1:k)/(k+1)
  a_seq_rho <- a_seq^(-rho)

  temp1 <- sapply(a_seq, function(i){ell_np_rescaled(data, i, x, k)})
  temp2 <- sum(k*a_seq_rho*temp1)-sum(temp1*sum(a_seq_rho))
  temp3 <- sum(k*(a_seq_rho^2)) - sum(a_seq_rho*sum(a_seq_rho))

  temp2/temp3
}
```

```
ell_np_est_BC3 <- function(data, tau = 5, x = NULL, k = 0.1, k1 = nrow(data)-10){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- round(k*n)}
  if(k1 < 1){k1 <- round(k1*n)}

  rhot <- rho_hat_Kernel(data, a=0.4, tau, r=0.4, x, k1)
  alphat <- alphaM_hat(data, rhot, x, k1)

  temp1 <- sapply(c(1:k)/(k+1), function(i) Kernel(i,tau)*(i^(-rhot)))
  temp2 <- sapply(c(1:k)/(k+1), function(i) Kernel(i,tau))

  res <- tilde_ell(data,tau,x,k) - ((k1/k)^rhot)*alphat*(sum(temp1)/k)
  finalval <- res/(sum(temp2)/k)

  finalval
  #return(ifelse(finalval < max(x), max(x), ifelse(finalval > sum(x), sum(x),
      finalval)))
}
```

E.2 Testing functions

Adjusted empirical STDF

```
### Husler-Li estimator function
ell_Husler_Li <- function(data, x=NULL, theta=1, k=0.1){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- round(nrow(data)/(theta+1))
  m <- nrow(data)-n
  if(k < 1){k <- k*m}

  test <- as.data.frame(matrix(NA, n, 0))
  for(i in 1:ncol(data)){
    test <- cbind(test, data[1:n,i] > quantile(data[(n+1):(n+m),i], min((m+1-k*x[i
        ])/m, 1)))
  }
  return((1/k)*theta*sum(rowSums(test)>0))
}
```

Adjusted kernel-smoothed empirical STDF

```
Kernel <- function(t,tau){return((tau + 1)*(t^tau))}
ell_np_kernel_est2 <- function(data, x = NULL, theta=1, k = 0.1, tau=5){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- round(k*n)}

  a <- (1:k)/(k+1)
  s1 <- sum(sapply(a, function(aj){Kernel(aj,tau)*(1/aj)*ell_Husler_Li(data, aj*x,
      theta=1, k=k)}))
  s2 <- sum(sapply(a, function(aj){Kernel(aj,tau)}))
  s1/s2
}
```

Adjusted bias-corrected empirical STDF

```
Kernel <- function(t,tau){return((tau + 1)*(t^tau))}

HL_ell_np_est <- function(data, x = NULL, theta=1, k = 0.1){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- round(nrow(data)/(theta+1))
  m <- nrow(data)-n
  if(k < 1){k <- k*m}

  test <- as.data.frame(matrix(NA, n, 0))
  for(i in 1:ncol(data)){
    test <- cbind(test, data[1:n,i] > quantile(data[(n+1):(n+m),i], min((m+1-k*x[i
        ])/m, 1)))
  }
  return((1/k)*theta*sum(rowSums(test)>0))
}
```

```
ell_np_rescaled <- function(data, a, x, k){
  HL_ell_np_est(data, x = a*x, k = k)/a
}

tilde_ell <- function(data, tau, x, k){
  sum(sapply(c(1:k)/(k+1), function(i){Kernel(i,tau)*ell_np_rescaled(data, a=i, x,
      k)}))/k
}

Delta <- function(data, a, x, k){
  ell_np_rescaled(data, a, x, k)-HL_ell_np_est(data, x, k)
}

Delta_Kernel <- function(data, a, tau, x, k){
  tilde_ell(data, tau, a*x, k)/a - tilde_ell(data, tau, x, k)
}

rho_hat <- function(data, a, r, x, k){
  res <- 1 - (1/log(r))*log(abs(Delta(data,a,r*x,k)/Delta(data,a,x,k)))
  return(ifelse(res > -0.1, -1, res)) #from Kiriliouk
  #return(min(res, 0))
}

rho_hat_Kernel <- function(data, a, tau, r, x, k){
  res <- 1 - (1/log(r))*log(abs(Delta_Kernel(data,a,tau,r*x,k)/Delta_Kernel(data,a,
      tau,x,k)))
  return(ifelse(res > -0.1, -1, res)) #from Kiriliouk
  #return(min(res, 0))
}

alphaM_hat <- function(data, rho, x, k){
  a_seq <- (1:k)/(k+1)
  a_seq_rho <- a_seq^(-rho)

  temp1 <- sapply(a_seq, function(i){ell_np_rescaled(data, i, x, k)})
  temp2 <- sum(k*a_seq_rho*temp1)-sum(temp1*sum(a_seq_rho))
  temp3 <- sum(k*(a_seq_rho^2)) - sum(a_seq_rho*sum(a_seq_rho))

  temp2/temp3
}

ell_np_est_BC3 <- function(data, tau = 5, x = NULL, k = 0.1, k1 = nrow(data)/2-10){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- nrow(data)
  if(k < 1){k <- round(k*n)}
  if(k1 < 1){k1 <- round(k1*n)}

  rhot <- rho_hat_Kernel(data, a=0.4, tau, r=0.4, x, k=k1)
  alphat <- alphaM_hat(data, rhot, x, k=k1)

  temp1 <- sapply(c(1:k)/(k+1), function(i) Kernel(i,tau)*(i^(-rhot)))
  temp2 <- sapply(c(1:k)/(k+1), function(i) Kernel(i,tau))

  res <- tilde_ell(data,tau,x,k) - ((k1/k)^rhot)*alphat*(sum(temp1)/k)
  finalval <- res/(sum(temp2)/k)

  #finalval
  return(ifelse(finalval < max(x), max(x), ifelse(finalval > sum(x), sum(x),
      finalval)))
}
```

Test statistics

```
### One-point TDC test statistic based on data
Tn <- function(data = data, x=NULL, theta=1, k=0.1){
  if(is.null(x)){x <- rep(1, ncol(data))}
  n <- round(nrow(data)/(theta+1))
  m <- nrow(data)-n
  if(k < 1){k <- k*m}
  if(is.null(dim(x))){
    sqrt(k)*(ell_test_fun(data = data, x=x, k=k)-sum(x))
  } else {
    result <- c()
```

```
    for(i in 1:nrow(x)){
      result[i] <- sqrt(k)*(ell_test_fun(data = data, x=x[i,], k=k)-sum(x[i,]))
    }
    return(result)
  }
}


### Integrated test statistic based on data
In <- function(data = data, theta=1, k=0.1){
  f <- function(x){(Tn(data = data, x=x, theta=theta, k=k))^2}
  adaptIntegrate(f, lowerLimit = rep(0, ncol(data)), upperLimit = rep(1, ncol(data)
      ),
                 absError = 0.1, maxEval = 2000)$integral
}

### Supremum test statistic based on data
Sn <- function(data = data, theta=1, k=0.1){
  f <- function(x){-abs(Tn(data = data, x=x, theta=theta, k=k))}
  -bobyqa(rep(1,ncol(data)), f, lower = rep(0, ncol(data)), upper = rep(1, ncol(
      data)))$fval
}
```

References

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

Abdi, H. (2010). Holm's Sequential Bonferroni Procedure. *Encyclopedia of Research Design*, pages 1–8.

Abdous, B. and Ghoudi, K. (2005). Non-parametric estimators of multivariate extreme dependence functions. *Journal of Nonparametric Statistics*, 17(8):915–935.

Almeida, C. and Czado, C. (2010). Efficient Bayesian inference for stochastic time-varying copula models.

Almeida, C., Czado, C., and Manner, H. (2012). Modeling high dimensional time-varying dependence using D-vine SCAR models. *Pre-print*, (February):1–32.

Balkema, A. A. and Resnick, S. I. (1977). Max-Infinite Divisibility. *Journal of Applied Probability*, 14(2):309–319.

Basrak, B. and Planinić, H. (2018). A note on vague convergence of measures. pages 1–15.

BCBS (2010). Developments in modelling risk aggregation. Technical report, Basel: Bank for International Settlements.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):245–268.

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

Beirlant, J., Escobar-Bach, M., Goegebeur, Y., and Guillou, A. (2016). Bias-corrected and robust estimation of the bivariate stable tail dependence function. *Journal of Multivariate Analysis*, 26(2):284–307.

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., De Waal, D., and Ferro, C. (2004). *Statistics of Extremes: Theory and Applications*.

Berman, S. M. (1961). Convergence to bivariate limiting extreme value distributions.

Blakesley, R. E., Mazumdar, S., Dew, M. A., Houck, P. R., Tang, G., Reynolds III, C. F., and Butters, M. A. (2009). Comparisons of methods for multiple hypothesis testing. *Neuropsychology*, 23(2):255–264.

Boldi, M. O. and Davison, A. C. (2007). A Mixture Model for Multivariate Extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):217–229.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bücher, A. and Dette, H. (2013). Multiplier bootstrap of tail copulas with applications. *Bernoulli*, 19(5A):1655–1687.

Camilo, D. C., De Carvalho, M., Wadsworth, J. L., Castro Camilo, D., De Carvalho, M., and Wadsworth, J. L. (2014). Time-Varying Extreme Value Dependence With Applications To Leading European Stock Markets. (2009):1–23.

Capéraà, P. and Fougères, A.-L. (2000). Estimation of a Bivariate Extreme Value Distribution. *Extremes*, 3(4):311—-329.

Charpentier, A. and Segers, J. (2009). Tails of multivariate Archimedean copulas. *YJMVA*, 100(7):1521–1537.

Chen, S. X. and Huang, T. M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, 35(2):265–282.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley, New York.

Chollete, L., Heinen, A., and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime-switching copula. *Journal of financial econometrics*, 7(4):437–480.

Coles, S. G., Heffernan, J. E., and Tawn, J. A. (1999). Dependence Measures for Extreme Value Analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Tawn, J. A. (1991). Modelling Extreme Multivariate Events. *Source Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):377–392.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3):519–547.

Czado, C. (2010). Pair-Copula Constructions of Multivariate Copulas. In *Copula Theory and Its Applications*, pages 93–109. Springer Berlin Heidelberg.

Czado, C., Schepsmeier, U., and Min, A. (2012). Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Statistical Modelling*, 12(3):229–255.

De Haan, L. (1984). A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12(4):1194–1204.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

de Haan, L., Mercadier, C., and Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics*, 20(2):321–354.

De Haan, L., Neves, C., and Peng, L. (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis*, 99(6):1260–1275.

De Haan, L. and Omey, E. (1983). Integrals and derivatives of regularly varying function in Rd and domains of attraction of stable distributions II. *Stochastic Processes and their Applications*, 16:157–170.

De Haan, L. and Resnick, S. I. (1977). Limit Theory for Multivariate Sample Extremes. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 337:317–337.

De Haan, L. and Resnick, S. I. (1987). On regular variation of probability densities. *Stochastic Processes and their Applications*, 25(C):83–93.

De Haan, L. and Resnick, S. I. (1993). Estimating the limit distribution of multivariate extremes. *Communications in Statistics. Stochastic Models*.

De Haan, L. and Zhou, C. (2011). Extreme Residual Dependence for Random Vectors and Processes. *Advances in Applied Probability*, 43(1):217–242.

De Oliviera, T. (1958). Extremal distributions. *Rev. Fac. Cienc. Lisboa*, 3:215–227.

De Oliviera, T. (1962). Structure theory of bivariate extremes: extensions. *Estudos de Matematica, Estatistica e Econometria*, 7:165—-195.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.*, 65(5):274–292.

Della Corte, P., Sarno, L., Schmeling, M., and Wagner, C. (2016). Exchange Rates and Sovereign Risk. pages 117–129.

Draisma, G., Drees, H., Ferreira, A., and De Haan, L. (2004). Dependence in Asymptotic Independence. *Source: Bernoulli*, 10(2):251–280.

Drees, H. and Huang, X. (1998). Best Attainable Rates of Convergence for Estimators of the Stable Tail Dependence Function. *Journal of Multivariate Analysis*, 64(1):25–47.

Einmahl, J. H., De Haan, L., and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Annals of Statistics*, 34(4):1987–2014.

Einmahl, J. H., De Haan, L., and Piterbarg, V. I. (2001). Nonparametric Estimation of the Spectral Measure of an Extreme Value Distribution. *The Annals of Statistics*, 29(5):1401–1423.

Einmahl, J. H., De Haan, L., and Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Processes and their Applications*, 70(2):143–171.

Einmahl, J. H., de Haan, L., and Zhou, C. (2016a). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 78(1):31–51.

Einmahl, J. H., Kiriliouk, A., Krajina, A., and Segers, J. (2016b). An M-estimator of spatial tail dependence. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 78(1):275–298.

Einmahl, J. H., Kiriliouk, A., and Segers, J. (2017). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, pages 1–29.

Einmahl, J. H., Krajina, A., and Segers, J. (2008). A method of moments estimator of tail dependence. *Bernoulli*, 14(4):1003–1026.

Einmahl, J. H., Krajina, A., and Segers, J. (2012). An m-estimator for tail dependence in arbitrary dimensions. *Annals of Statistics*, 40(3):1764–1793.

Einmahl, J. H. and Segers, J. (2009). Maximum Empirical Likelihood Estimation of the Spectral Measure of an Extreme-Value Distribution. *The Annals of Statistics*, 37(5):2953–2989.

Embrechts, P. (2002). Extremes in Economics and the Economics of Extremes. pages 1–22.

Embrechts, P., McNeil, A. J., and Straumann, D. (1999). Correlation and dependence in risk management: properties and pitfalls.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 77(1):239–265.

Falk, M., Hüsler, J., and Reiss, R. D. (2011). *Laws of small numbers: Extremes and rare events*. Birkhauser, Basel, 3rd editio edition.

Falk, M. and Michel, R. (2006). Testing for tail independence in extreme value models. *Annals of the Institute of Statistical Mathematics*, 58(2):261–290.

Fang, K.-T., Kotz, S., and Ng, K.-W. (1992). *Symmetric Multivariate and Related Distributions*, volume 34.

Fermanian, J.-d., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860.

Fermanian, J.-d. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, 5:25–54.

Fils-villetard, A., Guillou, A., and Segers, J. (2008). Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics*, 36(3):369–382.

Fink, H., Klimova, Y., Czado, C., and Stöber, J. (2016). Regime switching vine copula models for global equity and volatility indices. pages 1–38.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press*, 24(2):180—-190.

Fougères, A.-L., De Haan, L., and Mercadier, C. (2015). Bias correction in multivariate extremes. *Annals of Statistics*, 43(2):903–934.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics*. John Wiley & Sons, Wiley Series in Probability and Mathematical Statistics, New York-Chichester-Brisbane.

Geenens, G. (2014). Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association*, 109(505):346–358.

Geffroy, J. (1958). *Contribution à la théorie des valeurs extrêmes*.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions. *Biometrika*, 82(3):543–552.

Genest, C. and Mackay, J. (1986). Copules archimedienne et familles de lois bidimensionnelles dont les marges sont donnees. *The Canadian Journal of Statistics*, 14(2):145–159.

Genest, C. and Rivest, L.-P. (1989). A characterization value distributions of gumbel's family of extreme. *Statistics & Probability Letters*, 8:207–211.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics - Theory and Methods*, 19(2):445–464.

Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423—-453.

Gudendorf, G. and Segers, J. (2009). Extreme-Value Copulas. pages 1–20.

Gudendorf, G. and Segers, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference*, 142(12):3073–3085.

Guillotte, S., Perron, F., and Segers, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(3):377–406.

Hall, P. and Tajvidi, N. (2000). Distribution and Dependence-Function Estimation for Bivariate Extreme-Value Distributions. *Bernoulli*, 6(5):835.

Hartmann, P., Straetmans, S., and Vries, C. D. (2006). Banking system stability: a cross-Atlantic perspective.

Hashorva, E. (2006). On the multivariate Hüsler–Reiss distribution attracting the maxima of elliptical triangular arrays. *Statistics & Probability Letters*, 76:2027–2035.

Hashorva, E. (2010). On the residual dependence index of elliptical distributions. *Statistics and Probability Letters*, 80(13-14):1070–1078.

Heffernan, J. E. (2000). A Directory of Coefficients of Tail Dependence. *Extremes*, 3(3):279—-290.

Hobæk Haff, I., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction - Simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.

Hua, L. and Joe, H. (2011). Tail order and intermediate tail dependence of multivariate copulas. *Journal of Multivariate Analysis*, 102(10):1454–1471.

Huang, X. (1992). *Statistics of bivariate extreme values*. Phd thesis, Tinbergen Institute Rotterdam.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34:587–608.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown – Resnick process. *Biometrika*, 100(1):1–9.

Huser, R. and Genton, M. G. (2016). Non-Stationary Dependence Structures for Spatial Extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(3):470–491.

Hüsler, J. and Li, D. (2009). Testing asymptotic independence in bivariate extremes. *Journal of Statistical Planning and Inference*.

Hüsler, J. and Reiss, R. D. (1989). Maxima of bivariate random vectors: Between independence and complete dependence. *Statistics and Probability Letters*, 21(5):385–394.

Hutchinson, T. and Lai, C. (1990). *Continuous Multivariate Distributions, Emphasising Applications*. Rumsby Publishing Co., Adelaide.

Joe, H. (1993). Parametric families of multivariate distributions with given margins.

Joe, H. (1997). *Multivariate models and dependence concepts*. Number 1960.

Joe, H. (2015). Dependence Modeling with Copulas. *CRC Press*, page 479.

Joe, H. and Hu, T. (1996). Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, 57(2):240–265.

Joe, H., Li, H., and Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270.

Killiches, M., Kraus, D., and Czado, C. (2017). Examination and visualisation of the simplifying assumption for vine copulas in three dimensions. *Australian and New Zealand Journal of Statistics*, 59(1):95–117.

Kiriliouk, A. (2016). Vignette for the tailDepFun package. Technical report.

Kiriliouk, A., Segers, J., and Tafakori, L. (2018). An estimator of the stable tail dependence function based on the empirical beta copula. *Extremes*.

Kiriliouk, A., Segers, J., and Warchol, M. (2014). Nonparametric estimation of extremal dependence. pages 1–22.

Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. World Scientific.

Kurowicka, D. and Cooke, R. M. (2006). *Uncertainty Analysis With High Dimensional Dependence Modelling*.

Larsson, M. and Nešlehová, J. (2011). Extremal Behavior of Archimedean Copulas. 43(1):195–216.

Leadbetter, M. R. and Rootzén, H. (1988). Extremal Theory for Stochastic Processes. *The Annals of Probability*, 16(2):431–478.

Lehmann, E. L. (1966). Some concepts of dependence. *Statistics (Ber).*, 37(3):688–697.

Li, H. (2009). Orthant tail dependence of multivariate extreme value distributions. *Journal of Multivariate Analysis*, 100(1):243–256.

Li, H. (2013). Toward a Copula Theory for Multivariate Regular Variation. *Lecture Notes in Statistics-Proceedings*, (July):294.

Li, H. and Hua, L. (2015). Higher order tail densities of copulas and hidden regular variation. *Journal of Multivariate Analysis*, 138(July):143–155.

Li, H. and Wu, P. (2013). Extremal dependence of copulas: A tail density approach. *Journal of Multivariate Analysis*, 114(1):99–111.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, 56(2):649–676.

Mai, J. F. and Scherer, M. (2014). *Financial Engineering with Copula Explained*. Palgrave Macmillan, London.

Manner, H. and Reznikova, O. (2012). A survey on time-varying copulas: Specification, simulations, and application. *Econometric Reviews*, 31(6):654–687.

Marcon, G., Padoan, S. A., Naveau, P., Muliere, P., and Segers, J. (2017). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17.

Mardia, K. (1970). *Families of bivariate distributions*. Griffin's Statistical Monographs, Griffin, London.

Marshall, A. W. and Olkin, I. (1983). Domains of Attraction of Multivariate Extreme Value Distributions. *The Annals of Probability*, 2(5):347–370.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7(3-4):271–300.

McNeil, A. J. and Nešlehová, J. (2009). Multivariate archimedean copulas, d-monotone functions and l1-norm symmetric distributions. *Annals of Statistics*, 37(5B):3059–3097.

Molchanov, I. (2008). Convex geometry of max-stable distributions. *Extremes*, 11(3):235–259.

Nelsen, R. B. (2006). *An Introduction to Copulas*, volume 42. Springer Series in Statistics, New York, 2nd ed. edition.

Niculescu, C. P. and Persson, L.-E. (2004). Convex Functions and Their Applications: A Contemporary Approach.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes*.

Oakes, D. (1982). A model for association in bivariate survival data. *J.R. Statist. Soc. B*, 44:412–422.

Padoan, S., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Patton, A. J. (2013). Copula methods for forecasting multivariate time series. *Handbook of Economic Forecasting*, (May).

Peng, Y. and Lon, W. (2012). Analysing financial contagion and asymmetric market dependence with volatility indices via copulas. *Annals of Finance*, 8(1):49–74.

Pickands, J. (1981). Multivariate extreme value distributions. *Proceedings 43rd Session International Statistical Institute*, 2:859–878.

Puccetti, G. and Rüschendorf, L. (2012). Bounds for joint portfolios of dependent risks. *Statistics and Risk Modeling*, 29(2):107–132.

Puccetti, G., Rüschendorf, L., Small, D., and Vanduffel, S. (2017). Reduction of Value-at-Risk bounds via independence and variance information. *Scandinavian Actuarial Journal*, 2017(3):245–266.

Ranaldo, A. and Söderlind, P. (2010). Safe haven currencies. *Review of Finance*, 14(3):385–407.

Rényi, A. (1959). On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungaricae*, 10(3-4):441–451.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*.

Resnick, S. I. (2004). On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability*, 41 A(SPEC. ISSUE):191–212.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

Ressel, P. (2013). Homogeneous distributions-And a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis*, 117:246–256.

Rvačeva, E. L. (1962). On domains of attraction of multi-dimensional distributions. In *Selected Transl. Math. Statist. Probab.*, pages 183–205.

Sancetta, A. and Satchell, S. (2004). The Bernstein Copula and Its Applications to Modeling and Approximations of Multivariate Distributions. *Econometric Theory*, 20(3):535–562.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156.

Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research*, 55(2):301–327.

Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2):307–335.

Schoenberg, I. J. (1938). Metric Spaces and Completely Monotone Functions. *Annals of Mathematics*, 39(4):811–841.

Schweizer, B. and Wolff, E. F. (1981). On Nonparametric Measures of Dependence for Random Variables. 9(4):879–885.

Schweizer, R. and Sklar, A. (1983). *Probabilistic Metric Spaces*. North-Holland Publishing, New York.

Segers, J. (2012a). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782.

Segers, J. (2012b). Max-stable models for multivariate extremes. *REVSTAT - Statistical Journal*, 10:61–82.

Segers, J., Sibuya, M., and Tsukahara, H. (2017). The Empirical Beta Copula. *Journal of Multivariate Analysis*, 155:35–51.

Sibuya, M. (1960). Bivariate Extreme Statistics, I.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Smith, R. L. (1994). Multivariate threshold methods. *Extreme Value Theory and Applications*, 1(7):225–284.

Stöber, J. and Czado, C. (2014). Detecting regime switches in the dependence structure of high dimensional financial data. *Computational Statistics & Data Analysis*, 76:672–686.

Stöber, J., Joe, H., and Czado, C. (2013). Simplified Pair Copula Constructions — Limits and Extensions. *Journal of Multivariate Analysis*, 119:101–118.

Stute, W. (1984). The Oscillation Behavior of Empirical Processes: The Multivariate Case. *The Annals of Probability*, 12(2):361–379.

Tawn, J. A. (1988). Bivariate Extreme Value Theory: Models and Estimation. *Biometrika*, 75(3):397–415.

Tawn, J. A. and Ledford, A. W. (1996). Statistics for Near Independence in Multivariate Extreme Values. *Biometrika*, 83(1):169–187.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer Series in Statistics, New York.

Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 23(4):245–253.

Vettori, S., Huser, R., and Genton, M. G. (2017). A comparison of dependence function estimators in multivariate extremes. *Statistics and Computing*, pages 1–14.

Wang, Z. R., Chen, X. H., Jin, Y. B., and Zhou, Y. J. (2010). Estimating risk of foreign exchange portfolio: Using VaR and CVaR based on GARCHEVT-Copula model. *Physica A: Statistical Mechanics and its Applications*, 389(21):4918–4928.

Wen, K. and Wu, X. (2018). Transformation-Kernel Estimation of Copula Densities. *Journal of Business & Economic Statistics*, pages 1–36.

Wystup, U. (2006). *FX Options and Structured Products.* John Wiley & Sons, Ltd., London.