On Dependence and Extremes

Gabriel Kuhn

Center for Mathematical Sciences

Munich University of Technology

85747 Garching bei München, Germany

2006

Zentrum Mathematik

Lehrstuhl für Mathematische Statistik
der Technischen Universität München

ON DEPENDENCE AND EXTREMES

GABRIEL KUHN

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München genehmigten Dissertation zur Erlangung des akademischen Grades eines

Doktor der Naturwissenschaften (Dr. rer. nat.)

Vorsitzender		Univ.-Prof. Dr. Rudi Zagst
Prüfer der Dissertation	1.	Univ.-Prof. Dr. Claudia Klüppelberg
	2.	Univ.-Prof. Dr. Michael Falk, Bayerische Julius-Maximiliansuniversität Würzburg

Die Dissertation wurde am 01. Februar 2006 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 22. Mai 2006 angenommen.

Nichts ist, was nicht schon immer war.
Teilhard de Chardin

Um sein Nichtwissen wissen, ist das Höchste.

Nicht wissen, was Wissen ist, ist Leiden.
Eben wenn man an diesem Leiden leidet,
So leidet man dadurch nicht mehr.

Der Berufene leidet nicht.
Weil er an seinem Leiden leidet,
Darum leidet er nicht.
Laotse

Abstract

This thesis deals with various topics on multivariate dependence structures and extremes.

The first chapter investigates nonparametric estimation of multivariate extremes, where a new dependence function is developed, which allows for an easy understanding of multivariate extreme dependence. An additional focus there is the visualization of multivariate extremes and a new concept is introduced. In contrast to many articles dealing with 'multivariate extreme dependence' only in the bivariate situation, we extend the estimation procedure and dependence function to arbitrary high dimensions.

A problem arising when nonparametrically estimating multivariate extremes in higher dimensions is instability, hence there is an interest in flexible and finitely parameterized distribution classes; elliptical distributions and copulae are. Chapter 3 develops an estimator of the tail copula (measuring extreme dependence) under the assumption of data with an elliptical distribution. After deriving its asymptotic properties it is compared to the nonparametric estimator and the improvement is shown. Chapter 4 develops a tail copula estimator under the weaker assumption of only having data with an elliptical copula. There, also the tail copula of an elliptical copula with arbitrary dimension is shown together with a three-dimensional estimation example.

A prominent question concerning dependence structures is how to model and interpret it. And how to reduce dimensions. A classical tool in multivariate statistics is correlation structure analysis, and a linear structure of the data is assumed. Chapter 5 extends the method of correlation structure analysis to copulae whereby the main drawbacks of having a linear structure, same types of margins or existence of moments can be avoided. This approach is extended to extremes, where the use of elliptical copulae allows for dimension reduction and an interpretation of the dependence structure in the extremes. In a factor analysis setting, the theoretical results of the new estimators are verified, also showing an improvement in the performance in comparison to existing methods.

Concerning the influence of different dependence structures to some (financial) outcome, Chapter 6 gives an example of a portfolio of credit defaults. Using a one-factor model with different underlying distributions the limiting extreme value distribution of the portfolio is determined. A simulation study then shows large differences in the portfolio outcomes when using different (but with similar properties) dependence structures.

Furthermore, throughout this thesis, all of the developed methods and procedures are applied to financial data and their possible use for risk management is explained.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit unterschiedlichen Aspekten von multivariaten Abhängigkeitsstrukturen und Extremwerten.

Das erste Kapitel behandelt die nichtparametrische Schätzung multivariater Extrema und entwickelt hierfür ein neues Abhängigkeitsmass und ein Konzept zur Visualisierung mehrdimensionaler Extrema. Ein Problem beim nichtparametrischen Schätzen in hohen Dimensionen ist Instabilität, folglich besteht ein Interesse an flexiblen und endlich parametrisierbaren Verteilungsklassen. In Kapitel 2 wird ein Schätzer für die Tail Copula unter elliptischen Verteilungen entwickelt, sowie dessen asymptotischen Eigenschaften. Diese werden dann sowohl theoretisch als auch mittels Simulation mit dem empirischen Schätzer verglichen. Ähnlich wird in Kapitel 3 ein Tail Copula Schätzer entwickelt unter der schwächeren Annahme einer elliptischen Copula. Ebenso wird dort die Tail Copula einer Elliptischen Copula in beliebigen Dimensionen bestimmt und in einer ausführlichen Simulationsstudie wird der neue Schätzer mit dem empirischen verglichen. Kapitel 4 zeigt eine Erweiterung der Korrelations-Struktur-Analyse, indem Verteilungsgleichheit der Daten mit einem linearen Modell durch die schwächere Voraussetzung der Copula-Gleichheit ersetzt wird. Dazu werden neue, Copula basierte Schätzer entwickelt, ihre asymptotischen Eigenschaften gezeigt und in einer Simulationsstudie deren asymptotisches Verhalten überprüft. Im letzten Kapitel 5 wird die Extremwertverteilung eines Credit-Default Portfolios für unterschiedliche unterliegende Verteilungen bestimmt. Dabei stellt sich heraus, dass das Verhalten 1. Ordnung gleich ist und lediglich im langsam variierenden Anteil tauchen Unterschiede auf. Ebenfalls wird ein verbessertes Verfahren zur Anpassung der Portfolioverteilung gezeigt.

Acknowledgement

First of all, I would like to thank my advisor Prof. Claudia Klüppelberg for having confidence in me over all the years, for the fruitful discussions, the support and for spurring me again and again.

Further, thanks to Prof. Tailen Hsing for his help at the beginning of my PhD-time to get in touch with my main research topic 'multivariate extremes'. Another special thanks goes to Prof. Liang Peng for his stimulating discussions and his collaboration helping me to learn a lot and to solve many problems.

I would like to thank my colleagues at Prof. Claudia Klüppelberg's research group for having an enjoyable life at the department.

Last but not least, I thank my family and friends for their support, especially during the last months helping me to keep my head free for work.

My work was financially supported by the SFB386 of the Deutsche Forschungsgemeinschaft.

Contents

Introduction

Before we give an overview of the thesis, we first give a short review on the topics that can be found in this thesis.

Copulae

The concept of copulae provides a class of distribution functions which completely describe the dependence structure of an arbitrary random vector. In particular, a copula is a distribution function $C : [0,1]^d \mapsto [0,1]$ with standard uniform margins. *Sklar's Theorem* (see Sklar (1996)) shows that each distribution function can be separated in its univariate marginal distributions and its copula. This means, for a d-dimensional distribution function F with univariate margins F_j, $1 \leq j \leq d$, there exists a copula C_F (unique on the support of F) such that

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

for all $x_1, \ldots, x_d$ in the support of F. Therefore, copulae allow to study the dependence structure of an arbitrary random vector independent of its margins, and they can also be used to construct multivariate distributions. For a textbook treatment of copulae, see Nelsen (1999).

Elliptical distributions

When dealing with dependence structures, one is often interested in a class of distributions, where dependence can be modeled flexible and easy. A widely used class of flexible distributions is given by the *elliptical distributions*. By Fang, Kotz, and Ng (1990, Corollary 1), a random vector $\boldsymbol{X}$ is elliptically distributed if and only if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^d$, a matrix $\mathbf{A} \in \mathbb{R}^{k \times d}$, a positive random variable $G > 0$ and a random vector

$\boldsymbol{U}^{(k)} \sim \text{unif}\{\boldsymbol{s} \in \mathbb{R}^k : \boldsymbol{s}^T\boldsymbol{s} = 1\}$, independent of G, such that

$$\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + G\mathrm{A}^T\boldsymbol{U}^{(k)}.$$

The best well-known elliptical distribution is the normal distribution, where $G \overset{\mathrm{d}}{=} \sqrt{C}$ and C is a χ^2-distributed random variable with d degrees of freedom. Of course, these distributions have some drawbacks, i.e. they are symmetric, and all margins belong to the same class of distributions. Therefore, if someone is only interested in the dependence structure of a random vector, an elliptical copula being the copula of an elliptical distribution can be considered instead of the full distribution.

Extreme value theory

In statistics, extreme value theory offers methods to describe rare events of extreme random outcomes. Univariate extreme value theory starts with the asymptotic theory for maxima

$$M_1 = X_1 \quad \text{and} \quad M_n = \bigvee_{i=1}^{n} X_i, \quad n > 1,$$

where $X_1, \ldots, X_n$ are iid random variables with distribution function F. The *extremal types theorem* exhibits the possible limit distributions of the suitable normalized maxima. Assume that sequences $a_n > 0$ and $b_n \in \mathbb{R}$ exist such that

$$P\left(a_n^{-1}(M_n - b_n) \leq x\right) \overset{n\to\infty}{\longrightarrow} F(x), \quad x \in \mathbb{R},$$

and F is non-degenerate. Then there exist constants $a > 0$ and $b \in \mathbb{R}$ such that $F(ax+b)$ is one of the three extreme value distributions. These are the *Fréchet* distribution $\Phi_\alpha(x) = \exp(-x^{-\alpha})\mathbf{1}_{\{x>0\}}$, $\alpha > 0$, the *Weibull* distribution $\Psi_\alpha(x) = \mathbf{1}_{\{x>0\}} + \exp(-(-x)^\alpha)\mathbf{1}_{\{x\leq 0\}}$, $\alpha > 0$, and the *Gumbel* distribution $\Lambda(x) = \exp(-e^{-x})$. For a textbook treatment of univariate extremes, see Embrechts, Klüppelberg, and Mikosch (1997, Chapter 3).

Switching to the multivariate setting we consider an iid sequence of d-dimensional random vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^T$, $1 \leq i \leq n$. Then, multivariate extreme value theory considers the limiting distribution of the componentwise maxima

$$\begin{aligned} \boldsymbol{M}_1 &= (X_{1,1}, \ldots, X_{1,d})^T \quad \text{and} \\ \boldsymbol{M}_n &= (M_{n,1}, \ldots, M_{n,d})^T = \left(\bigvee_{i=1}^{n} X_{i,1}, \ldots, \bigvee_{i=1}^{n} X_{i,d}\right)^T, \quad n > 1. \end{aligned}$$

Similarly to the univariate case, assume that sequences $a_{n,j} > 0$ and $b_{n,j} \in \mathbb{R}$, $1 \le j \le d$, exist such that

$$P\left(a_{n,j}^{-1}(M_{n,j} - b_{n,j}) \le x_j,\, 1 \le j \le d\right) \stackrel{n\to\infty}{\longrightarrow} F(x_1,\ldots,x_d) \;=:\; F(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^d,$$

and F is non-degenerate. Then, F is a multivariate extreme value distribution if and only if the univariate margins of F are univariate extreme value distributions and the copula C_F of F satisfies

$$C_F(\boldsymbol{u}) \;=\; \exp\left(\int_{\mathcal{S}_d}\left(\bigwedge_{j=1}^{d} w_j(\ln u_j)\right)\mu(d\boldsymbol{w})\right), \quad \boldsymbol{u} \in [0,1]^d,$$

where $\mathcal{S}_d = \{\boldsymbol{s} \ge \boldsymbol{0} : \|s\| = 1\}$, $\|\cdot\|$ is an arbitrary norm and μ is a finite measure satisfying

$$\int_{\mathcal{S}_d} w_j\,\mu(d\boldsymbol{w}) = 1, \quad 1 \le j \le d.$$

The measure μ can be interpreted as the dependence measure, i.e. it shows, where the extremes are to be found. For more theoretical background, see Resnick (1987).

There exist many equivalent representations of the dependence measure μ and one of them we call the *tail copula* $\lambda(\boldsymbol{x})$, i.e.

$$\lambda(\boldsymbol{x}) \;:=\; \lim_{t\to 0}\frac{1}{t}P\left(1 - F_1(X_1) \le tx_1, \ldots, 1 - F_d(X_d) \le tx_d\right), \quad \boldsymbol{x} > \boldsymbol{0},$$

where F_j denote the univariate marginal distributions of $\boldsymbol{X}$. Statistical questions about multivariate extremes concern the appropriate choice of a dependence measure, the estimation of this measure and its interpretation. In this setting, bivariate extremes have been intensively investigated. However, there are many open questions when multivariate extremes for $d \ge 3$ or, even more realistically, high dimensions are considered. Many of the procedures for the bivariate case do not work properly in larger dimensions, e.g. estimators become unstable, expensive to compute and their interpretation becomes difficult. An alternative question deals with (semi)parametric distribution models for describing extreme dependence, i.e. one aims at a multivariate model, which is flexible enough to describe a complex extreme dependence structure, but is also easy to handle. Again, many models from the bivariate setting have problems, when they are extended to higher dimensions: they are either not flexible enough or not easy to handle. One compromise offers the class of elliptical copulae having dependence parameters for each bivariate margin and are quite easy to handle. One drawback of this class is that it is not possible to model asymmetric extreme dependence structures.

Overview of the thesis

This thesis consists of five chapters based on the articles Hsing, Klüppelberg, and Kuhn (2004), Klüppelberg, Kuhn, and Peng (2005a, 2005b), Klüppelberg and Kuhn (2006) and Kuhn (2005). Chapter 1 develops a new extreme dependence measure, a method for visualization of multivariate extremes. Chapters 2 and 3 develop new estimators of the tail copula under elliptical distributions and elliptical copulae, respectively, and also investigates the second order behavior. Chapter 4 extends correlation structure analysis to copula structure analysis and shows two different methods of fitting an elliptical copula to data, one of them based on extreme observations. Finally, Chapter 5 is devoted to credit default portfolios and shows the extremal behavior of the limiting portfolio for a large class of underlying 1-factor models.

In the following, we present a guideline to the thesis:
Chapter 1. In Section 1.1 the framework of multivariate extreme value theory is introduced and some equivalent representations of a multivariate extreme value distribution are given. Section 1.2 shows in the bivariate case how to measure extreme sets and also gives a nonparametric estimator for these extreme sets. To estimate an extreme dependence measure completely it suffices to estimate this measure for set in a measure determining class and two of such classes are provided in Section 1.3. Next, in Section 1.4, the new extreme dependence measure 'tail dependence function' is defined. This measure captures the complete extreme dependence of a random vector and properties, interpretations and an estimator are given for the bivariate case. In a simulation study the good performance of the estimator is shown. For visualization of extreme dependence we suggest to plot the reciprocal of the ranks of the data. Using this method, large values become visible and dependent extremes are to be found in the middle of the plot. Further, extremes of one component being independent from the other components will be close on the axes. Section 1.5 extends the results of the previous section to the general multivariate case and a simulation study shows a good performance of the estimator in a three-variate example. Finally, in Section 1.6 the procedures are applied to a financial data set of swap rates.

Chapter 2. This chapter deals with the extremes of bivariate elliptical distributions. First, in Section 2.1, the basic notions of a tail copula, elliptical distributions and an empirical estimator of extreme dependence is given. In Section 2.2 the main results are given and a new estimator for the tail copula under elliptical distributions developed. Under elliptical distributions it turns out that the tail copula is determined by the copula correlation and α, the index of regular variation of the generating variate. Therefore, this new tail copula estimator is based on the estimation of these parameters. To compare the new estimator with the existing nonparametric estimator, the second order behavior of both estimators are calculated explicitly. Further, the optimal asymptotic mean squared error is determined. In Section 2.3, asymptotic variance and asymptotic optimal mean squared error of both estimators are compared in an example and together with a simulation study it turned out that the new estimator always performs better than the standard empirical estimator. In Section 2.4, all proofs of this chapter are given.

Chapter 3. Similarly to the previous chapter, a new estimator for the tail copula is developed whereas the assumption of an elliptical distribution is replaced by the weaker assumption of an elliptical copula and in Section 3.1 the basic notations are given. In Section 3.2 the new estimator is shown together with theoretical results about asymptotic and second order behavior. However, contrary to the previous chapter, the index α of regular variation of the generating variate cannot be observed from the data directly. Therefore, α is estimated by inverting the theoretical tail copula and using the empirical tail copula. This estimation is done for all possible directions and by smoothing over all directions, the final estimator of α is obtained. Then, the new tail copula estimator is the value of the theoretical tail copula calculated from the estimated index of regular variation and copula correlation. In Section 3.3, the asymptotic variance and mean squared error of the new and the empirical tail copula are calculated and in Section 3.4 a simulation study is conducted to compare both estimators. It turns out that except from small areas, the new estimators performs better. Since both estimators are based on the k largest order statistics an unstable behavior can be observed when different k's or different directions are considered. There, the new estimator performs much better than the empirical one, i.e. it is smooth with respect to different directions an smoother than the empirical estimator with respect to a different number of order statistics k. Finally, in Section 3.5 the theoretical elliptical tail copula and its estimator is extended to the arbitrary multivariate setting and also a simulation example is given. Since the bivariate results from the previous sections in this chapter do not rely on two dimensions, all results also hold in the multivariate case. The proofs of this chapter are given in Section 3.6.

Chapter 4. This chapter is devoted to dimension reduction and Section 4.1 introduces the settings. The classical approach of decomposing a dependence structure to reduce dimensions or to understand the dependence is done by correlation structure analysis. There, a linear model is assumed to hold for the observed data. This model describes the correlation matrix as a function of some lower dimensional parameter vector. The drawbacks of these structure models are the fact that only linear dependence can be modeled, only similar classes of marginal distributions are admissible and that the existence of the 4th moment of the data is required. To overcome these restrictions, we use elliptical copulae having a correlation matrix as dependence parameter. The basic notations of elliptical copulae are given in Section 4.2 and the copula structure model is introduced in Section 4.3.1. Section 4.3.2 and 4.3.3 show how to estimate the parameters and select a proper model using some test statistic, respectively. For the latter procedures an estimator of the copula correlation matrix and an estimator of its asymptotic covariance matrix is needed. Section 4.4.1 introduces two dependence concepts, where the copula correlation matrix can be obtained from, i.e. Kendall's tau and the tail copula. In Section 4.4.2 the Kendall's tau based correlation estimator is given, its limiting distribution is determined as well as an asymptotically normal estimator of the asymptotic covariance matrix. Similarly, Section 4.4.3 determines a copula correlation estimator based on the tail copula, its limiting distribution and an estimator of the asymptotic covariance matrix. The concept of the tail copula based correlation estimator is similar to that of Chapter 3, i.e. the index of regular variation α is estimated as in the chapter before and the correlation is estimated by inverting the theoretical tail copula using the empirical tail copula and smoothing over all directions. Finally, Section 4.5 shows the performance of the new method and compares it to the classical approach using a factor model. In a simulation study, the estimated test statistics based on the two copula correlation estimators from Section 4.4 are compared to their limiting χ^2 distribution. It turns out that the distribution of the test statistic is close to their limiting distribution, i.e. the copula concept works well. Finally, in an example with a real financial data set the differences between the copula based estimators and the classical estimator assuming a linear factor model are worked out. The proofs are summarized in Section 4.6.

Chapter 5. The last chapter of this thesis deals with the extreme behavior of a portfolio being a sum of credit default indicators. In the framework of Chapter 4, the underlying latent variables of the default indicators follow a 1-factor model and we are interested in the influence of the dependence structure of the latent variables on the extremal behavior of the limiting portfolio. In Section 5.1 the model is described, where the distribution class of the classical Credit Metrics model is extended by introducing a 'global risk factor'. The main result are given in Section 5.2, i.e. the extreme value distribution of the limiting portfolio is shown for large classes of latent variables. It turns out that the limiting portfolio is always in the maximum domain of attraction of the Weibull distribution, i.e. the tails of the limiting portfolio with support $[0, 1]$ are always polynomially decreasing. Also for some specific distributions of the latent variables the second order behavior (i.e. the slowly varying part) of the limiting portfolio is shown. To see the influence of the second order behavior, a simulation study in Section 5.3 compares 4 different portfolios. These portfolios have light- or heavy tailed latent variables and different global risk factors. We also fix the first order tail behavior, the portfolio mean and the correlation structure in three different settings. It turns out that the tails of the portfolios behave completely different, i.e. the slowly varying part plays an important role. To overcome this problem which model to choose we compare in Section 5.4 the portfolio distribution of the models to the beta distribution. As we can choose the distribution of the global risk factor, we choose it such that it fits the variance of the beta distribution. Then, we observe a similar behavior of the beta distribution and the model where the latent variables follow a multivariate t distribution. Hence, this suggest the t model as a substantial improvement both of the beta model and the standard Credit Metrics model. Section 5.5 explains why one should be careful when using heavy tailed factors, i.e. it may happen that the limiting portfolio degenerates when the factors are not chosen properly. Finally, all proofs of this chapter are given in Section 5.7.

Chapter 1

Dependence estimation and visualization in multivariate extremes with applications to financial data

Summary

We investigate extreme dependence in a multivariate setting with special emphasis on financial applications. We introduce a new dependence function which allows us to capture the complete extreme dependence structure and present a nonparametric estimation procedure. The new dependence function is compared with existing measures including the spectral measure and other devices measuring extreme dependence. We also apply our method to a financial data set of zero coupon swap rates and estimate the extreme dependence in the data.

1.1 Extreme dependence structure

One of the general goals of statistical extreme value theory is to understand the behavior of the *extreme* observations in a set of data generated by a random process and how that information can be used to draw inference about the corresponding aspect of the true distribution. Extreme observations here may be very large or very small observations, or more generally, observations in some *rare* set. Some considerable progress has been made in past decades on the statistical inference of extremes. See Coles (2001), Embrechts, Klüppelberg, and Mikosch (1997) and Smith (2003). In this chapter, we focus on the very large observations in a data set when the observations are multivariate. Specifically, let d be a positive integer and consider an iid sequence of random vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})$, $i \in \mathbb{N}$. We are interested in the statistical inference of the joint distribution of the componentwise maxima

$$M_{n,j} = \bigvee_{i=1}^{n} X_{i,j}\,, \qquad 1 \leq j \leq d\,,$$

for large n. This topic is of relevance in many problems of practical interest; examples can be found in Tawn (1988) (sea levels data), Coles and Tawn (1991) (tidal wave data), Schlather and Tawn (2003) (rainfall data), de Haan and de Ronde (1998) (sea-level and wind-speed data), Dacorogna, Hauksson, Domenig, Müller, and Samorodnitsky (2001) (currency exchange rate data), to name a few.

Among the most important problems in multivariate statistical extremes are the description and inference of dependence between the components of $\boldsymbol{M}_n := (M_{n,1}, \ldots, M_{n,d})$ when n is large. For example, in designing an investment portfolio it is crucial to understand the relative behavior of the various assets in the portfolio in the event of large losses so that the risks can be balanced, or in the event of possible floods, it is important to understand of how extreme rainfall leads to dangerously high river levels so that losses of lives can be prevented.

It is well known that the dependence structure of a random vector can be fully captured by the *copula* or *dependence function*. A *copula* C is a multivariate cumulative distribution function (cdf) with standard uniform marginals. The copula C_G of an arbitrary random vector $(X_1, \ldots, X_d)$ with a joint cdf G and marginal cdf's G_j is given by

$$C_G(u_1, \ldots, u_d) = P\left(X_1 \leq G_1^{\leftarrow}(u_1), \ldots, X_d \leq G_d^{\leftarrow}(u_d)\right), \quad (u_1, \ldots, u_d) \in [0,1]^d\,, \tag{1.1.1}$$

where $G_j^{\leftarrow}(u) := \inf\{x \in \mathbb{R} : G_j(x) \geq u\}$ denotes the left-continuous inverse of G_j. See Joe (1997) for details. We focus on the copula of $\boldsymbol{M}_n$ for large n. Assume that there exist linear normalizing functions $f_{n,1}, \ldots, f_{n,d}$, such that

$$\lim_{n\to\infty} P\left(M_{n,j} \leq f_{n,j}(x_j), 1 \leq j \leq d\right) = F(x_1, \ldots, x_d), \tag{1.1.2}$$

where F is a nondegenerate m-variate cdf. Any possible limit cdf F in (1.1.2) is called a *multivariate extreme value cdf* (mevdf). It can be seen that a cdf F is a mevdf if and only if the marginals F_j, $1 \le j \le m$, are one-dimensional extreme value cdf's (cf. Embrechts et al. 1997) and the copula C satisfies (Joe 1997, Section 6.2)

$$C^t(u_1, \ldots, u_d) = C(u_1^t, \ldots, u_d^t), \quad (u_1, \ldots, u_d) \in [0,1]^d, \ t > 0. \tag{1.1.3}$$

Any copula C satisfying (1.1.3) is called an *extreme copula.*

Since applying monotone transformations to the marginals do not change the copula, (1.1.2) implies that the copula of $\boldsymbol{M}_n$, for large n, can be approximated by that of F and hence approximately satisfies (1.1.3). By the same token, it is clear that the particular normalizations $f_{n,1}, \ldots, f_{n,d}$ in (1.1.2) do not play a role in (1.1.3). Consequently, (1.1.3) is a very general property for the limiting copula of $\boldsymbol{M}_n$.

It is also known that any extreme copula can be written in the form of the *Pickands representation* (Resnick 1997, Section 5.4):

$$C(u_1, \ldots, u_d) = \exp\left\{ \int_{\mathcal{S}_d} \left(\bigwedge_{j=1}^{d} w_j (\ln u_j) \right) \mu(d\boldsymbol{w}) \right\}, \ (u_1, \ldots, u_d) \in [0,1]^d \tag{1.1.4}$$

where μ is a finite measure on $\mathcal{S}_d = \{\boldsymbol{y} \ge 0 : \sum_{i=1}^d y_i = 1\}$ satisfying

$$\int_{\mathcal{S}_d} w_j \mu(d\boldsymbol{w}) = 1, \quad j = 1, \ldots, d.$$

Further, by changing the variable $\boldsymbol{w}$ in the integral in (1.1.4), the extreme copula can be described in infinitely many different but equivalent forms; for instance, Einmahl, de Haan, and Piterbarg (2001) adopts the following representation for the case $d = 2$:

$$C(u_1, u_2) = \exp\left\{ \int_{[0,\pi/2]} \left(\frac{\ln u_1}{1 \vee \cot\theta} \wedge \frac{\ln u_2}{1 \vee \tan\theta} \right) \Phi(d\theta) \right\}, \ u_1, u_2 \in [0,1], \tag{1.1.5}$$

where Φ is a finite measure, called *spectral measure*, on $[0, \pi/2]$ satisfying

$$\int_{[0,\pi/2]} (1 \wedge \tan\theta)\Phi(d\theta) = \int_{[0,\pi/2]} (1 \wedge \cot\theta)\Phi(d\theta) = 1.$$

The focal point of this chapter is the inference of the copula of F in (1.1.2), namely the limiting copula of $\boldsymbol{M}_n$, based on a random sample. In view of (1.1.4), this is equivalent to the inference of the measure μ in the Pickands representation. We will discuss a purely nonparametric approach of estimating the extreme copula. In conjunction, we will introduce a method to visualize extreme tail dependence, a topic which has not received much attention. We believe that simple and effective visualization tools are crucial in this context in order to bridge theory and application. The literature of multivariate extremes has focused almost exclusively on the bivariate case $d = 2$. See Section 2 for a brief review of the literature of this case. The case $d \geq 3$ in contrast has received little attention. Our approach of estimating dependence can be implemented for any general d. Needless to say the curse of dimensionality is even stronger here than in most other contexts so that the general procedure will not achieve the intended purpose unless enough data are available. We will illustrate our procedures by theoretical computations as well as simulations. We will also apply the results on the analysis of a portfolio of zero coupon swap rates.

Throughout the chapter we write $a(u) \sim b(u)$ as $u \to \infty$, if $a(u)/b(u) \to 1$ as $u \to \infty$; we write $a(u) \approx b(u)$ for crude approximations.

1.2 Measuring bivariate extreme sets

As mentioned, the statistical estimation of F in (1.1.2) is of substantial interest in applications. There are three main approaches. Coles and Tawn (1991) and Tawn (1988) assume a parametric form for F and approach the estimation problem by maximum likelihood. While the parametric approach is efficient when the model is correct, the conclusion can be grossly misleading if the model is incorrect. The second approach estimates the measure μ in (1.1.4) based on the empirical measure for the transformed data where the transformation involves parameter estimation on the marginals. Such a procedure is semiparametric in nature and examples of it can be found in Embrechts, de Haan, and Huang (2000), Einmahl, de Haan, and Sinha (1997), de Haan and Resnick (1977) and de Haan and de Ronde (1998). A completely nonparametric approach for estimating μ was introduced in Einmahl, de Haan, and Piterbarg (2001). We next review this approach in detail.

Consider the bivariate case where $d = 2$. Suppose that the $\boldsymbol{X}_i = (X_{i,1}, X_{i,2})$ are iid random vectors with continuous marginal cdf's G_1, G_2. Assume that there exist continuous and nondecreasing normalizing functions $f_{n,1}, f_{n,2}$ such that

$$\lim_{n\to\infty} P\left(\bigvee_{i=1}^{n} X_{i,1} \leq f_{n,1}(x_1),\ \bigvee_{i=1}^{n} X_{i,2} \leq f_{n,2}(x_2)\right) = F(x_1, x_2)\,, \quad x_1, x_2 \in \mathbb{R}\,, \quad (1.2.1)$$

where F has continuous margins. As explained in Section 1.1, the copula C_F of F is an extreme copula and it is independent of the normalizations $(f_{n,1}, f_{n,2})$. Hence we consider instead the normalized limit of $(\bigvee_{i=1}^{n} G_1(X_{i,1}),\ \bigvee_{i=1}^{n} G_2(X_{i,2}))$. Since $G_j(X_{i,j}) \sim \text{unif}[0,1]$, $j = 1, 2$, we have for $j = 1, 2$,

$$\lim_{n\to\infty} P\left(\bigvee_{i=1}^{n} G_j(X_{i,j}) \leq \frac{1}{n}\ln u + 1\right) = u\,, \quad u \in [0,1]\,.$$

Consequently, the following computations yield the copula of F in (1.2.1):

$$P\left(\bigvee_{i=1}^{n} G_1(X_{i,1}) \le \frac{1}{n}\ln u_1 + 1, \bigvee_{i=1}^{n} G_2(X_{i,2}) \le \frac{1}{n}\ln u_2 + 1\right)$$

$$\to \quad C_F(u_1,u_2) \;=\; \exp\left\{\int_0^{\pi/2}\left(\frac{\ln u_1}{1\vee\cot\theta}\wedge\frac{\ln u_2}{1\vee\tan\theta}\right)\Phi(d\theta)\right\},$$

where the representation (1.1.5) is adopted in order to be consistent with the presentation of Einmahl et al. (2001). Let $\overline{G}_j(x) = P(X_{1,j} > x),\ j = 1,2$, then it follows that

$$\begin{aligned} & nP\left(G_1(X_{1,1}) > 1 - \frac{x_1}{n} \text{ or } G_2(X_{1,2}) > 1 - \frac{x_2}{n}\right) \\ = \quad & nP\left(n(\overline{G}_1(X_{1,1}), \overline{G}_2(X_{1,2})) \in ([x_1,\infty]\times[x_2,\infty])^C\right) \\ \overset{n\to\infty}{\longrightarrow} \quad & \int_0^{\pi/2}\left(\frac{x_1}{1\vee\cot\theta}\vee\frac{x_2}{1\vee\tan\theta}\right)\Phi(d\theta). \end{aligned} \tag{1.2.2}$$

Since $P(\cdot)$ is monotone, the discrete index $n\to\infty$ in (1.2.2) can be replaced by a continuous index $t\to\infty$ and the limit remains the same. On $[0,\infty]^2\backslash\{(\infty,\infty)\}$ define the measures Λ_t and Λ on the Borel σ-algebra of $[0,\infty]^2\backslash\{(\infty,\infty)\}$ by

$$\Lambda_t(\boldsymbol{A}) = tP\left(t(\overline{G}_1(X_{1,1}), \overline{G}_2(X_{1,2})) \in \boldsymbol{A}\right),$$

and

$$\Lambda\left(([x_1,\infty]\times[x_2,\infty])^C\right) = \int_0^{\pi/2}\left(\frac{x_1}{1\vee\cot\theta}\vee\frac{x_2}{1\vee\tan\theta}\right)\Phi(d\theta)\,,\ x_1,x_2\in[0,\infty). \tag{1.2.3}$$

Note that the latter relation indeed defines a measure since the sets $([x_1,\infty]\times[x_2,\infty])^C$, $0<x_1,x_2<\infty$, form a π-class which generates the Borel σ-algebra of $[0,\infty]^2\backslash\{(\infty,\infty)\}$. It follows from the continuous-index version of (1.2.2) that for all Borel sets $\boldsymbol{A}\subset[0,\infty]^2\backslash\{(\infty,\infty)\}$ with $\Lambda(\partial\boldsymbol{A}) = 0$, we have (cf. Resnick 1987, Proposition 5.17)

$$\lim_{t\to\infty}\Lambda_t(\boldsymbol{A}) = \lim_{t\to\infty} tP\left(t(\overline{G}_1(X_{1,1}), \overline{G}_2(X_{1,2})) \in \boldsymbol{A}\right) = \Lambda(\boldsymbol{A}). \tag{1.2.4}$$

Given an iid sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$, where $\boldsymbol{X}_i = (X_{i,1}, X_{i,2})$, and a Borel set $\boldsymbol{A}$ in $[0,\infty]^2\backslash\{(\infty,\infty)\}$, an intuitive estimator of $\Lambda_t(\boldsymbol{A})$ is

$$\widetilde{\Lambda}_{t,n}(\boldsymbol{A}) = tP_n\left(t\left(\overline{G}_1(X_{1,1}), \overline{G}_2(X_{1,2})\right) \in \boldsymbol{A}\right) := \frac{t}{n}\sum_{i=1}^{n} I\left(t\left(\overline{G}_1(X_{i,1}), \overline{G}_2(X_{i,2})\right) \in \boldsymbol{A}\right).$$

Furthermore, each $\overline{G}_j(X_{i,j})$ is uniformly distributed on $[0,1]$ and hence can be estimated by $R_{i,j}/n$ where $R_{i,j}$ is the rank of $-X_{i,j}$ among $-X_{1,j},\ldots,-X_{n,j}$. Writing $\varepsilon = t/n$, the estimator $\widetilde{\Lambda}_{t,n}$ is approximated by

$$\widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{A}) = \varepsilon \sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1}, R_{i,2}\right) \in \boldsymbol{A}\right). \tag{1.2.5}$$

This simple and natural estimator works very well both in theory and in practice. The fact that it does not require estimating the marginal tail distributions eliminates an important source of error in the estimation of tail dependence. Generally speaking, the variance and bias of the estimator increases and decreases with ε, respectively, and ε should satisfy $\varepsilon \to 0$ and $n\varepsilon \to \infty$ in order for consistent estimation to be achieved. A result in Einmahl et al. (2001) shows that the estimator can achieve a quick rate of convergence in estimating $\Lambda(\boldsymbol{A})$ for $\boldsymbol{A}$ of a certain form when ε is chosen properly. See Einmahl et al. (2001), Huang (1992) and Qi (1997) for additional details on the theoretical aspects of this estimation approach.

However, in practice when the procedure is implemented we have to select a suitable ε from the data. This is always a difficult issue. In the examples in the next section, we show how to do this by a practical approach.

1.3 Inference of dependence through measure determining classes

We continue our discussions from Section 2 and use the notation developed there. To fully estimate the measure Λ, it suffices to estimate $\Lambda(\boldsymbol{A})$ for sets $\boldsymbol{A}$ in a measure-determining class of Λ. There are obviously infinitely many such classes. The key criteria for selecting such a class are that the measures $\Lambda(\boldsymbol{A})$ are easy to interpret, directly useful for describing tail probabilities, and can be estimated efficiently. Below we mention two examples of such classes for the case $d = 2$.

Definition 1.3.1. *For* $\theta \in [0, \pi/2]$,

$$\begin{aligned} \boldsymbol{C}_\theta &:= \left\{(x_1,x_2) \in [0,\infty]^2 : x_1 \wedge x_2 \le 1,\ x_2 \le x_1 \tan\theta\right\} \quad \textit{and} \\ \boldsymbol{D}_\theta &:= \left\{(x_1,x_2) \in [0,\infty]^2 : x_1 \wedge x_2 \tan\theta \le 1\right\}. \end{aligned}$$

Both sets $\boldsymbol{C}_\theta$ and $\boldsymbol{D}_\theta$ have clear geometric interpretations. For $\theta_1 < \theta_2$ in $[0,\pi/2]$, $\boldsymbol{C}_{\theta_2} \backslash \boldsymbol{C}_{\theta_1}$ contains those points in $[0,\infty]^2$ for which at least one of the components is no bigger than

1 and are trapped in the cone between angles θ_1 and θ_2; $\boldsymbol{D}_\theta$ defines the union of two sets

$$\{(x_1, x_2) : 0 \le x_1 \le 1, 0 \le x_2 \le \infty\} \cup \{(x_1, x_2) : 0 \le x_1 \le \infty, 0 \le x_2 \le \cot\theta\}$$

where the factor $\cot\theta$ allows us to control the boundary of the second set. Define

$$\Lambda(x_1, x_2) = \Lambda\left(([x_1, \infty] \times [x_2, \infty])^C\right), \quad x_1, x_2 \in [0, \infty]^2. \tag{1.3.1}$$

Immediately by (1.2.3),

$$\Lambda(x_1, x_2) = x_1 \Lambda(1, x_2/x_1). \tag{1.3.2}$$

The following holds.

Proposition 1.3.2. *For each $\theta \in [0, \pi/2]$,*

(1) $\Lambda(\boldsymbol{C}_\theta) = \Phi[0, \theta]$, *where Φ is the spectral measure in (1.2.3), and*

(2) $\Lambda(\boldsymbol{D}_\theta) = \Lambda(1, \cot\theta)$. □

In view of Proposition 1.3.2(2) together with (1.3.2), $\{\boldsymbol{D}_\theta : 0 \le \theta \le \pi/2\}$ is measure-determining for Λ. The corresponding result of (1), which is proved in the Appendix, shows that $\{\boldsymbol{C}_\theta : 0 \le \theta \le \pi/2\}$ is also measure-determining for Λ. We note that Proposition 1.3.2(1) was obtained in Einmahl et al. (2001) from an entirely different perspective.

Definition 1.3.3. *For all $\theta \in [0, \pi/2]$ we define*

$$\Phi(\theta) = \Lambda(\boldsymbol{C}_\theta) = \Phi[0, \theta], \quad \textit{and} \quad \psi(\theta) = \Lambda(\boldsymbol{D}_\theta) = \Lambda(1, \cot\theta).$$

□

By Proposition 1.3.2 and (1.2.4),

$$\Phi(\theta) = \lim_{t\to\infty} \Lambda_t(\boldsymbol{C}_\theta) \quad \text{and} \quad \psi(\theta) = \lim_{t\to\infty} \Lambda_t(\boldsymbol{D}_\theta),$$

provided that $\Lambda(\partial \boldsymbol{C}_\theta) = \Lambda(\partial \boldsymbol{D}_\theta) = 0$, and therefore $\Phi(\theta)$ and $\psi(\theta)$ can be estimated statistically by the nonparametric procedures $\widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{C}_\theta), \widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{D}_\theta)$, respectively, if an iid sample is available. From this perspective, we discuss below the relevance of $\Phi(\theta)$ and $\psi(\theta)$.

Estimating $\Phi(\theta)$ is a central theme in Einmahl et al. (2001). Let $\overline{G}_j(x) = P(X_j > x) = 1/x\,, x > 1$. Observe that for $0 < \theta_1 < \theta_2 < \pi/2$,

$$
\begin{aligned}
& P(X_1 \vee X_2 > n, \tan\theta_1 < X_1/X_2 \le \tan\theta_2) \\
= \; & P(n\overline{G}_1(X_1) \wedge n\overline{G}_2(X_2) < 1, \tan\theta_1 < \overline{G}_2(X_2)/\overline{G}_1(X_1) \le \tan\theta_2) \\
= \; & n^{-1}(\Lambda_n(\boldsymbol{C}_{\theta_2}) - \Lambda_n(\boldsymbol{C}_{\theta_1})) \\
\sim \; & n^{-1}(\Phi(\theta_2) - \Phi(\theta_1)),
\end{aligned}
$$

provided $\Lambda(\partial \boldsymbol{C}_{\theta_i}) = 0, i = 1, 2$. However, if the $\overline{G}_i$ are highly non-linear, the quantity $\Phi(\theta_2) - \Phi(\theta_1)$ may be difficult to interpret. It is also somewhat cumbersome to use an estimated $\Phi(\theta)$ to estimate the distribution of the coordinate-wise maxima

$$
P\left(\bigvee_{i=1}^{n} G_j(X_{i,j}) \le \frac{1}{n}u_j + 1,\ j = 1, 2\right);
$$

one could conceivably proceed with this using the integral representation of the copula, but in doing this $\Phi(\theta)$ has to be estimated for every θ followed by a numerical integration. The function $\psi(\theta)$ complements $\Phi(\theta)$ in that respect, as explained below.

Suppose that $x_i = x_{i,n},\ i = 1, 2$, are such that

$$
0 < \liminf_{n\to\infty} n\overline{G}_i(x_i) \le \limsup_{n\to\infty} n\overline{G}_i(x_i) < \infty,\ i = 1, 2.
$$

Then it follows from (1.2.2) that for $n \to \infty$,

$$
\begin{aligned}
P(X_1 > x_1 \text{ or } X_2 > x_2) \;&\sim\; \frac{1}{n}\Lambda(n\overline{G}_1(x_1), n\overline{G}_2(x_2)) \\
= \; \overline{G}_1(x_1)\Lambda\left(1, \frac{\overline{G}_2(x_2)}{\overline{G}_1(x_1)}\right) \;&=\; \overline{G}_1(x_1)\psi\left(\arctan\left(\frac{\overline{G}_1(x_1)}{\overline{G}_2(x_2)}\right)\right). \qquad (1.3.3)
\end{aligned}
$$

As a result,

$$
\begin{aligned}
P^n(X_1 \le x_1, X_2 \le x_2) \;&\approx\; \exp\left(-n\overline{G}_1(x_1)\psi\left(\arctan\left(\frac{\overline{G}_1(x_1)}{\overline{G}_2(x_2)}\right)\right)\right) \\
&\approx\; P^{n\xi}(X_1 \le x_1), \qquad (1.3.4)
\end{aligned}
$$

where

$$
\xi = \xi(x_1, x_2) = \psi\left(\arctan\left(\frac{\overline{G}_1(x_1)}{\overline{G}_2(x_2)}\right)\right).
$$

If $G_1 = G_2$ then

$$
\xi(x, x) = \psi(\pi/4), \qquad (1.3.5)
$$

which is what Schlather and Tawn (2000) refers to as *extremal coefficient*, a notion related to the extremal index (cf. Leadbetter, Lindgren, and Rootzén (1983) or Embrechts et al. (1997)) in univariate extreme value theory for time series.

1.4 Bivariate tail dependence function

In this section we continue to explore the properties of $\psi(\theta)$ defined in Definition 1.3.3 and how it can be useful for describing multivariate extremes. First, we have:

Proposition 1.4.1. **(1)** *ψ is convex.*

(2) $\psi_1(\theta) \le \psi(\theta) \le \psi_0(\theta), \quad \theta \in [0, \pi/2]$, *where* $\psi_0(\theta) := 1 + \cot\theta$ *corresponds to independence and* $\psi_1(\theta) := 1 \vee \cot\theta$ *to complete dependence.*

Proof: $\psi(\theta) = \int_0^{\pi/2} (1/(1 \vee \cot\gamma)) \vee (\cot\theta/(1 \vee \tan\gamma))\, \Phi(d\gamma)$ and since the integrand is convex with respect to θ, ψ is, hence (1) holds. An equivalent expression of ψ is given by *Pickands Dependence function* D (see Pickands (1981)), i.e.

$$\psi(\theta) \quad = \quad (1 + \cot\theta) D\left(\frac{1}{1 + \cot\theta}\right),$$

where $D : [0, 1] \to [1/2, 1]$ is a convex function satisfying $D(0) = D(1) = 1$ as well as $z \vee (1 - z) \le D(z) \le 1$, $z \in [0, 1]$, hence (2) holds. □

The function ψ becomes a much more effective tool for visualizing dependence if it is normalized, as follows.

Definition 1.4.2. *We define the bivariate* tail dependence function *as*

$$\rho(\theta) = \frac{\psi_0(\theta) - \psi(\theta)}{\psi_0(\theta) - \psi_1(\theta)} \quad = \quad \frac{1 + \cot\theta - \psi(\theta)}{1 \wedge \cot\theta}, \quad \theta \in (0, \pi/2). \tag{1.4.1}$$

□

By Proposition 1.4.1(2) the function $\rho(\theta)$ takes values in $[0, 1]$, with $\rho(\theta)$ being close to 0/1 corresponds to weak/strong dependence.

A similar approach to ρ is to be found in the *canonical dependence function* or *tail dependence function* ξ, defined in Falk, Hüsler, and Reiss (2004, Section 6.4), i.e.

$$\begin{aligned} \rho(\theta) \quad &= \quad \frac{1 + \cot\theta - \psi(\theta)}{1 + \cot\theta - 1 \vee \cot\theta} \quad = \quad \frac{1 - D\left((1 + \cot\theta)^{-1}\right)}{1 - \left[(1 + \cot\theta)^{-1} \vee (1 - (1 + \cot\theta)^{-1})\right]} \\ &= \quad \xi\left(\frac{1}{1 + \cot\theta}\right). \end{aligned}$$

The quantity $\rho(\pi/4) = 2 - \psi(\pi/4)$ (cf. (1.3.5)) is referred to as the *(upper) tail dependence coefficient* in Joe (1997), which, as the name suggests, is meant to describe the degree of dependence in the upper tails of the marginals. Thus, the function ρ extends this notion from a single direction, $\pi/4$, to all directions in $(0, \pi/2)$. This is illustrated by the following example, which is similar to an example in Ledford and Tawn (1996).

Example 1.4.3. Let $X_1 \sim G_1$, $X_2 \sim G_2$ where G_1 and G_2 are continuous distributions. Note that $\left(1/\overline{G}_1(X_1), 1/\overline{G}_2(X_2)\right)$ has Pareto(1) margins and the same copula as (X_1, X_2). It follows from (1.2.4) and Definition 1.3.3 that for all $\theta \in (0, \pi/2)$, we have $\Lambda_t(1, \cot\theta) \to \Lambda(1, \cot\theta) = \psi(\theta)$ as $t \to \infty$, and hence

$$
\begin{aligned}
&\lim_{t\to\infty} P\left(\frac{1}{\overline{G}_2(X_2)} > t\tan\theta \,\middle|\, \frac{1}{\overline{G}_1(X_1)} > t\right) \\
&= \lim_{t\to\infty} t\left(1 - P\left(\frac{1}{\overline{G}_1(X_1)} \le t\right) - P\left(\frac{1}{\overline{G}_2(X_2)} \le t\tan\theta\right)\right. \\
&\quad \left. + P\left(\frac{1}{\overline{G}_1(X_1)} \le t\,,\ \frac{1}{\overline{G}_2(X_2)} \le t\tan\theta\right)\right) \\
&= 1 + \cot\theta - \lim_{t\to\infty} t\left(1 - P\left(\frac{1}{\overline{G}_1(X_1)} \le t\,,\ \frac{1}{\overline{G}_2(X_2)} \le t\tan\theta\right)\right) \\
&= 1 + \cot\theta - \lim_{t\to\infty} tP\left(t\left(\overline{G}_1(X_1), \overline{G}_2(X_2)\right) \in \left([1,\infty] \times [\cot\theta, \infty]\right)^C\right) \\
&= 1 + \cot\theta - \psi(\theta) = (1 \wedge \cot\theta)\,\rho(\theta)\,.
\end{aligned}
$$

Hence for all $\theta \in (0, \pi/2)$,

$$
\lim_{t\to\infty} P\left(X_2 > G_2^{-1}\left(1 - \frac{1}{t\tan\theta}\right) \,\middle|\, X_1 > G_1^{-1}\left(1 - \frac{1}{t}\right)\right) = (1 \wedge \cot\theta)\,\rho(\theta)\,.
$$

□

Our examples below show that ρ provides an effective tool to visualize dependence in the extreme tails of the bivariate distribution. In practice, when G is unknown, $\rho(\theta)$ can be estimated from a set of iid data $(X_i, Y_i), 1 \le i \le n$, by the nonparametric estimate

$$
\widehat{\rho}_{\varepsilon,n}(\theta) = \frac{\psi_0(\theta) - \widehat{\psi}_{\varepsilon,n}(\theta)}{\psi_0(\theta) - \psi_1(\theta)} = \frac{1 + \cot\theta - \widehat{\psi}_{\varepsilon,n}(\theta)}{1 \wedge \cot\theta}\,, \quad \theta \in (0, \pi/2),
$$

where

$$
\begin{aligned}
\widehat{\psi}_{\varepsilon,n}(\theta) &:= \widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{D}_\theta) \\
&= \varepsilon \sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1}, R_{i,2}\right) \in \boldsymbol{D}_\theta\right) \\
&= \varepsilon \sum_{i=1}^{n} I(R_{i,1} \le \varepsilon^{-1} \text{ or } R_{i,2} \le \varepsilon^{-1}\cot\theta).
\end{aligned}
$$

As mentioned in Section 2, theoretically ε and $1/(n\varepsilon)$ should be both small in order for the estimator to perform well. In practice, we will plot $\widehat{\rho}_{\varepsilon,n}(\theta)$ for ε in some sensible range for which ε and $1/(n\varepsilon)$ are "small" and pick an ε_0 for which the estimates $\widehat{\psi}_{\varepsilon,n}(\pi/4)$

behave stably in the neighborhood of ε_0. While it is convenient to use the same ε for all θ, allowing ε to vary with θ in simple ways may improve the quality of the estimation. Indeed, when θ approaches $\pi/2$, increasingly fewer points of $\varepsilon\,(R_{i,1}, R_{i,2})$ are captured by $\boldsymbol{D}_\theta$, which has the effect of inflating the variance of the estimate in that region. A practical way to overcome this is to choose a baseline $\varepsilon = \varepsilon_0$ at $\theta = \pi/4$ and allow ε to decrease slightly as θ approaches $\pi/2$. Another practical consideration is a simple smoothing. At least visually if not theoretically, the quality of the estimate of $\widehat{\rho}_{\varepsilon,n}(\theta)$ improves if some smoothing is incorporated. In that regard, one can perform a simple averaging over a box window or use something more sophisticated such as spline smoothing.

We also recommend plotting $(1/R_{i,1}, 1/R_{i,2})$, $1 \le i \le n$, alongside that of $\widehat{\rho}_{\varepsilon,n}(\theta)$ to fully appreciate the information in the latter. Recall that

$$\lim_{n\to\infty} P^n\left(1/(n\overline{G}_1(X_1)) \le 1, 1/(n\overline{G}_2(X_2)) \le \tan\theta\right) = \rho(\theta).$$

As such, $\widehat{\rho}_{\varepsilon,n}(\theta)$ describes the degree of dependence reflected by the pattern of points of $(1/R_{i,1}, 1/R_{i,2})$, $1 \le i \le n$, in the box $[0,1] \times [0, \tan\theta]$. The following simple example demonstrates these points.

Example 1.4.4. Let $p_1, p_2 \in (0,1)$ and consider the model

$$X_1 = p_1 Z_1 \vee (1-p_1)Z_2 \quad \text{and} \quad X_2 = p_2 Z_1 \vee (1-p_2)Z_3,$$

with Z_1, Z_2, Z_3 distributed as iid Pareto(1). Clearly, the dependence between X_1 and X_2 arises from the common component Z_1. Hence the dependence is stronger for larger values of p_1, p_2. It is easy to see that both X_1 and X_2 are asymptotically distributed as Pareto(1) in the tails. It is also easy to see that

$$P(X_1 > x \text{ or } X_2 > x\tan\theta) \sim \frac{1}{x}\left(1 + \cot\theta - p_1 \wedge p_2 \cot\theta\right).$$

Applying (1.3.3), we have

$$\psi(x) = 1 + \cot\theta - p_1 \wedge p_2 \cot\theta,$$

and

$$\rho(\theta) \quad = \quad \frac{p_1 \wedge p_2 \cot\theta}{1 \wedge \cot\theta}. \tag{1.4.2}$$

In Figure 1.1 we simulated this model for $n = 10\,000$ iid observations of (X_1, X_2). The three sets of plots on the three rows correspond to the cases: $p_1 = 0.7$, $p_2 = 0.3$, $p_1 =$

0.5, $p_2 = 0.5$ and $p_1 = 0.2$, $p_2 = 0.8$. On each row the left-most plot is the true functions $\rho(\theta)$ in (1.4.2) (dashed line) overlaid with the smoothed version of $\widehat{\rho}_{\varepsilon,n}(\theta)$ (solid line) based on one simulated sample of size 10 000, where ε is 1/200 for $\theta \in [0, \pi/4]$ and thereafter, ε decreases linearly to 1/210 when θ reaches $\pi/2$. We computed $\widehat{\rho}_{\varepsilon,n}(\theta)$ for $\theta \in \{\theta_i = i\pi/200, 1 \le i \le 100\}$ and produced the smoothed version $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_i)$ by averaging $\widehat{\rho}_{\varepsilon,n}(\theta_j), |j-i| \le s = 5$, i.e.

$$\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_i) = \frac{1}{2s+1} \sum_{j=-s}^{s} \widehat{\rho}_{\varepsilon,n}(\theta_{i-j}) \, .$$

The plots in the middle column illustrate the root of the mean squared error

$$\mathrm{MSE}(\theta_i) = \sum_{k=1}^{100} \left(\widehat{\rho}^{(s),k}_{\varepsilon,n}(\theta_i) - \rho(\theta_i) \right)^2$$

for the three cases based on 100 simulations with $n = 10\,000$ iid observations each and $\widehat{\rho}^{(s),k}_{\varepsilon,n}(\theta_i)$ represents the smoothed estimator of simulation k, $1 \le k \le 100$. The right-most plots contain the simulated points $(1/R_{i,1}, 1/R_{i,2})$, $1 \le i \le n$, of one single sample of size 10 000 but with points close to (1,1) truncated for easy viewing.

In the first row of plots, ρ is larger for small θ than for large θ; this is reflected by the right-most plot in which the violation of independence can be seen to be more severe below the diagonal. In the second row of plots, ρ is constant; which is reflected by having a portion of extreme points lined up on the diagonal in the right-most plot. The third row of plots is the converse of the first row of plots which is reflected by the pattern of extreme points above the diagonal. This is an example of a situation where Joe's tail dependence coefficient does not convey a good picture of extreme dependence, in that $\rho(\pi/4)$ is not sufficient to describe the full dependence structure of this model. □

Example 1.4.5. Let $\boldsymbol{X} = (X_1, X_2)$ be a bivariate random vector with dependence structure given by a *Gumbel*-copula

$$C_{\boldsymbol{X}}(u,v) = \exp\left\{ - \left[(-\ln u)^{\delta} + (-\ln v)^{\delta} \right]^{1/\delta} \right\} , \quad \delta \in [1, \infty) \, . \tag{1.4.3}$$

The dependence arises from δ. It is a symmetric model and by Example 1.4.3 it has (upper) tail dependence coefficient $\rho(\pi/4) = \lambda_U = \ln 2 / \ln(2-\delta)$. Since $C_{\boldsymbol{X}}$ is an extreme copula, $\psi(\theta) = \left(1 + (\cot\theta)^{\delta}\right)^{1/\delta}$ and hence

$$\rho(\theta) = \frac{1 + \cot\theta - \left(1 + (\cot\theta)^{\delta}\right)^{1/\delta}}{1 \wedge \cot\theta} , \quad \theta \in (0, \pi/2) \, .$$

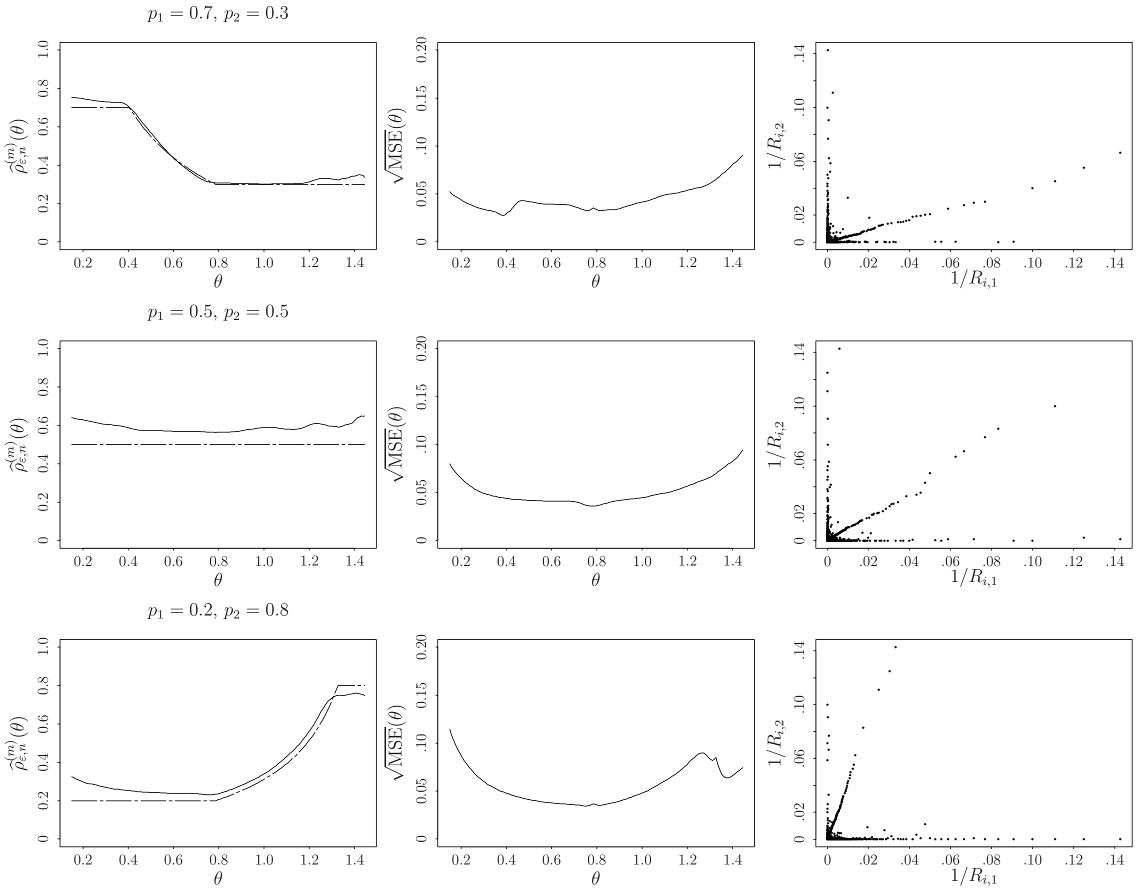

Figure 1.1: Left column: smoothed version of $\widehat{\rho}^{(5)}_{\varepsilon,n}(\theta)$ (solid line) overlaid with true function $\rho(\theta)$.
Middle column: $\sqrt{\mathrm{MSE}}(\theta)$.
Right column: plots of $(1/R_{i,1}, 1/R_{i,2})$, with points close to (1,1) truncated, $p_1 = 0.7$, $p_2 = 0.3$ (upper row), $p_1 = 0.5$, $p_2 = 0.5$ (middle row) and $p_1 = 0.2$, $p_2 = 0.8$ (lower row).

We simulated this model for $n = 10\,000$, and in Figure 1.2 the plots are given in the same order as in Figure 1.1 based on Example 1.4.4. We have chosen $\rho(\pi/4) = 0.3$ (upper row), $\rho(\pi/4) = 0.7$ (middle row) and $\rho(\pi/4) = 0.9$ (lower row). The level of dependence is manifested by the data scattered around the diagonal. $\square$

1.5 Multivariate extensions

One advantage of the functions of ψ and ρ in Definitions 3.3 and 4.2 is that they can be readily extended to higher dimensions by incorporating additional angles θ_j. Let $d \geq 2$ and $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})$ be iid with a distribution G, where the margins G_j are assumed to be continuous. Assume that (1.1.2) holds and the copula of F has the representation

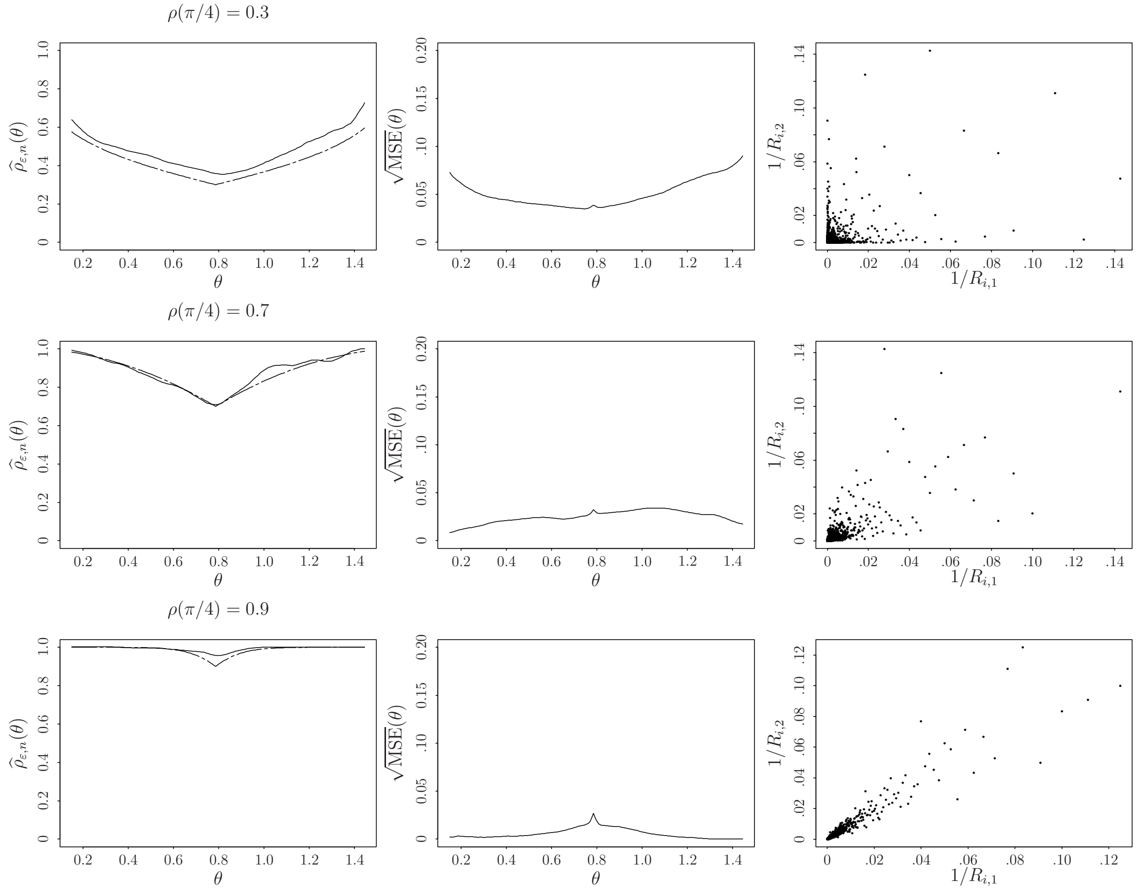

Figure 1.2: Left column: smoothed versions $\widehat{\rho}^{(5)}_{\varepsilon,n}(\theta)$ (solid line) overlaid with true function $\rho(\theta)$.
Middle column: $\sqrt{\text{MSE}}(\theta)$.
Right column: plots of $1/R_{i,j}$, $j = 1, 2$, for $C_{\boldsymbol{X}}$ given in (1.4.3) and $\rho(\pi/4) = 0.3$ (upper row), $\rho(\pi/4) = 0.7$ (middle row), $\rho(\pi/4) = 0.9$ (lower row).

(1.1.4). Define the measures Λ_t and Λ on the Borel σ-algebra of $[0,\infty]^d \backslash \{\infty, \ldots, \infty\}$ by

$$\Lambda_t(\boldsymbol{A}) := tP\left(t\left(\overline{G}_1(X_{1,1}), \ldots, \overline{G}_d(X_{1,d})\right) \in \boldsymbol{A}\right),$$

and

$$\Lambda\left(([x_1,\infty] \times \cdots \times [x_d,\infty])^C\right) := \int_{S_d} \left(\bigvee_{j=1}^{d} w_j x_j\right) \mu(d\boldsymbol{w}) .$$

As in the two-dimensional case, we have

$$\lim_{t\to\infty} \Lambda_t(\boldsymbol{A}) = \Lambda(\boldsymbol{A})$$

for any Borel set $\boldsymbol{A} \subset [0,\infty]^d \backslash \{\infty, \ldots, \infty\}$ with $\Lambda(\partial \boldsymbol{A}) = 0$. Now set

$$\Lambda(x_1, \ldots, x_d) := \Lambda\left(([x_1,\infty] \times \cdots \times [x_d,\infty])^C\right),$$

and, for $\theta_2, \ldots, \theta_d \in [0, \pi/2]$,

$$\boldsymbol{D}_{\theta_2,\ldots,\theta_d} := \left\{(x_1, \ldots, x_d) \in [0,\infty]^d : \, x_1 \wedge x_2 \tan\theta_2 \wedge \cdots \wedge x_d \tan\theta_d \leq 1\right\} .$$

The sets $\boldsymbol{D}_{\theta_2,\ldots,\theta_d}$ for $\theta_2, \ldots, \theta_d \in [0, \pi/2]$ are measure-determining for Λ. Define

$$\psi(\theta_2, \ldots, \theta_d) := \Lambda\left(\boldsymbol{D}_{\theta_2,\ldots,\theta_d}\right) = \Lambda(1, \cot\theta_2, \ldots, \cot\theta_d).$$

Hence by the same arguments as in Proposition 1.4.1, ψ is convex and

$$\psi_1(\theta_2, \ldots, \theta_d) \leq \psi(\theta_2, \ldots, \theta_d) \leq \psi_0(\theta_2, \ldots, \theta_d), \quad \theta_2, \ldots, \theta_d \in [0, \pi/2], \tag{1.5.1}$$

where

$$\begin{aligned} \psi_0(\theta_2, \ldots, \theta_d) &= 1 + \cot\theta_2 + \cdots + \cot\theta_d, \\ \psi_1(\theta_2, \ldots, \theta_d) &= 1 \vee \cot\theta_2 \vee \cdots \vee \cot\theta_d; \end{aligned}$$

ψ_0 and ψ_1 correspond to the independent and completely dependent cases, respectively.

Definition 1.5.1. *The* tail dependence function*, for $d \geq 2$, is defined as*

$$\rho(\theta_2, \ldots, \theta_d) = \frac{(1 + \cot\theta_2 + \cdots + \cot\theta_d) - \psi(\theta_2, \ldots, \theta_d)}{(1 + \cot\theta_2 + \cdots + \cot\theta_d) - (1 \vee \cot\theta_2 \vee \cdots \vee \cot\theta_d)}.$$

□

By (1.5.1), ρ is in $[0, 1]$ and ρ being close to 0 and 1 correspond to weak and strong dependence, respectively. Again, a similar approach to ρ in this multivariate case with $d \geq 2$ is given by the *canonical dependence function* ξ defined in Falk et al. (2004).

In practice, when G is unknown, $\Lambda(\boldsymbol{A})$ can be estimated for any Borel set $\boldsymbol{A}$ from a set of data $(X_{i,1}, \ldots, X_{i,d})$, $1 \leq i \leq n$, using the nonparametric estimator

$$\widehat{\Lambda}_{\varepsilon,n}(\boldsymbol{A}) = \varepsilon \sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1}, \ldots, R_{i,d}\right) \in \boldsymbol{A}\right). \tag{1.5.2}$$

The theoretical properties of the bivariate estimator as explained after (1.2.5) can also be verified in higher dimensions. Accordingly, the estimate $\rho(\theta_2, \ldots, \theta_d)$ is defined as

$$\widehat{\rho}_{\varepsilon,n}(\theta_2, \ldots, \theta_d) := \frac{\psi_0(\theta_2, \ldots, \theta_d) - \widehat{\Lambda}_{\varepsilon,n}\left(\boldsymbol{D}_{\theta_2,\ldots,\theta_d}\right)}{\psi_0(\theta_2, \ldots, \theta_d) - \psi_1(\theta_2, \ldots, \theta_d)},$$

where

$$\widehat{\Lambda}_{\varepsilon,n}\left(\boldsymbol{D}_{\theta_2,\ldots,\theta_d}\right) = \varepsilon \sum_{i=1}^{n} I\left(\varepsilon\left(R_{i,1}, \ldots, R_{i,d}\right) \in \boldsymbol{D}_{\theta_2,\ldots,\theta_d}\right).$$

All practical considerations made in the previous section continue to be applicable here. To visualize extreme dependence in the data, plot $\widehat{\rho}_{\varepsilon,n}(\theta_2, \ldots, \theta_d)$ for a discrete set of $(\theta_2, \ldots, \theta_d)$. When $d \geq 3$, plotting the estimated ρ requires considerable creativity. In the following example the tail dependence function can be calculated explicitly.

Example 1.5.2. Let $c_{ji} \in [0, 1]$ for $1 \leq j \leq d, 1 \leq i \leq k$, such that $\sum_{i=1}^{k} c_{ji} = 1$ for all j. Consider

$$X_j = \bigvee_{i=1}^{k} c_{ji} Z_i \,, \quad j = 1, \ldots, d,$$

where $Z_1, \ldots, Z_k$ are iid Pareto(1). Generalizing (1.3.3), we obtain

$$P\left(X_1 > x \text{ or } X_2 > x \tan\theta_2 \text{ or } \cdots \text{ or } X_d > x \tan\theta_d\right) \sim \frac{1}{x}\psi(\theta_2, \ldots, \theta_d)\,, \quad x \to \infty\,.$$

On the other hand,

$$\begin{aligned}
& P\left(X_1 > x \text{ or } X_2 > x \tan\theta_2 \text{ or } \cdots \text{ or } X_d > x \tan\theta_d\right) \\
&= 1 - P\left(X_1 \leq x, X_2 \leq x \tan\theta_2, \ldots, X_d \leq x \tan\theta_d\right) \\
&= 1 - \prod_{j=1}^{k} P\left(Z_j \leq x \left(c_{1j}^{-1} \wedge c_{2j}^{-1} \tan\theta_2 \wedge \cdots \wedge c_{dj}^{-1} \tan\theta_d\right)\right) \\
&\sim \frac{1}{x} \sum_{j=1}^{k} c_{1j} \vee c_{2j} \cot\theta_2 \vee \cdots \vee c_{dj} \cot\theta_d\,.
\end{aligned}$$

Hence,

$$\psi(\theta_2, \ldots, \theta_d) = \sum_{i=1}^{k} \left(c_{1i} \vee c_{2i} \cot\theta_2 \vee \cdots \vee c_{di} \cot\theta_d\right)\,,$$

and

$$\begin{aligned}
& \rho(\theta_2, \ldots, \theta_d) \\
&= \frac{(1 + \cot\theta_2 + \cdots + \cot\theta_d) - \sum_{i=1}^{k} \left(c_{1i} \vee c_{2i} \cot\theta_2 \vee \cdots \vee c_{di} \cot\theta_d\right)}{(1 + \cot\theta_2 + \cdots + \cot\theta_d) - (1 \vee \cot\theta_2 \vee \cdots \vee \cot\theta_d)}\,.
\end{aligned}$$

Note that this example generalizes Example 1.4.4 which is the special case of $d = 2$, $k = 3$, $c_{11} = p_1$, $c_{12} = 1 - p_1$, $c_{13} = 0$, $c_{21} = p_2$, $c_{22} = 0$, $c_{23} = 1 - p_2$. □

Example 1.5.3. We estimate the dependence structure of the model given in Example 1.5.2 with $d = 3$ and $k = 5$. We choose the constants c_{ji}, $1 \leq j \leq 3$, $1 \leq i \leq 5$, as

$c_{11} = 0.2$	$c_{12} = 0.2$	$c_{13} = 0$	$c_{14} = 0.6$	$c_{15} = 0$
$c_{21} = 0.6$	$c_{22} = 0$	$c_{23} = 0.2$	$c_{24} = 0$	$c_{25} = 0.2$
$c_{31} = 0.2$	$c_{32} = 0.6$	$c_{33} = 0.2$	$c_{34} = 0$	$c_{35} = 0$

Figures 1.3 and 1.4 contain the simulation results of this model for $n = 10\,000$ iid observations of (X_1, X_2, X_3). We chose $\varepsilon = 1/200$ and computed the estimate $\widehat{\rho}_{\varepsilon,n}(\theta_2, \theta_3)$ for $\theta_2, \theta_3 \in \{\theta_i = i\pi/200,\, 1 \le i \le 100\}$ and smoothed $\widehat{\rho}_{\varepsilon,n}(\theta_i, \theta_j)$ by averaging $\widehat{\rho}_{\varepsilon,n}(\theta_k, \theta_l)$, $|k - i|, |l - j| \le s = 3$, i.e.

$$\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_i, \theta_j) = \frac{1}{(2s+1)^2} \sum_{k,l=-s}^{s} \widehat{\rho}_{\varepsilon,n}(\theta_{i-k}, \theta_{j-l})\,.$$

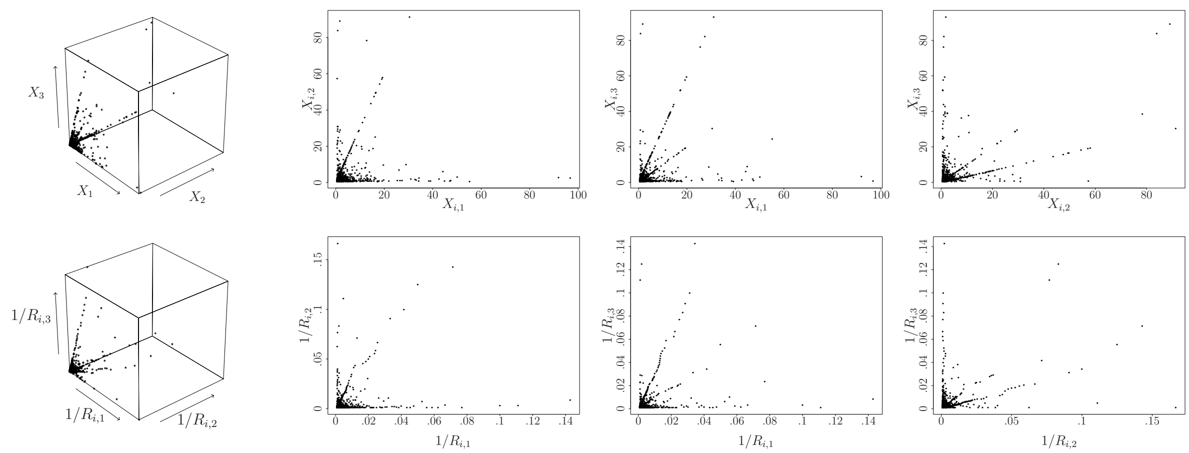

Figure 1.3: First row: data (3-d and 2-d projections) of the model given in Example 1.5.2. Second row: ranks $1/R_{i,j}$, $1 \le j \le 3$, in the same order as in row one.

In the first row of Figure 1.3 the data are plotted, where in the left-most plot we show the 3-dimensional data, the three plots on the right hand side show the projections of the data, $(X_{i,1}, X_{i,2})$, $(X_{i,1}, X_{i,3})$ and $(X_{i,2}, X_{i,3})$. The second row is given in the same order as the first row, showing the reciprocal ranks $1/R_{i,j}$, $1 \le j \le 3$. The first row of Figure 1.4 shows the estimate $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2, \theta_3)$, where the left plot is a perspective plot, the middle one is a contour plot and the right one is a grey-scale image plot. To see how the estimator performs the second row presents the true tail-dependence function $\rho(\theta_2, \theta_3)$ for this model.

Remark 1.5.4. Let $\rho_{1,2,3}$ be the tail dependence function of three rvs X_1, X_2, X_3 and $\rho_{1,j}$ be the tail dependence function of X_1, X_j, $j = 2, 3$, hence by definition $\rho_{1,2}(\theta_2) = \rho_{1,2,3}(\theta_2, \pi/2)$ and $\rho_{1,3}(\theta_3) = \rho_{1,2,3}(\pi/2, \theta_3)$ holds $\forall \theta_2, \theta_3 \in (0, \pi/2)$. Therefore $\rho_{1,2}$ can be estimated by the cross section of the estimated trivariate tail dependence function at a large and fixed angle θ_2, and similarly for $\rho_{1,3}$. To identify $\rho_{2,3}$ recall that $\Lambda_{1,2,3}(0, a, b) =$

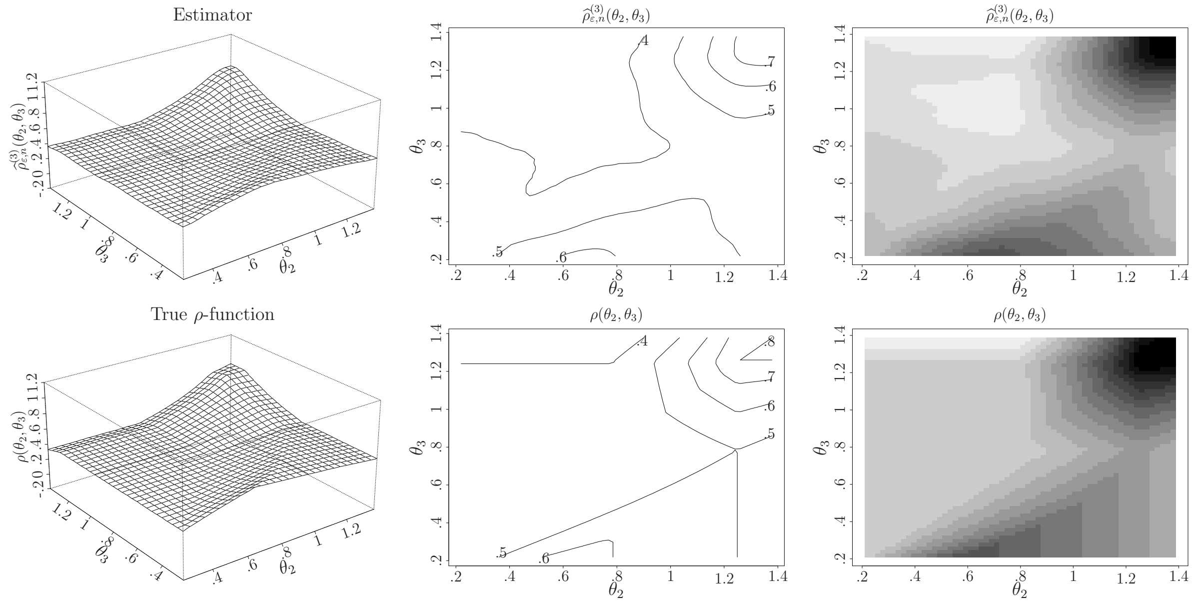

Figure 1.4: First row: smoothed estimate $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2, \theta_3)$ of the simulated data (see Figure 1.3), with perspective plot (left-most), contour plot (middle) and image plot (right-most).
Second row: true tail dependence function $\rho(\theta_2, \theta_3)$ for this model.

$\Lambda_{2,3}(a, b)$, hence

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \rho_{1,2,3}\left(\arctan\varepsilon, \arctan\left(\varepsilon\tan\theta\right)\right)\\
&= \lim_{\varepsilon\to 0} \frac{1 + 1/\varepsilon + \cot\theta/\varepsilon - \psi_{1,2,3}\left(\arctan\varepsilon, \arctan\left(\varepsilon\tan\theta\right)\right)}{1 + 1/\varepsilon + \cot\theta/\varepsilon - 1 \vee 1/\varepsilon \vee \cot\theta/\varepsilon}\\
&= \lim_{\varepsilon\to 0} \frac{\varepsilon + 1 + \cot\theta - \varepsilon\Lambda_{1,2,3}\left(1, 1/\varepsilon, \cot\theta/\varepsilon\right)}{\varepsilon + 1 + \cot\theta - \varepsilon \vee 1 \vee \cot\theta}\\
&= \lim_{\varepsilon\to 0} \frac{\varepsilon + 1 + \cot\theta - \Lambda_{1,2,3}\left(\varepsilon, 1, \cot\theta\right)}{\varepsilon + 1 + \cot\theta - \varepsilon \vee 1 \vee \cot\theta}\\
&= \frac{1 + \cot\theta - \Lambda_{2,3}\left(1, \cot\theta\right)}{1 + \cot\theta - 1 \vee \cot\theta} = \rho_{2,3}(\theta)\,.
\end{aligned}$$

□

1.6 The swap rate data

The data consist of returns (daily differences) of *Annually Compounded Zero Coupon Swap Rates* with different maturities (between 7 days and 30 years) and different currencies (EUR, USD and GBP). Each of the time series consists of 257 daily returns during the year 2001. In an exploratory data analysis we investigated first each single time series.

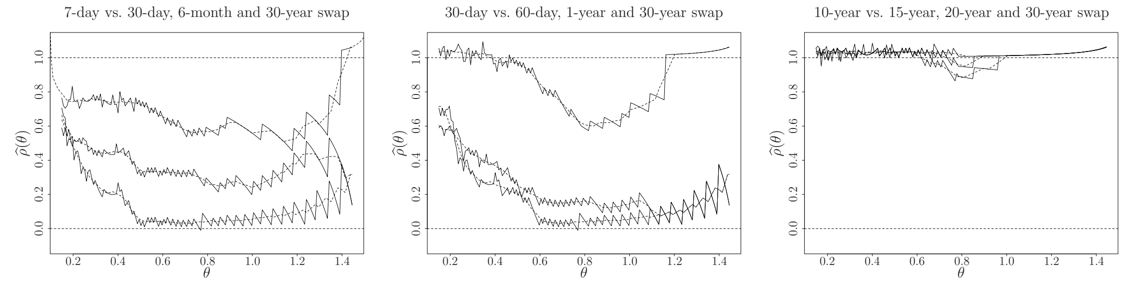

Figure 1.5: Estimates $\widehat{\rho}$ for some swap rates with smoothed versions (dashed lines).
Left plot: $\widehat{\rho}$ for 7-day vs 30-day, 7-day vs 6-month, and 7-days vs 30-year.
Middle plot: $\widehat{\rho}$ for 30-day vs 60-day, 30-day vs 1-year and 30-day vs 30-year.
Right plot: $\widehat{\rho}$ for 10-year vs 15-year, 10-year vs 20-year and 10-year vs 30-year.

Plots of the autocorrelation functions of the single time series, their moduli and squares exhibited no significant temporal dependence structure; hence we assume the data being iid. Moreover, the histograms and a tail analysis showed that the marginals are well modelled (at least in the tails) by a two-sided exponential distribution. Concerning multivariate (spatial) dependence, for swap rates in the same currency we observed a high dependence for similar maturities, and a low dependence between very different maturities. Between different currencies we observed only very little dependence except for similarly long maturities, where we detected some moderate dependence. For plots and details on these effects we refer to Kuhn (2002).

To see the estimator $\widehat{\rho}$ at work we show plots of $\widehat{\rho}_{\varepsilon,n}(\theta)$, $\theta \in (0, \pi/2)$, as defined in (1.4.1) for the *swap rate data* described above for EUR. We use the nonparametric estimator given in (1.2.5). We stay away from the boundaries $\theta = 0$ and $\theta = \pi/2$ since $\widehat{\psi}_{\varepsilon,n}(\theta)$ tends to ∞ as $\theta \to 0$, and for θ near $\pi/2$ there is a lack of data.

In Figure 1.5 the tail dependence function is estimated for various combinations of swap rates of different maturities with $\widehat{\rho}_{\varepsilon,n}(\theta_i)$ (zigzag-line) and the smoothed version $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_i)$ (dashed line) for $\varepsilon = 0.06$, $s = 5$ and $\theta_i = \frac{i}{200}\frac{\pi}{2}$, $1 \le i \le 200$. The left plot shows strong dependence between the 7-day and 30-day rates, moderate dependence between the 7-day and 6-month rates, but very weak dependence between the 7-day and 30-year rates. The middle plot shows moderate dependence between the 30-day and 60-day rates for θ close to $\pi/4$ and exceptionally high dependence for θ small or large, but weak dependence between the 30-day and 1-year and 30-day and 30-year rates. The right plot shows strong dependence between the 10-year,15-year, 20-year and 30-year rates. $\square$

Example 1.6.1. Figure 1.6 shows a comparison of the tail dependence function with the spectral measure Φ as defined in (1.3.2). We recall that in case of independence

$$\int_0^{\frac{\pi}{2}} \left(\frac{x}{1 \vee \cot \gamma} \vee \frac{y}{1 \vee \tan \gamma} \right) \Phi(d\gamma) = x + y,$$

and hence

$$\Phi(\theta) = \Phi([0, \theta]) = \begin{cases} 1, & \theta < \pi/2\,, \\ 2, & \theta = \pi/2\,, \end{cases}$$

and in case of complete dependence

$$\int_0^{\frac{\pi}{2}} \left(\frac{x}{1 \vee \cot \gamma} \vee \frac{y}{1 \vee \tan \gamma} \right) \Phi(d\gamma) = x \vee y,$$

and hence

$$\Phi(\theta) = \Phi([0, \theta]) = \begin{cases} 0, & \theta < \pi/4\,, \\ 1, & \pi/4 \leq \theta \leq \pi/2\,. \end{cases}$$

These results allow us to interpret the plots. We consider the 20-year vs. 30-year, 7-day vs. 30-day, and 7-day vs. 30-year swap rates. In the first row (high dependence) the estimated spectral measure Φ equals 0 for $\theta < 0.4$ and then quickly jumps to 1. In the third row (low dependence) the estimated Φ jumps quickly to 1 and remains there until close to $\pi/2$ where it jumps to 2. The middle row (moderate dependence) is a mixture of high and low dependence case. □

Example 1.6.2. Figures 1.7 and 1.8 show two trivariate examples. The first example is generated by the low dependent swap rates with 7 day maturity and currencies USD, EUR and GBP; $X_{i,1}$ corresponds to USD, $X_{i,2}$ to EUR and $X_{i,3}$ to GBP. In the first row we plotted the ranks $1/R_{i,j}$, $1 \leq j \leq 3$, where $R_{i,j} = \text{rank}(-X_{i,j})$. In the left-most plot we show the 3-dimensional data, the three plots on the right hand side show the two-dimensional projections $(1/R_{i,1}, 1/R_{i,2})$, $(1/R_{i,1}, 1/R_{i,3})$ and $(1/R_{i,2}, 1/R_{i,3})$. The second row shows the smoothed estimator $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_2, \theta_3)$ for $n = 257$, $\varepsilon = 0.06$ and $s = 3$; the left plot is a perspective plot, the middle one is a contour plot and the right one is a grey-scale image plot.

These 7-day swap rates show low and symmetric tail dependence which is reflected by many points lying near to the axes and the rest is scattered roughly uniformly with respect to the angles θ_2, θ_3 (first row of figure 1.7). The estimator $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_2, \theta_3)$ (second row) is therefore between 0.15 and 0.35 showing no significant difference between small and large angles.

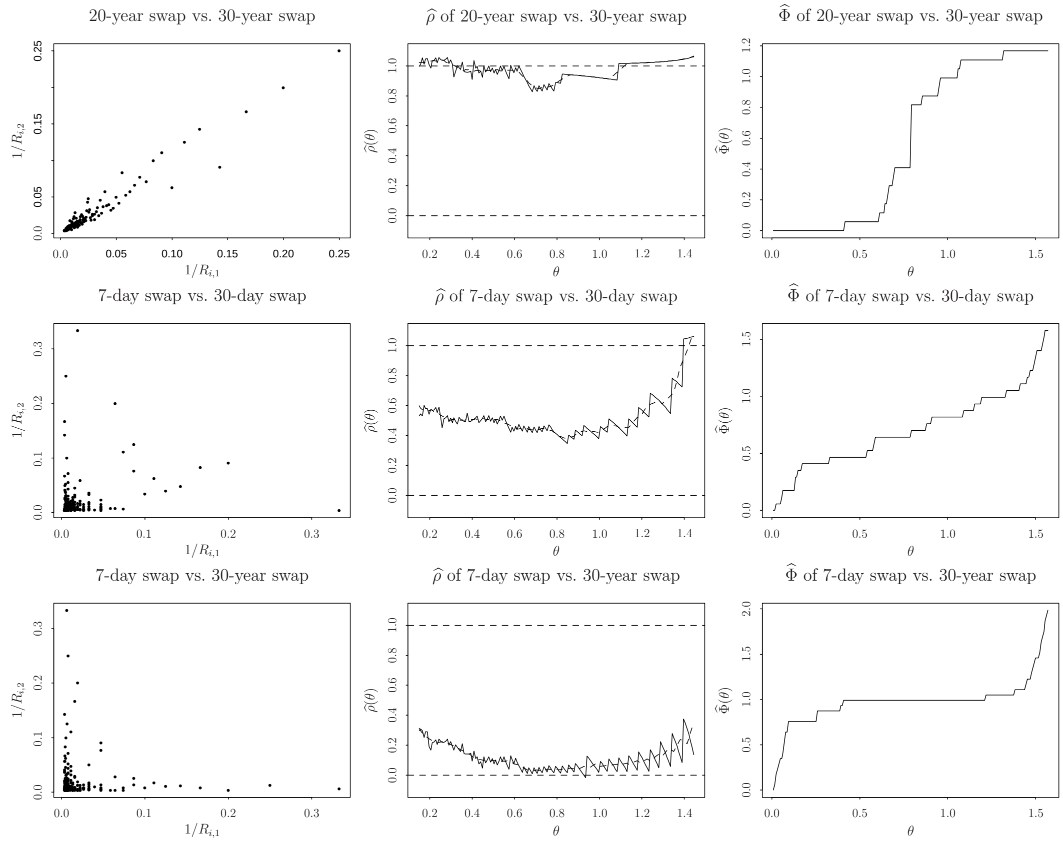

Figure 1.6: Estimators of ρ and Φ for some swap rates: 20-year vs. 30-year (first row), 7-day vs. 30-day (second row), and 7-day vs. 30-year swap rates (third row).
Left plots: transformed ranks $1/R_{i,j}, j = 1, 2$.
Middle plots: estimated tail dependence function $\widehat{\rho}$.
Right plots: estimated spectral measure $\widehat{\Phi}$ of the data.
In the first row we see high, in the second middle and in the third row low dependence.

Figure 1.8 shows the same as figure 1.7 for the high dependent EUR swap rates with maturities 5, 6 and 7 years. These swap rates with long and similar maturities show high and symmetric tail dependence which is reflected by all points lying near the diagonal (first row of figure 1.7). The estimator $\widehat{\rho}^{(s)}_{\varepsilon,n}(\theta_2, \theta_3)$ (second row) is therefore almost everywhere close to 1, only for angles θ_2, θ_3 near $\pi/4$ the estimator becomes smaller which is illustrated by the points that are away from the diagonal. □

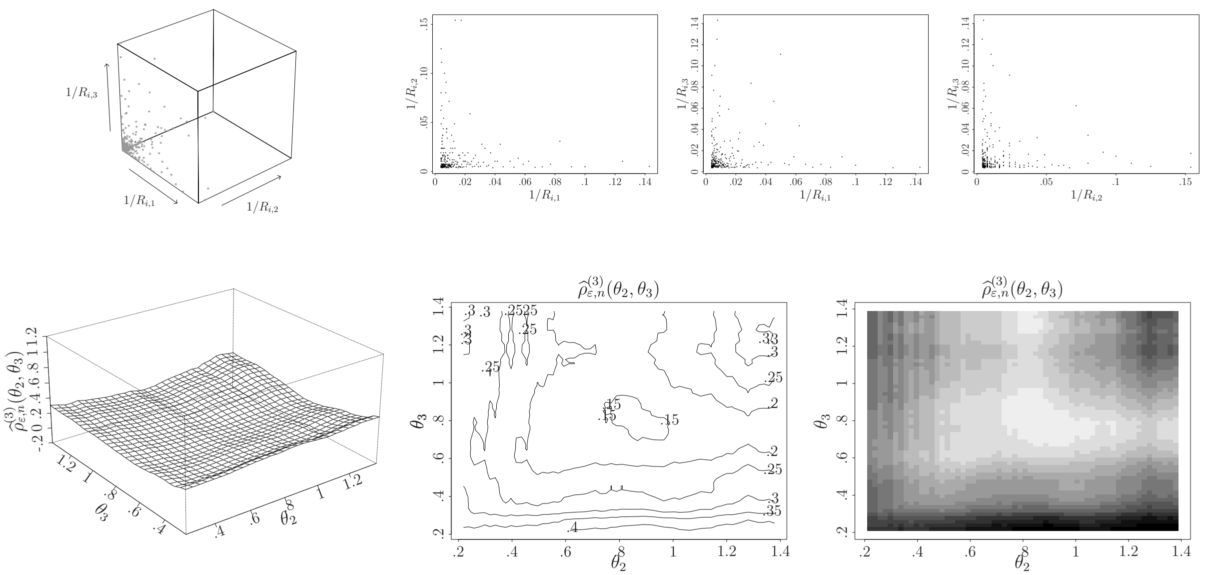

Figure 1.7: First row: data (3-d and 2-d projections) of the ranks $1/R_{i,j}$, $1 \le j \le 3$, of the low dependent 7-day swap rates rates in USD, EUR and GBP.
Second row: smoothed estimator $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2, \theta_3)$, perspective plot (left-most), contour plot (middle) and grey scale image plot (right-most)

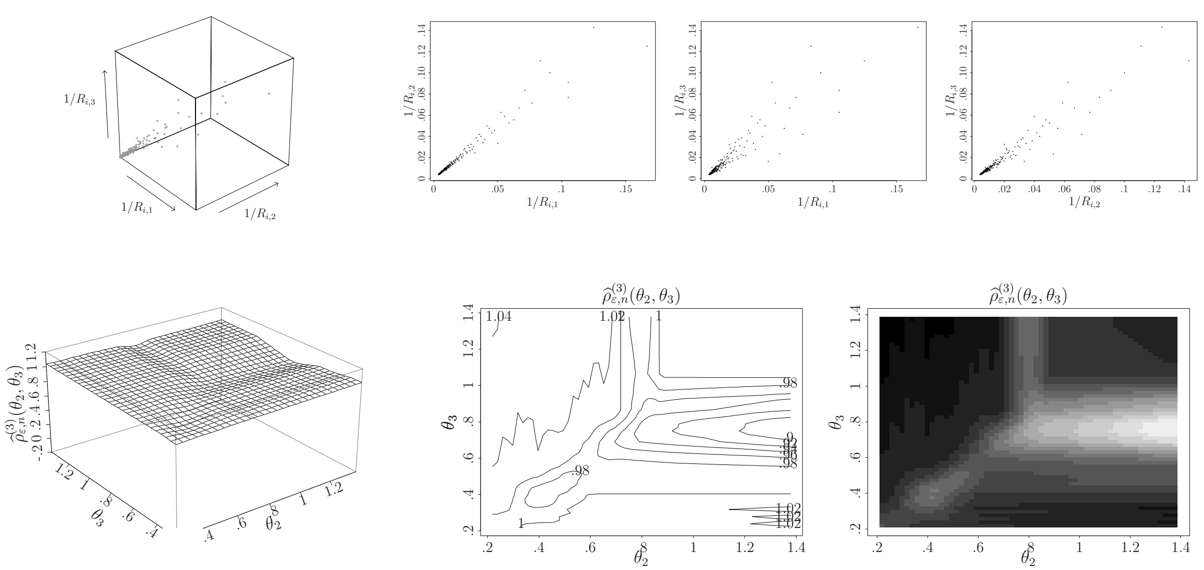

Figure 1.8: First row: data (3-d and 2-d projections) of the ranks $1/R_{i,j}$, $1 \le j \le 3$, of the high dependent 5-year, 6-year and 7-year EUR swap rates.
Second row: smoothed estimator $\widehat{\rho}^{(3)}_{\varepsilon,n}(\theta_2, \theta_3)$, perspective plot (left-most), contour plot (middle) and grey scale image plot (right-most)

1.7 Proofs

Proof of Proposition 1.3.2 (1): The case $\theta = \pi/2$ is obvious. Let $y_i = i/n$ if $\theta \in (0, \pi/4]$ and $x_i = i/n$ if $\theta \in (\pi/4, \pi/2)$, $1 \le i \le n$, then

$$\Lambda(C_\theta) = \begin{cases} \lim_{n\to\infty} \sum_{i=1}^{n} \left[\Lambda(y_i \cot\theta, y_i) - \Lambda(y_i \cot\theta, y_{i-1})\right], & \theta \in (0, \frac{\pi}{4}] \\ \lim_{n\to\infty} \sum_{i=1}^{n} \left[\Lambda(x_i, x_i \tan\theta) - \Lambda(x_i, x_{i-1}\tan\theta)\right] + \Lambda(1,1) - \Lambda(1, \tan\theta), & \theta \in (\frac{\pi}{4}, \frac{\pi}{2}). \end{cases}$$

Consider first $\theta \in (0, \pi/4]$. Note that for $x_1, x_2 \in [0, \infty]$,

$$\Lambda(x_1, x_2) = x_1 \int\limits_{\left(\arctan\frac{x_2}{x_1}, \frac{\pi}{2}\right]} \frac{1}{1 \vee \cot\gamma} \Phi(d\gamma) + x_2 \int\limits_{\left[0, \arctan\frac{x_2}{x_1}\right]} \frac{1}{1 \vee \tan\gamma} \Phi(d\gamma). \tag{1.7.1}$$

Thus, letting $\theta_i := \arctan\left(\frac{i-1}{i}\tan\theta\right)$,

$$\begin{aligned} &\Lambda\left(y_i \cot\theta, y_i\right) - \Lambda\left(y_i \cot\theta, y_{i-1}\right) \\ &= \frac{1}{n}\left(i \int\limits_{(\theta_i, \theta]} [(1 \wedge \cot\gamma) - (\cot\theta)(1 \wedge \tan\gamma)]\Phi(d\gamma) + \int\limits_{[0,\theta_i]} (1 \wedge \cot\theta)\,\Phi(d\gamma)\right) \\ &= \frac{1}{n}\left(i \int\limits_{(\theta_i, \theta]} [1 - (\cot\theta)(\tan\gamma)]\Phi(d\gamma) + \Phi[0, \theta_i]\right). \end{aligned}$$

Observe that $\sup_{\gamma\in(\theta_i,\theta]} i[1 - (\cot\theta)(\tan\gamma)] \le 1$. Since $\theta_i \to \theta$, we have

$$\limsup_{i\to\infty} i \int\limits_{(\theta_i,\theta]} [1 - (\cot\theta)(\tan\gamma)]\Phi(d\gamma) \le \Phi(\{\theta\}),$$

whereas $\Phi[0, \theta_i] \to \Phi[0, \theta)$ as $i \to \infty$. Applying Cesaro's mean value theorem we conclude that $\Lambda(C_\theta) = \Phi[0, \theta]$ for all $\theta \in (0, \pi/4]$ with $\Phi(\{\theta\}) = 0$. The case $\theta \in (\pi/4, \pi/2)$ can be dealt with similarly and the two cases combine to give $\Lambda(C_\theta) = \Phi[0, \theta]$ for all $\theta \in (0, \pi/2)$ with $\Phi(\{\theta\}) = 0$. Note that both $\Lambda(C_\theta)$ and $\Phi[0, \theta]$ are nondecreasing and right-continuous functions in θ. Since they agree on a dense subset of points in $[0, \pi/2]$ they must agree on the entire interval of $[0, \pi/2]$. This concludes the proof. □

Chapter 2

Estimating tail dependence of elliptical distributions

Summary

Recently there has been an increasing interest in applying elliptical distributions to risk management. Under weak conditions, Hult and Lindskog (2002) showed that a random vector with an elliptical distribution is in the domain of attraction of a multivariate extreme value distribution. In this chapter we study two estimators for the tail dependence function, which are based on extreme value theory and the structure of an elliptical distribution, respectively. After deriving second order regular variation estimates and proving asymptotic normality for both estimators, we show that the estimator based on the structure of an elliptical distribution is better than that based on extreme value theory in terms of both asymptotic variance and optimal asymptotic mean squared error. Our theoretical results are confirmed by a simulation study.

2.1 Introduction

Let $(X, Y), (X_1, Y_1), (X_2, Y_2), \cdots$ be independent random vectors with common distribution function F and continuous marginals F_X and F_Y. Define the *tail copula*

$$\lambda(x, y) \quad := \quad \lim_{t \to 0} \frac{1}{t} P\left(1 - F_X(X) \leq tx,\, 1 - F_Y(Y) \leq ty\right)$$

for $x, y \geq 0$, if the limit exists. Then $\lambda(1, 1)$ is called the *upper tail dependence coefficient*, see e.g. Joe (1997) and, by Huang (1992), $l(x, y) := x + y - \lambda(x, y)$ is called the *tail dependence function.* Assuming that (X, Y) is in the domain of attraction of a bivariate extreme value distribution, there exist several estimators for estimating the tail dependence function $l(x, y)$, see Huang (1992), Einmahl, de Haan, and Huang (1993) and de Haan and Resnick (1993). The optimal rate of convergence for estimating $l(x, y)$ is given by Drees and Huang (1998). An alternative method for estimating $l(x, y)$ is via estimating the spectral measure, see Einmahl, de Haan, and Sinha (1997) and Einmahl, de Haan, and Piterbarg (2001). For modeling dependence of extremes parametrically, we refer to Tawn (1988) and Ledford and Tawn (1997).

Triggered by financial risk management problems we observe an increasing interest in elliptical distributions as natural extensions of the normal family allowing for the modeling of heavy tails and extreme dependence. The vector (X, Y) is *elliptically distributed*, if

$$(X, Y)^T = \boldsymbol{\mu} + G\mathbf{A}\boldsymbol{U}^{(2)}, \tag{2.1.1}$$

where $\boldsymbol{\mu} = (\mu_X, \mu_Y)^{\mathrm{T}}$, $G > 0$ is a random variable, called *generating variable*, $\mathbf{A} \in \mathbb{R}^{2\times 2}$ is a deterministic matrix with

$$\mathbf{A}\mathbf{A}^T =: \boldsymbol{\Sigma} := \begin{pmatrix} \sigma^2 & \rho\sigma\upsilon \\ \rho\sigma\upsilon & \upsilon^2 \end{pmatrix}$$

and $\mathrm{rank}(\boldsymbol{\Sigma}) = 2$, $\boldsymbol{U}^{(2)}$ is a 2-dimensional random vector uniformly distributed on the unit hyper-sphere $\mathcal{S}_2 := \{\boldsymbol{z} \in \mathbb{R}^2 : \|\boldsymbol{z}\| = 1\}$, and $\boldsymbol{U}^{(2)}$ is independent of G.

Note that ρ is termed as the linear correlation coefficient of $\boldsymbol{\Sigma}$. Under some conditions, Hult and Lindskog (2002) showed that regular variation of $P(G > \cdot)$ with index $\alpha > 0$, i.e., $\lim_{t\to\infty} P(G > tx)/P(G > t) = x^{-\alpha}$ for all $x > 0$, (notation: $P(G > \cdot) \in \mathcal{R}_{-\alpha}$) implies that the regular variation of (X, Y) with the same index $\alpha > 0$ (see Resnick (1987) for the definition of multivariate regular variation). Moreover, if $P(G > \cdot) \in \mathcal{R}_{-\alpha}$, then

$$\lambda(1, 1) = \left(\int_{(\pi/2 - \arcsin\rho)/2}^{\pi/2} (\cos\phi)^\alpha \, \mathrm{d}\phi \right) \Big/ \left(\int_0^{\pi/2} (\cos\phi)^\alpha \, \mathrm{d}\phi \right). \tag{2.1.2}$$

Here we are interested in estimating the dependence function $\lambda(x, y)$ by assuming that $P(G > \cdot) \in RV_{-\alpha}$ for some $\alpha > 0$. Since $P(G > \cdot) \in RV_{-\alpha}$ implies that (X, Y) is in the domain of attraction of an extreme value distribution, a naive estimator is to apply Huang's estimator by ignoring the structure of the elliptical distribution, i.e.,

$$\widehat{\lambda}^{\mathrm{Hu}}_{k_{\mathrm{Hu}},n}(x, y) := \frac{1}{k_{\mathrm{Hu}}} \sum_{i=1}^{n} \boldsymbol{I}\left(X_i \geq X_{(n-\lfloor xk_{\mathrm{Hu}}\rfloor, n)},\ Y_i \geq Y_{(n-\lfloor yk_{\mathrm{Hu}}\rfloor, n)} \right), \tag{2.1.3}$$

where $X_{(1,n)} \le \cdots \le X_{(n,n)}$ and $Y_{(1,n)} \le \cdots \le Y_{(n,n)}$ denote the order statistics of $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$, respectively, $k_{\mathrm{Hu}} = k_{\mathrm{Hu}}(n) \stackrel{n\to\infty}{\longrightarrow} \infty$ and $k_{\mathrm{Hu}}/n \stackrel{n\to\infty}{\longrightarrow} 0$. The same estimator has been analyzed by Schmidt and Stadtmüller (2006); see their equation (4.14). The aim of this chapter is two-fold. Firstly, we suggest a new estimator, which exploits the structure of an elliptical distribution similar to (2.1.2). Secondly, we aim at determining the optimal number of order statistics to be used in both estimators. The choice will be based on the asymptotic mean squared error of the estimators.

Our chapter is organized as follows. We first derive an expression for $\lambda(x,y)$, which generalizes equation (2.1.2), and then construct a new estimator for $\lambda(x,y)$ via this expression; see section 2 for details. After deriving the second order behavior for elliptical distributions and the limiting distributions of both estimators in section 2, we show that the new estimator is better than the naive empirical estimator from Huang in terms of both asymptotic variance and optimal asymptotic mean squared error in section 3. More importantly, the optimal choice of the sample fraction for the new estimator is the same as that for Hill's estimator (Hill (1975)). That is, all data-driven methods for choosing the optimal sample fraction for Hill's estimator can be applied to our new estimator directly. A simulation study is provided in section 3 and all proofs are summarized in section 4.

2.2 Methodology and Main Results

The following theorem gives an expression for $\lambda(x,y)$, which will be employed to construct an estimator.

Theorem 2.2.1. *Suppose (X,Y) defined in (2.1.1) holds with $\sigma > 0$, $\upsilon > 0$, $|\rho| < 1$ and $1-G \in \mathcal{R}_{-\alpha}$ for some $\alpha > 0$. Further, define*

$$g(t) \quad := \quad \arctan\left((t-\rho)/\sqrt{1-\rho^2}\right) \in [-\arcsin\rho, \pi/2], \quad t \in \mathbb{R}.$$

Then

$$\lambda(x,y) \quad = \quad \left(\int_{-\pi/2}^{\pi/2} (\cos\phi)^\alpha \, \mathrm{d}\phi\right)^{-1} \left(\int_{g\left((x/y)^{1/\alpha}\right)}^{\pi/2} x(\cos\phi)^\alpha \, \mathrm{d}\phi + \int_{-\arcsin\rho}^{g\left((x/y)^{1/\alpha}\right)} y\left(\sin(\phi + \arcsin\rho)\right)^\alpha \, \mathrm{d}\phi\right).$$

□

In order to derive the asymptotic normality of $\widehat{\lambda}^{\mathrm{Hu}}_{k_{\mathrm{Hu}},n}(x,y)$, it is known that a second order condition is needed. Here we seek the relation of the second order behavior among the tail copula $\lambda(x,y)$, $\sqrt{X^2+Y^2}$ and G; see the next two theorems for details.

In the setting of (2.1.1) assume that there exists $A \in RV_\beta$ such that for all $x > 0$ and some $\beta \leq 0$ (hence $A(t) \to 0$)

$$\lim_{t\to\infty} \frac{P(G \geq tx)/P(G \geq t) - x^{-\alpha}}{A(t)} = x^{-\alpha}\frac{x^\beta - 1}{\beta}, \tag{2.2.1}$$

where $\beta \leq 0$ is called a *second order regular variation parameter*, see de Haan and Stadtmüller (1996). Additionally, we assume

$$\lim_{t\to\infty} t^2 A(t) =: a \in [-\infty, \infty]. \tag{2.2.2}$$

Since $A \in \mathcal{R}_\beta$, it holds that $(\cdot)^2 A(\cdot) \in RV_{2+\beta}$, therefore $t^2 A(t) \overset{t\to\infty}{\longrightarrow} a = 0$ for $\beta < -2$ and $t^2 A(t) \overset{t\to\infty}{\longrightarrow} a = \pm\infty$ for $\beta \in (-2, 0]$.

The following two theorems derive the corresponding second order condition for $\sqrt{X^2 + Y^2}$ and the tail copula $\lambda(x, y)$.

Theorem 2.2.2. *Assume that the conditions of Theorem 2.2.1, (2.2.1) and (2.2.2) hold. Further, define*

$$\begin{aligned} d_1(\phi) &= \sigma^2 \cos^2\phi + \upsilon^2 \sin^2(\phi + \arcsin\rho), \\ d_2(\phi) &= \mu_X \sigma \cos\phi + \mu_Y \upsilon \sin(\phi + \arcsin\rho). \end{aligned}$$

Then, for all $x > 0$,

$$\begin{aligned} &\lim_{t\to\infty} \frac{P\{\sqrt{X^2+Y^2} \geq tx\}/P\{\sqrt{X^2+Y^2} \geq t\} - x^{-\alpha}}{t^{-2} + |A(t)|} \\ &= x^{-\alpha}\left(\int_{-\pi}^{\pi} (d_1(\phi))^{\alpha/2}\,\mathrm{d}\phi\right)^{-1} \left\{\frac{a}{1+|a|}\frac{x^\beta - 1}{\beta}\left(\int_{-\pi}^{\pi} (d_1(\phi))^{(\alpha-\beta)/2}\,\mathrm{d}\phi\right)\right. \\ &\quad + \frac{1}{1+|a|}\frac{\alpha}{2}\left(x^{-2} - 1\right)\int_{-\pi}^{\pi} (d_1(\phi))^{\alpha/2-1} \times \\ &\quad\quad \left.\times \left[\alpha(d_2(\phi))^2 + d_1(\phi)(\mu_X^2 + \mu_Y^2)\right]\mathrm{d}\phi\right\}. \end{aligned} \tag{2.2.3}$$

Also, for all $x > 0$ and $V(x) := \inf\{y : P(\sqrt{X^2+Y^2} > y) \le x^{-1}\}$,

$$
\begin{aligned}
&\lim_{t\to\infty} \frac{V(tx)/V(t) - x^{1/\alpha}}{(F_Y^{\leftarrow}(1-t^{-1}))^{-2} + |A(F_Y^{\leftarrow}(1-t^{-1}))|} \\
&= x^{1/\alpha} \left(\int_{-\pi}^{\pi} (d_1(\phi))^{\alpha/2}\,\mathrm{d}\phi\right)^{-1} \left(\frac{\int_{-\pi}^{\pi}(d_1(\phi))^{\alpha/2}\,d\phi}{\upsilon^\alpha \int_0^\pi (\sin\phi)^\alpha\,d\phi}\right)^{-(2\wedge|\beta|)/\alpha} \times \\
&\quad \times \left\{\frac{a}{1+|a|}\,\frac{x^{\beta/\alpha}-1}{\alpha\beta}\left(\int_{-\pi}^{\pi}(d_1(\phi))^{(\alpha-\beta)/2}\,\mathrm{d}\phi\right)\right. \\
&\quad + \frac{1}{1+|a|}\frac{1}{2}\left(x^{-2/\alpha}-1\right)\int_{-\pi}^{\pi}(d_1(\phi))^{\alpha/2-1} \times \\
&\qquad \left. \times \left[\alpha(d_2(\phi))^2 + d_1(\phi)(\mu_X^2+\mu_Y^2)\right]\mathrm{d}\phi\right\} =: x^{1/\alpha}\mathcal{B}_{(2.2.4)}(x). \qquad (2.2.4)
\end{aligned}
$$

Especially, when $\mu_X = \mu_Y = 0$, we have for all $x > 0$

$$
\begin{aligned}
\lim_{t\to\infty} \frac{V(tx)/V(t) - x^{1/\alpha}}{A(F_Y^{\leftarrow}(1-t^{-1}))} &= x^{1/\alpha}\frac{x^{\beta/\alpha}-1}{\upsilon^\beta\alpha\beta}\left(\int_{-\pi}^{\pi}(d_1(\phi))^{\alpha/2}\,\mathrm{d}\phi\right)^{-1} \times \\
&\quad \times \left(\int_{-\pi}^{\pi}(d_1(\phi))^{(\alpha-\beta)/2}\,\mathrm{d}\phi\right)\left(\frac{\int_{-\pi}^{\pi}(d_1(\phi))^{\alpha/2}\,d\phi}{\int_0^\pi(\sin\phi)^\alpha\,d\phi}\right)^{\beta/\alpha} \\
&=: x^{1/\alpha}\mathcal{B}_{(2.2.5)}(x). \qquad (2.2.5)
\end{aligned}
$$

□

Theorem 2.2.3. *Assume that the conditions of Theorem 2.2.1 and (2.2.1) hold. Further, define*

$$
\begin{aligned}
\mathcal{S}_2^+ &:= \left\{\boldsymbol{z} \in \mathbb{R}^2 : \boldsymbol{z} \ge \mathbf{0} \text{ and } \|\boldsymbol{z}\| = 1\right\} \quad \textit{and} \\
\mathcal{B}_{(2.2.6)}(x) &:= -x\frac{x^{-\beta/\alpha}-1}{\beta}\left(\int_0^\pi(\sin\phi)^\alpha \mathrm{d}\,\phi\right)^{-1}\left(\int_0^\pi(\sin\phi)^{\alpha-\beta}\,\mathrm{d}\phi\right). \qquad (2.2.6)
\end{aligned}
$$

Then,

$$\begin{aligned}
&\lim_{t\to 0}\frac{t^{-1}P\left(F_X(X)\geq 1-tx,\ F_Y(Y)\geq 1-ty\right)-\lambda(x,y)}{A\left(F_Y^{\leftarrow}(1-t)\right)}\\
&= \upsilon^{-\beta}\left\{\frac{x}{\beta}\int_{g((x/y)^{1/\alpha})}^{\pi/2}\left[x^{-\beta/\alpha}(\cos\phi)^{\alpha-\beta}-(\cos\phi)^{\alpha}\right]\mathrm{d}\phi\right.\\
&\quad+\frac{y}{\beta}\int_{-\arcsin\rho}^{g((x/y)^{1/\alpha})}\left[y^{-\beta/\alpha}(\sin(\phi+\arcsin\rho))^{\alpha-\beta}-(\sin(\phi+\arcsin\rho))^{\alpha}\right]\mathrm{d}\phi\\
&\quad+\mathcal{B}_{(2.2.6)}(x)\int_{g((x/y)^{1/\alpha})}^{\pi/2}(\cos\phi)^{\alpha}\,d\phi\\
&\quad+\mathcal{B}_{(2.2.6)}(y)\int_{-\arcsin\rho}^{g((x/y)^{1/\alpha})}(\sin(\phi+\arcsin\rho))^{\alpha}\,d\phi\\
&\quad\left.-\lambda(x,y)\frac{1}{\beta}\int_{-\pi/2}^{\pi/2}(\cos\phi)^{\alpha}\left((\cos\phi)^{-\beta}-1\right)\mathrm{d}\phi\right\}\left(\int_{-\pi/2}^{\pi/2}(\cos\phi)^{\alpha}\,\mathrm{d}\phi\right)^{-1}\\
&=: \mathcal{B}_{(2.2.7)}(x,y)
\end{aligned}\tag{2.2.7}$$

holds for all $x,y\geq 0$ and uniformly on $\mathcal{S}_2^+$. □

Now we are ready to define our new estimator. Put $Z_i=\sqrt{X_i^2+Y_i^2}$ for $i=1,\ldots,n$ and let $Z_{(1,n)}\leq\cdots\leq Z_{(n,n)}$ denote their order statistics. First we estimate the index α by Hill's estimator, which is defined as

$$\widehat{\alpha}^{\mathrm{H}}_{k_{\mathrm{El}},n} := \left(\frac{1}{k_{\mathrm{El}}}\sum_{i=1}^{k_{\mathrm{El}}}\log Z_{(n-i+1,n)}-\log Z_{(n-k_{\mathrm{El}},n)}\right)^{-1},$$

where $k_{\mathrm{El}}=k_{\mathrm{El}}(n)\to\infty$ and $k_{\mathrm{El}}/n\to 0$ as $n\to\infty$. Now let (X,Y) and $(\widetilde{X},\widetilde{Y})$ be iid with elliptical distribution. Then, it follows from Hult and Lindskog (2002) that $\tau=(2/\pi)\arcsin\rho$, where τ is called *Kendall's tau* and defined by

$$\tau := P\left(\left(X-\widetilde{X}\right)\left(Y-\widetilde{Y}\right)>0\right)-P\left(\left(X-\widetilde{X}\right)\left(Y-\widetilde{Y}\right)<0\right).$$

As usual, we estimate Kendall's tau by

$$\widehat{\tau}_n := \frac{2}{n(n-1)}\sum_{1\leq i,j\leq n}\operatorname{sign}\left((X_i-X_j)(Y_i-Y_j)\right),$$

which results in estimating ρ by

$$\widehat{\rho}_n=\sin\left(\frac{\pi}{2}\widehat{\tau}_n\right).$$

Hence, we can estimate $\lambda(x,y)$ by replacing ρ and α in Theorem 2.2.1 by $\widehat{\rho}_n$ and $\widehat{\alpha}^{\mathrm{H}}_{k_{\mathrm{El}},n}$, respectively. Let us denote this estimator by

$$\widehat{\lambda}^{\mathrm{El}}_{k_{\mathrm{El}},n}(x,y). \tag{2.2.8}$$

We remark that $\widehat{\lambda}^{\mathrm{El}}_{k_{\mathrm{El}},n}(1,1)$ was mentioned by Schmidt (2003), but without further study. The following theorem shows the asymptotic normalities of $\widehat{\lambda}^{\mathrm{Hu}}_{k_{\mathrm{Hu}},n}(x,y)$ and $\widehat{\lambda}^{\mathrm{El}}_{k_{\mathrm{El}},n}(x,y)$, which allows us to compare these two estimators theoretically.

Theorem 2.2.4. *Assume that the conditions of Theorem 2.2.1 and (2.2.1) hold. Suppose* $k_{\mathrm{Hu}} = k_{\mathrm{Hu}}(n) \overset{n\to\infty}{\longrightarrow} \infty$, $k_{\mathrm{Hu}}/n \overset{n\to\infty}{\longrightarrow} 0$ *and*

$$\sqrt{k_{\mathrm{Hu}}}A(F_Y^{\leftarrow}(1-k_{\mathrm{Hu}}/n)) \overset{n\to\infty}{\longrightarrow}: \mathcal{K}_{\mathrm{Hu}},$$

for $|\mathcal{K}_{\mathrm{Hu}}|<\infty$. *Then, as* $n\to\infty$,

$$\sup_{0\le x,y\le T}\left|\sqrt{k_{\mathrm{Hu}}}\left(\widehat{\lambda}^{\mathrm{Hu}}_{k_{\mathrm{Hu}},n}(x,y)-\lambda(x,y)\right)-\mathcal{K}_{\mathrm{Hu}}\mathcal{B}_{(2.2.7)}(x,y)-B(x,y)\right| = o_p(1), \tag{2.2.9}$$

for any $T>0$, *where* $\mathcal{B}_{(2.2.7)}(x,y)$ *is defined in Theorem 2.2.3,*

$$B(x,y) = W(x,y)-\left(1-\frac{\partial\lambda(x,y)}{\partial x}\right)W(x,0)-\left(1-\frac{\partial\lambda(x,y)}{\partial y}\right)W(0,y),$$

and $W(x,y)$ *is a Wiener process with zero mean and covariance structure*

$$\begin{aligned}&E\left(W(x_1,y_1)W(x_2,y_2)\right)\\ &= x_1\wedge x_2+y_1\wedge y_2-\lambda(x_1\wedge x_2,y_1)-\lambda(x_1\wedge x_2,y_2)-\lambda(x_1,y_1\wedge y_2)\\ &\quad -\lambda(x_2,y_1\wedge y_2)+\lambda(x_1,y_2)+\lambda(x_2,y_1)+\lambda(x_1\wedge x_2,y_1\wedge y_2).\end{aligned}$$

Therefore, for any fixed $x,y>0$,

$$\sqrt{k_{\mathrm{Hu}}}\left(\widehat{\lambda}^{\mathrm{Hu}}_{k_{\mathrm{Hu}},n}(x,y)-\lambda(x,y)\right)\overset{d}{\longrightarrow}\mathcal{N}\left(\mathcal{K}_{\mathrm{Hu}}\mathcal{B}_{(2.2.7)}(x,y),\sigma^2_{\mathrm{Hu}}\right)$$

as $n\to\infty$, *where*

$$\begin{aligned}\sigma^2_{\mathrm{Hu}} &= x\left(\frac{\partial}{\partial x}\lambda(x,y)\right)^2+y\left(\frac{\partial}{\partial y}\lambda(x,y)\right)^2+2\lambda(x,y)\times\\ &\times\left(\frac{1}{2}-\frac{\partial}{\partial x}\lambda(x,y)-\frac{\partial}{\partial y}\lambda(x,y)+\left(\frac{\partial\lambda(x,y)}{\partial x}\right)\left(\frac{\partial\lambda(x,y)}{\partial y}\right)\right),\end{aligned} \tag{2.2.10}$$

$$\frac{\partial}{\partial x}\lambda(x,y) = \left(\int_{-\pi/2}^{\pi/2}(\cos\phi)^\alpha\,\mathrm{d}\phi\right)^{-1}\int_{g((x/y)^{1/\alpha})}^{\pi/2}(\cos\phi)^\alpha\,\mathrm{d}\phi \quad \textit{and} \tag{2.2.11}$$

$$\frac{\partial}{\partial y}\lambda(x,y) = \left(\int_{-\pi/2}^{\pi/2}(\cos\phi)^\alpha\,\mathrm{d}\phi\right)^{-1}\times \tag{2.2.12}$$

$$\times\int_{-\arcsin\rho}^{g((x/y)^{1/\alpha})}(\sin(\phi+\arcsin\rho))^\alpha\,\mathrm{d}\phi.$$

□

Theorem 2.2.5. *Assume that the conditions of Theorem 2.2.1 and (2.2.1) hold. Further assume (2.2.2) holds when* $\boldsymbol{\mu}\neq 0$*. Suppose* $k_{\mathrm{El}}=k_{\mathrm{El}}(n,\boldsymbol{\mu})\overset{n\to\infty}{\longrightarrow}\infty$*,* $k_{\mathrm{El}}/n\overset{n\to\infty}{\longrightarrow}0$ *and*

$$\sqrt{k_{\mathrm{El}}}\Big(\,(F_Y^{\leftarrow}(1-k_{\mathrm{El}}/n))^{-2}+|A(F_Y^{\leftarrow}(1-k_{\mathrm{El}}/n))|\,\Big)\overset{n\to\infty}{\longrightarrow}:\ \mathcal{K}_{\mathrm{El}},\quad \boldsymbol{\mu}\neq 0,$$

$$\sqrt{k_{\mathrm{El}}}A(F_Y^{\leftarrow}(1-k_{\mathrm{El}}/n))\overset{n\to\infty}{\longrightarrow}:\ \mathcal{K}_{\mathrm{El}},\quad \boldsymbol{\mu}=\mathbf{0},$$

for $|\mathcal{K}_{\mathrm{El}}|<\infty$ *Then, as* $n\to\infty$*,*

$$\sup_{0\le x,y\le T}\left|\sqrt{k_{\mathrm{El}}}\left(\widehat{\lambda}^{\mathrm{El}}_{k_{\mathrm{El}},n}(x,y)-\lambda(x,y)\right)-\mathcal{B}_{(2.2.15)}(x,y)Z_0\right| = o_p(1), \tag{2.2.13}$$

where $Z_0\sim\mathcal{N}\left(-\alpha^2\mathcal{K}_{\mathrm{El}}\mathcal{B}_{(2.2.14)},\,\alpha^2\right)$ *with*

$$\mathcal{B}_{(2.2.14)} := \begin{cases}\int_0^1\mathcal{B}_{(2.2.4)}(1/s)\,\mathrm{d}s, & \boldsymbol{\mu}\neq\mathbf{0},\\ \int_0^1\mathcal{B}_{(2.2.5)}(1/s)\,\mathrm{d}s, & \boldsymbol{\mu}=\mathbf{0},\end{cases} \tag{2.2.14}$$

$\mathcal{B}_{(2.2.4)}(s)$ *and* $\mathcal{B}_{(2.2.5)}(s)$ *are defined in Theorem 2.2.2 and*

$$\begin{aligned}\mathcal{B}_{(2.2.15)}(x,y) := \Bigg\{&\int_{g\left((x/y)^{1/\alpha}\right)}^{\pi/2}x(\cos\phi)^\alpha\ln(\cos\phi)\,\mathrm{d}\phi\\ &+\int_{-\arcsin\rho}^{g\left((x/y)^{1/\alpha}\right)}y(\sin(\phi+\arcsin\rho))^\alpha\ln(\sin(\phi+\arcsin\rho))\,\mathrm{d}\phi\\ &-\lambda(x,y)\left(\int_{-\pi/2}^{\pi/2}(\cos\phi)^\alpha\ln(\cos\phi)\,\mathrm{d}\phi\right)\Bigg\}\left(\int_{-\pi/2}^{\pi/2}(\cos\phi)^\alpha\,\mathrm{d}\phi\right)^{-1}.\end{aligned} \tag{2.2.15}$$

Therefore, for any fixed $x,y>0$*,*

$$\begin{aligned}&\sqrt{k_{\mathrm{El}}}\left(\widehat{\lambda}^{\mathrm{El}}_{k_{\mathrm{El}},n}(x,y)-\lambda(x,y)\right)\\ &\overset{d}{\longrightarrow}\ \mathcal{N}\left(-\alpha^2\mathcal{K}_{\mathrm{El}}\mathcal{B}_{(2.2.14)}\mathcal{B}_{(2.2.15)}(x,y),\,\alpha^2\left(\mathcal{B}_{(2.2.15)}(x,y)\right)^2\right).\end{aligned}$$

□

The next corollary gives the optimal choice of sample fraction for both estimators. As criterion we use the *asymptotic mean squared error* of $\widehat{\lambda}^{\mathrm{Hu}}_{k_{\mathrm{Hu}},n}$ and $\widehat{\lambda}^{\mathrm{El}}_{k_{\mathrm{El}},n}$, denoted by $\mathrm{amse}_{\mathrm{Hu}}(k_{\mathrm{Hu}})$ and $\mathrm{amse}_{\mathrm{El}}(k_{\mathrm{El}})$, respectively.

Corollary 2.2.6. *Assume that the conditions of Theorems 2.2.4 and 2.2.5 hold. Further, suppose that*

$$\begin{aligned} A(F_Y^{\leftarrow}(1-t)) &\sim b_0 t^{-\beta/\alpha}, \\ (F_Y^{\leftarrow}(1-t))^{-2} + |A(F_Y^{\leftarrow}(1-t))| &\sim b_1 t^{(2\wedge(-\beta))/\alpha} \end{aligned}$$

for some $b_0, b_1 > 0$ as $t \to 0$ and define

$$b_2 t^{-\beta_2/\alpha} := \begin{cases} b_1 t^{(2\wedge(-\beta))/\alpha} & \boldsymbol{\mu} \neq \mathbf{0}, \\ b_0 t^{-\beta/\alpha}, & \boldsymbol{\mu} = 0. \end{cases}$$

Then

$$\mathrm{amse}_{\mathrm{Hu}}(k_{\mathrm{Hu}}) = \sigma^2_{\mathrm{Hu}} k^{-1}_{\mathrm{Hu}} + \left(b(k_{\mathrm{Hu}}/n)^{-\beta/\alpha}\mathcal{B}_{(2.2.7)}(x,y)\right)^2$$

and

$$\mathrm{amse}_{\mathrm{El}}(k_{\mathrm{El}}) = (\mathcal{B}_{(2.2.15)}(x,y))^2 \left(\alpha^2 k^{-1}_{\mathrm{El}} + \left(\alpha^2 b_2 (k_{\mathrm{El}}/n)^{-\beta_2/\alpha}\mathcal{B}_{(2.2.14)}\right)^2\right).$$

Let $k^{\mathrm{opt}}_{\mathrm{Hu}}$ and $k^{\mathrm{opt}}_{\mathrm{El}}$ denote the optimal sample fraction in the sense of minimizing $\mathrm{amse}_{\mathrm{Hu}}$ *and* $\mathrm{amse}_{\mathrm{El}}$*, respectively. Then*

$$\begin{aligned} k^{\mathrm{opt}}_{\mathrm{Hu}} &= \left(\frac{-\alpha\sigma^2_{\mathrm{Hu}}}{2\beta b_0^2 \left(\mathcal{B}_{(2.2.7)}(x,y)\right)^2}\right)^{\alpha/(\alpha-2\beta)} n^{-2\beta/(\alpha-2\beta)}, \\ k^{\mathrm{opt}}_{\mathrm{El}} &= \left(-2\beta_2\alpha b_2^2 \left(\mathcal{B}_{(2.2.14)}\right)^2\right)^{-\alpha/(\alpha-2\beta_2)} n^{-2\beta_2/(\alpha-2\beta_2)}, \\ \mathrm{amse}^{\mathrm{opt}}_{\mathrm{Hu}} &:= \mathrm{amse}_{\mathrm{Hu}}\left(k^{\mathrm{opt}}_{\mathrm{Hu}}\right) = n^{2\beta/(\alpha-2\beta)} b_0^{2\alpha/(\alpha-2\beta)} \left(1 - \frac{\alpha}{2\beta}\right) \times \\ &\quad \times \left(\left(\sigma^2_{\mathrm{Hu}}\right)^{-\beta/\alpha} \mathcal{B}_{(2.2.7)}(x,y)\sqrt{-2\beta/\alpha}\right)^{2\alpha/(\alpha-2\beta)} \quad \text{and} \\ \mathrm{amse}^{\mathrm{opt}}_{\mathrm{El}} &:= \mathrm{amse}_{\mathrm{El}}\left(k^{\mathrm{opt}}_{\mathrm{El}}\right) = n^{2\beta_2/(\alpha-2\beta_2)} b_2^{2\alpha/(\alpha-2\beta_2)} \left(1 - \frac{\alpha}{2\beta_2}\right) \times \\ &\quad \times \alpha^2 (\mathcal{B}_{(2.2.15)}(x,y))^2 \left(\sqrt{-2\alpha\beta_2}\mathcal{B}_{(2.2.14)}\right)^{2\alpha/(\alpha-2\beta_2)}. \end{aligned}$$

□

Remark 2.2.7. Note that $k_{\mathrm{El}}^{\mathrm{opt}}$ is independent of x and y, but $k_{\mathrm{Hu}}^{\mathrm{opt}}$ depends on x and y. In case of $\boldsymbol{\mu} = \mathbf{0}$, both $\mathrm{amse}_{\mathrm{Hu}}^{\mathrm{opt}}$ and $\mathrm{amse}_{\mathrm{El}}^{\mathrm{opt}}$ depend on n, α, β, ρ, υ, x, y and b_0, $\mathrm{amse}_{\mathrm{El}}^{\mathrm{opt}}$ additionally depends on σ, but the ratio $\mathrm{amse}_{\mathrm{Hu}}^{\mathrm{opt}}/\mathrm{amse}_{\mathrm{El}}^{\mathrm{opt}}$ is independent of n and b_0. Since the optimal $k_{\mathrm{El}}^{\mathrm{opt}}$ is the same as that for Hill's estimator, when $\boldsymbol{\mu} = 0$, all data-driven methods for choosing the optimal sample fraction for Hill's estimator can be applied to $\widehat{\lambda}_{k_{\mathrm{El}},n}^{\mathrm{El}}(x,y)$ directly. Note that $\boldsymbol{\mu}$ is the median of (X,Y) and the mean of (X,Y) when $\alpha > 1$. Hence, we could estimate $\boldsymbol{\mu}$ by the sample median, say $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_X, \widehat{\mu}_Y)$. Therefore, consider the new estimator $\widehat{\lambda}_{k_{\mathrm{El}},n}^{\mathrm{El}}(x,y)$ with $Z_i = \sqrt{X_i^2 + Y_i^2}$ replaced by $\sqrt{(X_i - \widehat{\mu}_X)^2 + (Y_i - \widehat{\mu}_Y)^2}$. Similar to the proofs in Ling and Peng (2004), we can show that Theorem 2.2.5 and Corollary 2.2.6 hold with $\boldsymbol{\mu} = 0$ for this new estimator. □

2.3 Comparisons and Simulation Study

First we compare $\sigma_{\mathrm{Hu}}^2, \sigma_{\mathrm{El}}^2$ given in Theorem 2.2.4 and 2.2.5. Note that both only depend on α, ρ, x and y. In Figure 2.1, we plot the ratio $\sigma_{\mathrm{El}}^2(\alpha)/\sigma_{\mathrm{Hu}}^2(\alpha)$ for $x = y = 1$ as a function of α, and each curve therein corresponds to a different correlation $\rho \in \{0.1, \ldots, 0.9\}$. From Figure 2.1, we conclude that $\widehat{\lambda}_{k,n}^{\mathrm{El}}$ is always better in terms of asymptotic variance.

Second, we compare the two estimators in terms of optimal asymptotic mean squared errors. Since the ratio of the optimal asymptotic mean squared error depends on α, β, $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, x, y, we consider elliptical distributions with $\sigma = \upsilon = 1$, $\mu_X = \mu_Y = 0$. In Figure 2.2, we consider $G \sim$ Fréchet(α), i.e. $P(G > x) = \exp(-x^{-\alpha})$, $x > 0$, hence (2.2.1) is satisfied with $\beta = -\alpha$. In Figure 2.3, we consider $G \sim$ Pareto(α), i.e. $P(G > x) = (1+x)^{-\alpha}$ for $x > 0$, therefore, (2.2.1) is satisfied with $\beta = -1$. Under the above setup, the ratio of optimal asymptotic mean squared errors only depends on α, ρ, x, y. Similar to Figure 2.1, we plot the ratio $\mathrm{amse}_{\mathrm{El}}^{\mathrm{opt}}(\alpha)/\mathrm{amse}_{\mathrm{Hu}}^{\mathrm{opt}}(\alpha)$ for $x = y = 1$ as a function of α for different ρ's in Figures 2.2 and 2.3. We conclude from both Figures that $\widehat{\lambda}_{k,n}^{\mathrm{El}}$ always performs better than $\widehat{\lambda}_{k,n}^{\mathrm{Hu}}$ in terms of optimal asymptotic mean squared errors as well.

Third, we examine the influence of x and y to the ratio of asymptotic mean squared error. We plot the ratio $\mathrm{amse}_{\mathrm{El}}^{\mathrm{opt}}(\alpha)/\mathrm{amse}_{\mathrm{Hu}}^{\mathrm{opt}}(\alpha)$ for $\|(x,y)\| = \sqrt{2}$ and $G \sim$ Pareto(α) in Figure 2.4, where each curve corresponds to a different pair of $(\alpha,\, \rho) \in \{(20, 0.9), (10, 0.6), (5, 0.3), (1, 0.1)\}$. This figure further confirms that $\widehat{\lambda}_{k,n}^{\mathrm{El}}$ always has a smaller optimal asymptotic mean squared error than $\widehat{\lambda}_{k,n}^{\mathrm{Hu}}$.

Finally, we study the finite sample behavior of the two estimators $\widehat{\lambda}_{k,n}^{\mathrm{El}}(x,y)$ and $\widehat{\lambda}_{k,n}^{\mathrm{Hu}}(x,y)$. As above, we consider two elliptical distributions with $\sigma = \upsilon = 1$, $\mu_X =$

$\mu_Y = 0$, $G \sim$ Fréchet(α) in Figure 2.5 and $G \sim$ Pareto(α) in Figure 2.6. We generate 1000 random samples of size $n = 1000$ from these elliptical distributions with $(\alpha, \rho) \in \{(20, 0.9), (10, 0.6), (5, 0.3), (1, 0.1)\}$, and plot $\widehat{\lambda}^{\mathrm{El}}_{k,n}(1, 1)$ and $\widehat{\lambda}^{\mathrm{Hu}}_{k,n}(1, 1)$ against $k = 1, \ldots, 300$ for different pairs (α, ρ) in Figures 2.5 and 2.6, where the solid line corresponds to $\widehat{\lambda}^{\mathrm{El}}_{k,n}(1, 1)$ and the dashed line to $\widehat{\lambda}^{\mathrm{Hu}}_{k,n}(1, 1)$. This simulation study also confirms the better performance of $\widehat{\lambda}^{\mathrm{El}}_{k,n}$.

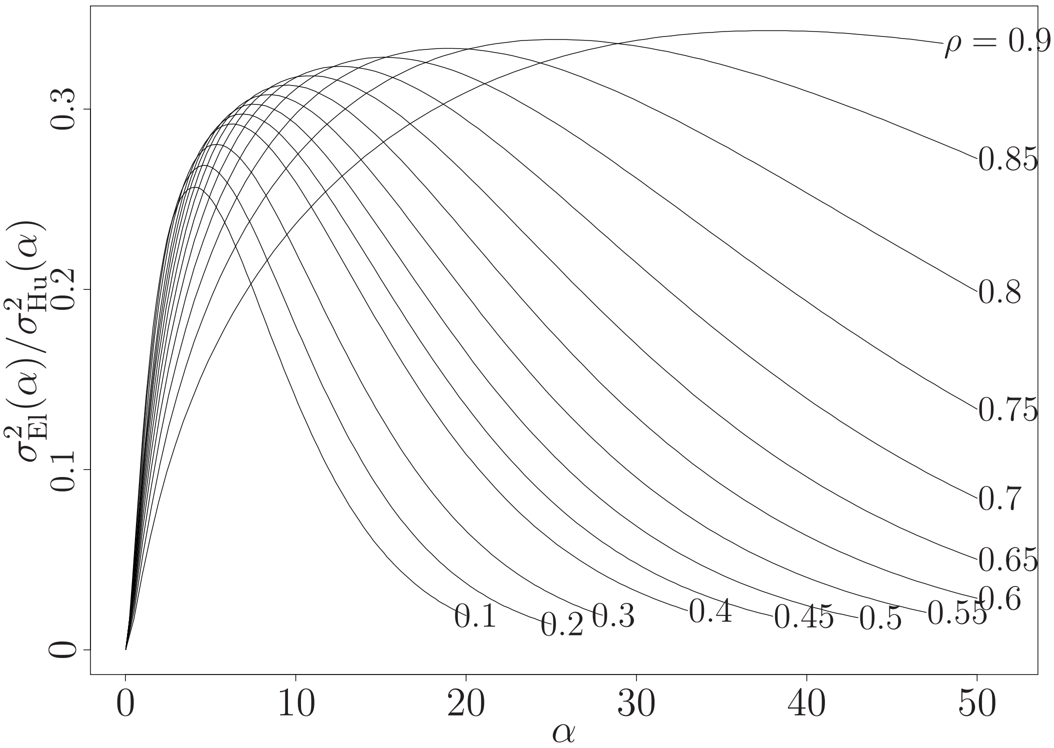

Figure 2.1: Ratio $\sigma^2_{\mathrm{El}}(\alpha)/\sigma^2_{\mathrm{Hu}}(\alpha)$ for different correlations ρ.

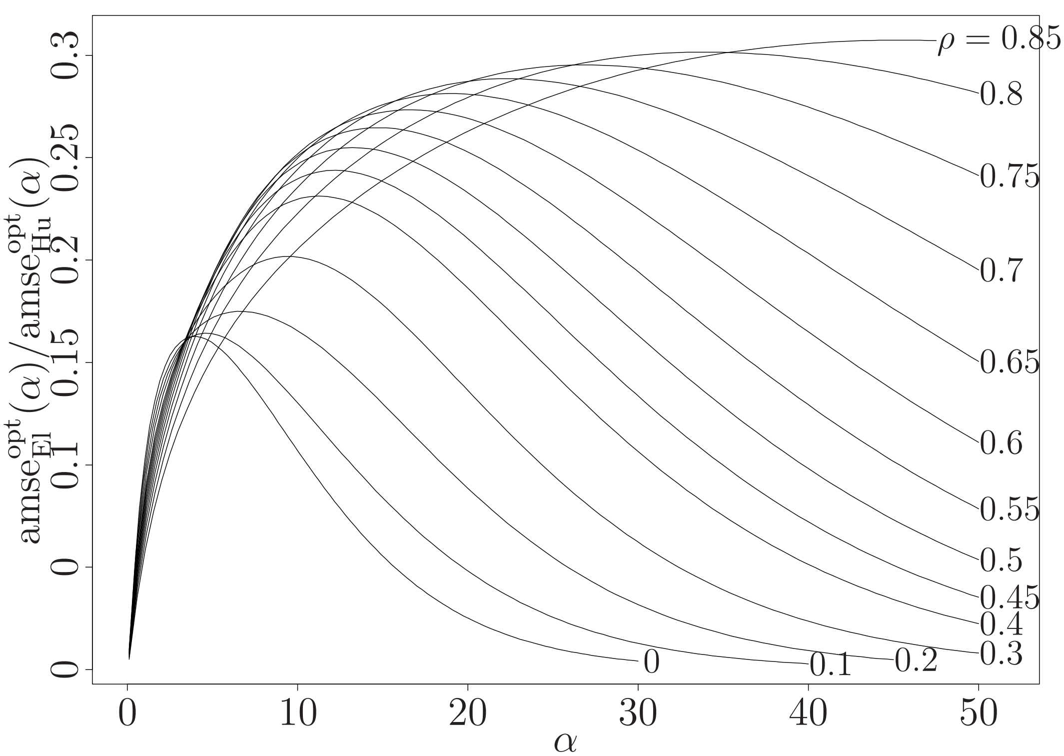

Figure 2.2: Ratio $\mathrm{amse}^{\mathrm{opt}}_{\mathrm{Hu}}(\alpha)/\mathrm{amse}^{\mathrm{opt}}_{\mathrm{El}}(\alpha)$ for different correlations ρ and $\beta = -\alpha$.

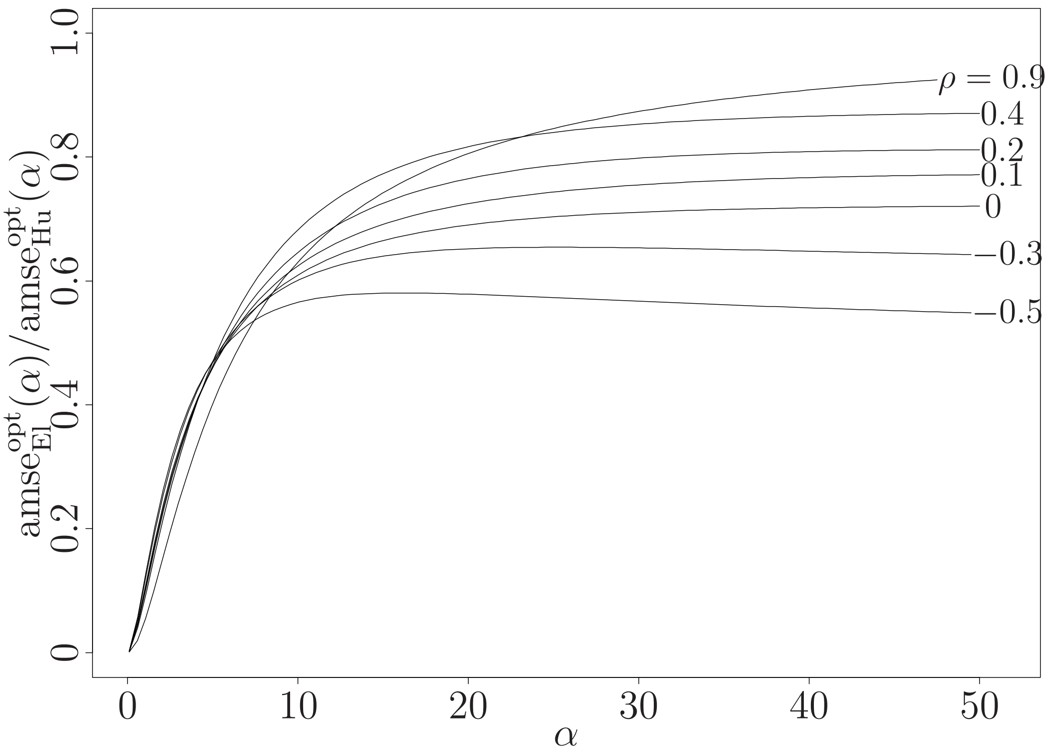

Figure 2.3: Ratio $\text{amse}^{\text{opt}}_{\text{Hu}}(\alpha)/\text{amse}^{\text{opt}}_{\text{El}}(\alpha)$ for different correlations ρ and $\beta = -1$.

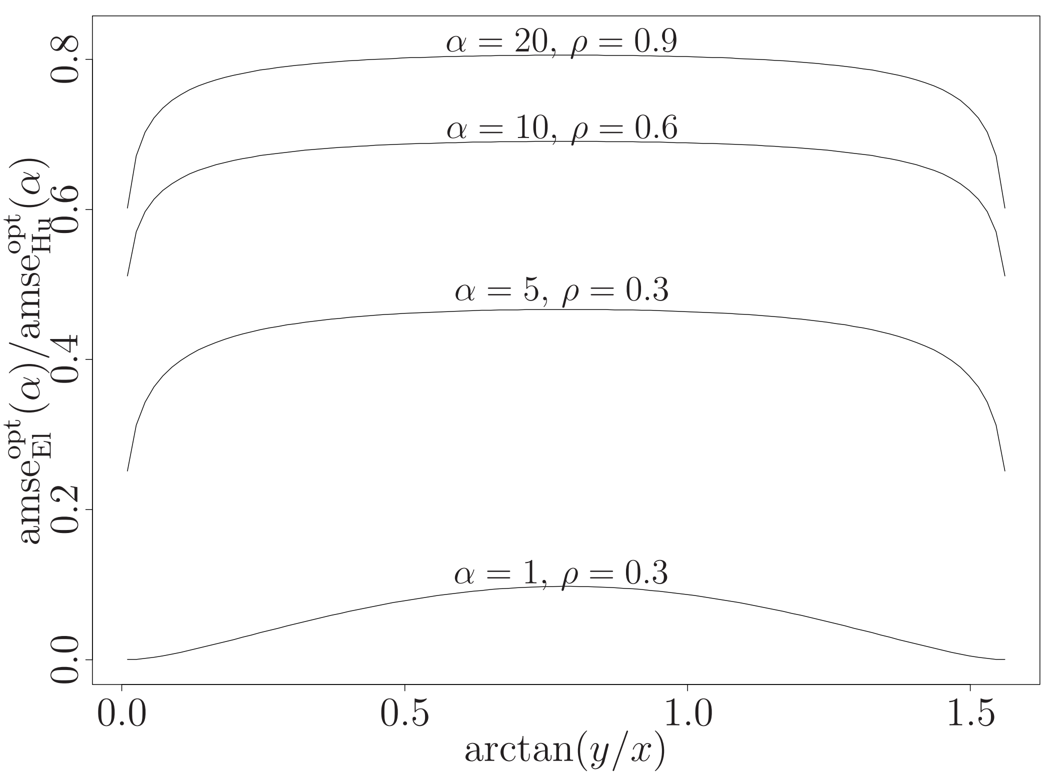

Figure 2.4: Ratio $\text{amse}^{\text{opt}}_{\text{El}}(\alpha)/\text{amse}^{\text{opt}}_{\text{Hu}}(\alpha)$, $x^2 + y^2 = 2$, for different $(\alpha,\, \rho)$ and $\beta = -1$.

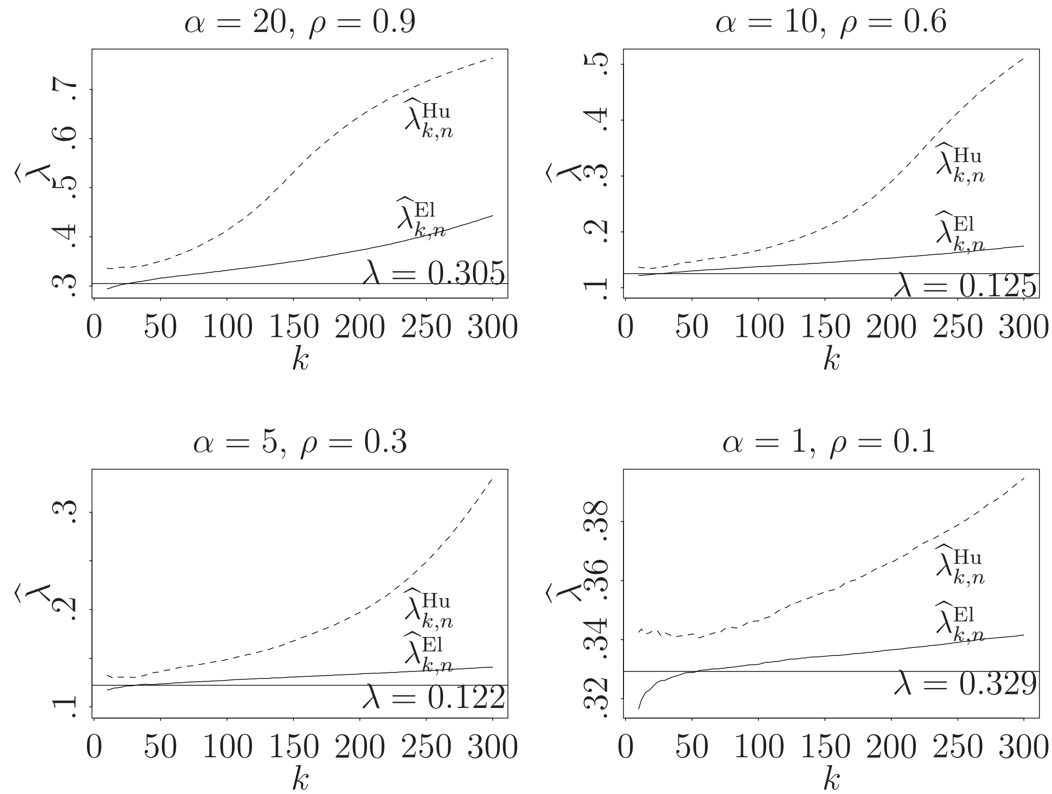

Figure 2.5: Mean of estimators $\widehat{\lambda}^{\mathrm{Hu}}_{k,n}(1,1)$ and $\widehat{\lambda}^{\mathrm{El}}_{k,n}(1,1)$ for 1 000 samples of length $n = 1\,000$ and different k with $\sigma = \upsilon = 1$, $\boldsymbol{\mu} = \mathbf{0}$, $G \sim$ Fréchet(α), and different (α, ρ).

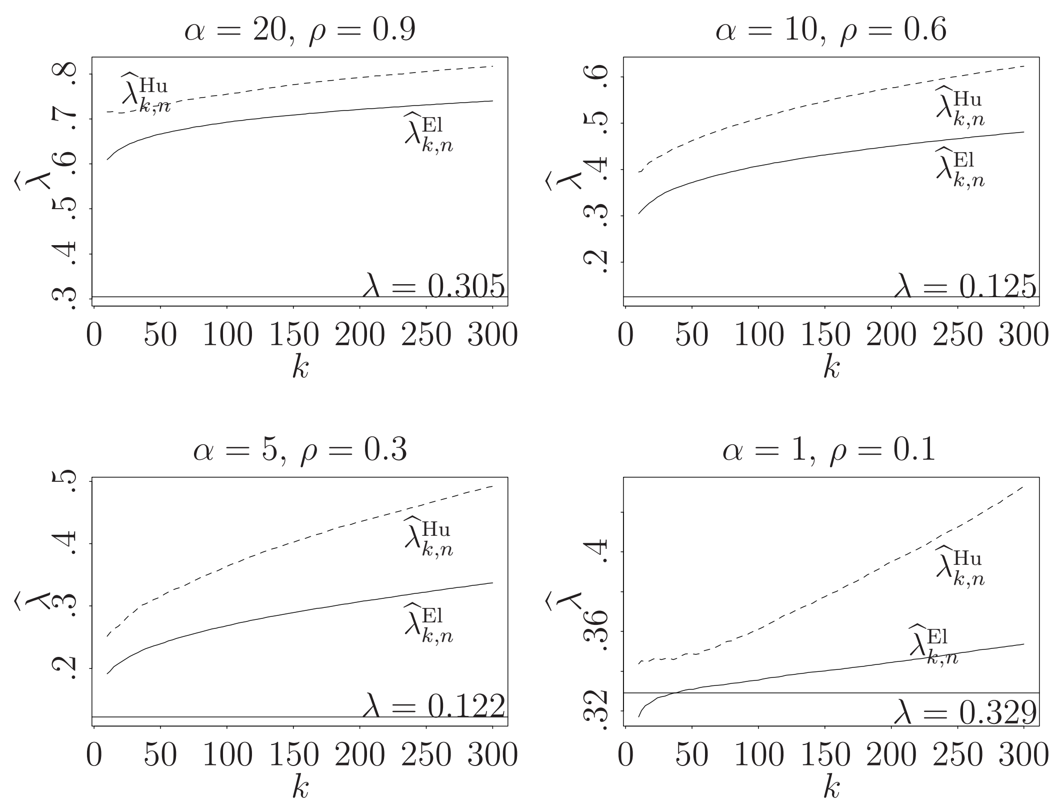

Figure 2.6: Mean of estimators $\widehat{\lambda}^{\mathrm{Hu}}_{k,n}(1,1)$ and $\widehat{\lambda}^{\mathrm{El}}_{k,n}(1,1)$ for 1 000 samples of length $n = 1\,000$ and different k with $\sigma = \upsilon = 1$, $\boldsymbol{\mu} = \mathbf{0}$, $G \sim$ Pareto(α) and different (α, ρ).

2.4 Proofs

Proof of Theorem 2.2.1: Without loss of generality, we assume $\boldsymbol{\mu} = \mathbf{0}$. Let $\Phi \sim \text{unif}(-\pi, \pi)$ be independent of G and $F_i^{\leftarrow}(x)$ denote the inverse of $F_i(x)$, $i = 1, 2$. Then, by Hult and Lindskog (2002),

$$\begin{aligned}
&F_X^{\leftarrow}(u) = \tfrac{\sigma}{\upsilon} F_Y^{\leftarrow}(u), && \text{for } 0 < u < 1, \\
&\lim_{t\to\infty}(1 - F_i(tx))/(1 - F_i(t)) = x^{-\alpha}, && \text{for } x > 0 \text{ and } i = 1, 2, \\
&(X, Y) \stackrel{\mathrm{d}}{=} (\sigma G \cos\Phi, \upsilon G \sin(\arcsin\rho + \Phi)).
\end{aligned} \tag{2.4.1}$$

Therefore,

$$\begin{aligned}
&t^{-1} P\left(F_X(X) \geq 1 - tx,\ F_Y(Y) \geq 1 - ty\right) \\
&= t^{-1} P\left(G\cos\Phi \geq F_Y^{\leftarrow}(1 - tx)/\upsilon, G\sin(\arcsin\rho + \Phi) \geq F_Y^{\leftarrow}(1 - ty)/\upsilon\right) \\
&= \frac{1}{2\pi t} \int_{-\arcsin\rho}^{\pi/2} P\left(G \geq \frac{F_Y^{\leftarrow}(1 - tx)}{\upsilon\cos\phi} \vee \frac{F_Y^{\leftarrow}(1 - ty)}{\upsilon\sin(\arcsin\rho + \phi)}\right) \mathrm{d}\phi.
\end{aligned} \tag{2.4.2}$$

Note that

$$\begin{aligned}
t &= P\left(X > F_X^{\leftarrow}(1 - t)\right) = P\left(G\cos\Phi > F_Y^{\leftarrow}(1 - t)/\upsilon\right) \\
&= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} P\left(G > \frac{F_Y^{\leftarrow}(1 - t)}{\upsilon\cos\phi}\right) \mathrm{d}\phi.
\end{aligned}$$

Further, $1 \geq P\left(G > x/\cos\phi\right)/P\left(G \geq x\right) \stackrel{x\to\infty}{\longrightarrow} (\cos\phi)^\alpha$. Hence, in the following formula we can apply the dominated convergence theorem and obtain

$$\begin{aligned}
\frac{1}{\mathcal{B}_{(2.4.3)}(t)} &:= \frac{1}{2\pi t} P\left(G \geq F_Y^{\leftarrow}(1 - t)/\upsilon\right) \\
&\stackrel{t\to 0}{\longrightarrow} \left(\int_{-\pi/2}^{\pi/2} (\cos\phi)^\alpha \,\mathrm{d}\phi\right)^{-1} =: \frac{1}{\mathcal{B}_{(2.4.3)}}.
\end{aligned} \tag{2.4.3}$$

Next, we obtain for $\phi \in (-\arcsin\rho, \pi/2)$

$$\frac{F_Y^{\leftarrow}(1 - tx)}{\upsilon\cos\phi} \geq \frac{F_Y^{\leftarrow}(1 - ty)}{\upsilon\sin(\arcsin\rho + \phi)} \quad \Leftrightarrow \quad \frac{F_Y^{\leftarrow}(1 - ty)}{F_Y^{\leftarrow}(1 - tx)} \leq \frac{\sin(\arcsin\rho + \phi)}{\cos\phi}.$$

Note that $\sin(\arcsin\rho+\phi)/\cos\phi$ is strictly increasing, hence its inverse exists and equals $\arctan\left((\cdot-\rho)/\sqrt{1-\rho^2}\right)$. Therefore,

$$\begin{aligned}
&\frac{F_Y^{\leftarrow}(1-tx)}{\upsilon\cos\phi} \geq \frac{F_Y^{\leftarrow}(1-ty)}{\upsilon\sin(\arcsin\rho+\phi)} \\
\Leftrightarrow\quad &\phi \geq \arctan\left(\frac{F_Y^{\leftarrow}(1-ty)/F_Y^{\leftarrow}(1-tx)-\rho}{\sqrt{1-\rho^2}}\right) \\
&=:\ g\left(\frac{F_Y^{\leftarrow}(1-ty)}{F_Y^{\leftarrow}(1-tx)}\right) =: h(x,y,t). \qquad (2.4.4)
\end{aligned}$$

Since $1-F_Y \in \mathcal{R}_{-\alpha}$, by Proposition 1.7(9) of Geluk and de Haan (1987) $F_Y^{\leftarrow}(1-tx)/F_Y^{\leftarrow}(1-t) \stackrel{t\to 0}{\longrightarrow} x^{-1/\alpha}$, i.e.,

$$h(x,y,t) \stackrel{t\to 0}{\longrightarrow} g\left((x/y)^{1/\alpha}\right). \qquad (2.4.5)$$

It follows from (2.4.2) and (2.4.4) that

$$\begin{aligned}
&t^{-1}P\left(F_X(X)\geq 1-tx,\ F_Y(Y)\geq 1-ty\right) \\
&= \frac{1}{\mathcal{B}_{(2.4.3)}(t)}\int_{h(x,y,t)}^{\pi/2}\frac{P\left(G\geq \frac{F_Y^{\leftarrow}(1-t)}{\upsilon\cos\phi}\frac{F_Y^{\leftarrow}(1-tx)}{F_Y^{\leftarrow}(1-t)}\right)}{P\left(G\geq F_Y^{\leftarrow}(1-t)/\upsilon\right)}\mathrm{d}\phi \\
&\quad + \frac{1}{\mathcal{B}_{(2.4.3)}(t)}\int_{-\arcsin\rho}^{h(x,y,t)}\frac{P\left(G\geq \frac{F_Y^{\leftarrow}(1-t)}{\upsilon\sin(\arcsin\rho+\phi)}\frac{F_Y^{\leftarrow}(1-ty)}{F_Y^{\leftarrow}(1-t)}\right)}{P\left(G\geq F_Y^{\leftarrow}(1-t)/\upsilon\right)}\mathrm{d}\phi.
\end{aligned} \qquad (2.4.6)$$

Hence, the theorem follows from (2.4.3), (2.4.5) and Potter's inequality, e.g. see (1.20) in Geluk and de Haan (1987). □

Proof of Theorem 2.2.2: Since

$$(X,Y) \quad \stackrel{\mathrm{d}}{=} \quad \left(\mu_X+\sigma G\cos\Phi,\ \mu_Y+\upsilon G\sin(\Phi+\arcsin\rho)\right),$$

we have $X^2+Y^2 \stackrel{\mathrm{d}}{=} G^2 d_1(\Phi)+2Gd_2(\Phi)+\mu_X^2+\mu_Y^2$. Define

$$d_3(x,\phi) \quad := \quad \frac{1}{d_1(\phi)}\left(-d_2(\phi)+\sqrt{d_2^2(\phi)-d_1(\phi)\left(\mu_X^2+\mu_Y^2-x^2\right)}\right).$$

Since $P(X^2+Y^2 \ge t) = P\left(G \ge d_3(t,\Phi)\right)$ holds for large t, we obtain

$$\begin{aligned}
&\frac{P\left(X^2+Y^2 \ge t^2x^2\right)}{P\left(X^2+Y^2 \ge t^2\right)} - x^{-\alpha} \\
= &\left(\int_{-\pi}^{\pi} \frac{P\left(G \ge d_3(tx,\phi)\right)}{P(G \ge t)} \mathrm{d}\phi\right)\left(\int_{-\pi}^{\pi} \frac{P\left(G \ge d_3(t,\phi)\right)}{P(G \ge t)} \mathrm{d}\phi\right)^{-1} - x^{-\alpha} \\
= &\left\{\int_{-\pi}^{\pi} \left[\frac{P\left(G \ge d_3(tx,\phi)\right)}{P(G \ge t)} - \left(\frac{1}{t}d_3(tx,\phi)\right)^{-\alpha}\right]\mathrm{d}\phi\right. \\
&+ \int_{-\pi}^{\pi} \left[-x^{-\alpha}\frac{P\left(G \ge d_3(t,\phi)\right)}{P(G \ge t)} + x^{-\alpha}\left(\frac{1}{t}d_3(t,\phi)\right)^{-\alpha}\right]\mathrm{d}\phi \\
&\left.+ \int_{-\pi}^{\pi}\left[\left(\frac{1}{t}d_3(tx,\phi)\right)^{-\alpha} - x^{-\alpha}\left(\frac{1}{t}d_3(t,\phi)\right)^{-\alpha}\right]\mathrm{d}\phi\right\} \times \\
&\times \left(\int_{-\pi}^{\pi}\frac{P\left(G \ge d_3(t,\phi)\right)}{P(G \ge t)}\mathrm{d}\phi\right)^{-1} \\
=: &\left(\mathcal{J}_1(t) + \mathcal{J}_2(t) + \mathcal{J}_3(t)\right)\left(\int_{-\pi}^{\pi} \frac{P\left(G \ge d_3(t,\phi)\right)}{P(G \ge t)}\,\mathrm{d}\phi\right)^{-1}.
\end{aligned}$$

Since $|\rho| < 1$, it is straightforward to check that

$$\begin{aligned}
&\lim_{t\to\infty} t^{-1}d_3(t,\phi) = (d_1(\phi))^{-1/2}, \\
&0 < \sup_{-\pi\le\phi\le\pi} d_1(\phi) < \infty, \quad \text{and} \\
&\sup_{-\pi\le\phi\le\pi} |d_2(\phi)| < \infty.
\end{aligned} \tag{2.4.7}$$

Hence, similarly to the proof of Theorem 2.2.1,

$$\lim_{t\to\infty}\int_{-\pi}^{\pi} \frac{P\left(G \ge d_3(t,\phi)\right)}{P(G \ge t)}\,\mathrm{d}\phi = \int_{-\pi}^{\pi} (d_1(\phi))^{\alpha/2}\,\mathrm{d}\phi. \tag{2.4.8}$$

Similar to the proof of Draisma, de Haan, Peng, and Pereira (1999, Lemma 5.2), for any $\varepsilon > 0$, there exists $t_0 > 0$ such that for all $t \ge t_0, d_3(tx,\phi) \ge t_0$

$$\begin{aligned}
&\left|\frac{\frac{P\left(G \ge d_3(tx,\phi)\right)}{P(G \ge t)} - \left(\frac{1}{t}d_3(tx,\phi)\right)^{-\alpha}}{A(t)} - \left(\frac{1}{t}d_3(tx,\phi)\right)^{-\alpha}\frac{\left(\frac{1}{t}d_3(tx,\phi)\right)^{\beta} - 1}{\beta}\right| \\
\le\ &\varepsilon\left(1 + \left(\frac{1}{t}d_3(tx,\phi)\right)^{-\alpha} + \left(\frac{1}{t}d_3(tx,\phi)\right)^{-\alpha+\beta}\exp\left\{\varepsilon\left|\ln\left(\frac{1}{t}d_3(tx,\phi)\right)\right|\right\}\right).
\end{aligned} \tag{2.4.9}$$

Using (2.4.7), for any fixed $x > 0$, we can choose t_0 large enough such that $d_3(tx, \phi) \geq t_0$ uniformly for $\phi \in [-\pi, \pi]$. That is, for any fixed $x > 0$, (2.4.9) holds uniformly for $\phi \in [-\pi, \pi]$. Therefore, by dominated convergence theorem and (2.4.8), for $x > 0$,

$$\lim_{t\to\infty} \frac{\mathcal{J}_1(t)}{A(t)} = \frac{1}{x^\alpha \beta} \int_{-\pi}^{\pi} \left(x^\beta \left(d_1(\phi)\right)^{(\alpha-\beta)/2} - \left(d_1(\phi)\right)^{\alpha/2} \right) \mathrm{d}\phi \quad \text{and} \quad (2.4.10)$$

$$\lim_{t\to\infty} \frac{\mathcal{J}_2(t)}{A(t)} = -\frac{1}{x^\alpha \beta} \int_{-\pi}^{\pi} \left(\left(d_1(\phi)\right)^{(\alpha-\beta)/2} - \left(d_1(\phi)\right)^{\alpha/2} \right) \mathrm{d}\phi. \qquad (2.4.11)$$

It follows from (2.4.7) and a Taylor expansion, for $x > 0$, that

$$\begin{aligned} \mathcal{J}_3(t) &= \frac{\alpha}{tx^\alpha} \left(x^{-1} - 1\right) \int_{-\pi}^{\pi} \left(d_1(\phi)\right)^{\alpha/2-1/2} d_2(\phi)\, \mathrm{d}\phi \; + \; o\left(t^{-2}\right) \\ &+ \frac{\alpha}{2t^2 x^\alpha} \left(x^{-2} - 1\right) \int_{-\pi}^{\pi} \left(d_1(\phi)\right)^{\alpha/2-1} \left[\alpha (d_2(\phi))^2 + d_1(\phi)(\mu_X^2 + \mu_Y^2)\right] \mathrm{d}\phi. \end{aligned} \qquad (2.4.12)$$

Note that $\sin(\phi + \arcsin\rho) = \sqrt{1-\rho^2}\sin\phi + \rho\cos\phi$. Then, splitting the integral into $[-\pi, -\pi/2)$, $[-\pi/2, 0)$, $[0, \pi/2)$, $[\pi/2, \pi]$ and using the symmetry of sin and cos, we obtain

$$\int_{-\pi}^{\pi} \left(d_1(\phi)\right)^{(\alpha-1)/2} d_2(\phi)\, \mathrm{d}\phi = 0. \qquad (2.4.13)$$

Hence (2.2.3) follows from (2.4.10), (2.4.11), (2.4.12) and (2.4.13). Note that

$$\lim_{t\to\infty} P\left(\sqrt{X^2+Y^2} > t\right) / P(G > t) = \int_{-\pi}^{\pi} \left(d_1(\phi)\right)^{\alpha/2} d\phi$$

and, since $Y \stackrel{\mathrm{d}}{=} \mu_Y + \upsilon G \sin\Phi$ with $\Phi \sim \mathrm{unif}(-\pi, \pi)$ holds,

$$\lim_{t\to\infty} P(Y > t)/P(G > t) = \upsilon^\alpha \int_0^\pi (\sin\phi)^\alpha\, d\phi.$$

Therefore, we have

$$V(t) \sim \inf\left\{ y : P(G > y) \leq t^{-1} \Big/ \int_{-\pi}^{\pi} \left(d_1(\phi)\right)^{\alpha/2} d\phi \right\} \quad \text{and}$$

$$F_Y^{\leftarrow}(1 - t^{-1}) \sim \inf\left\{ y : P(G > y) \leq t^{-1} \Big/ \left(\upsilon^\alpha \int_0^\pi (\sin\phi)^\alpha\, d\phi \right) \right\}.$$

Hence,

$$\lim_{t\to\infty} \frac{V(t)}{F_Y^{\leftarrow}(1 - t^{-1})} = \left(\frac{\int_{-\pi}^{\pi} \left(d_1(\phi)\right)^{\alpha/2} d\phi}{\upsilon^\alpha \int_0^\pi (\sin\phi)^\alpha\, d\phi} \right)^{1/\alpha},$$

i.e., since $t^2|A(t)| \stackrel{t\to\infty}{\longrightarrow} \infty$ for $-2 < \beta \le 0$,

$$\begin{aligned}
&\lim_{t\to\infty} \frac{(V(t))^{-2} + |A(V(t))|}{(F_Y^{\leftarrow}(1-t^{-1}))^{-2} + |A(F_Y^{\leftarrow}(1-t^{-1}))|} \\
&= \left(\frac{\int_{-\pi}^{\pi} (d_1(\phi))^{\alpha/2}\, d\phi}{\upsilon^\alpha \int_0^\pi (\sin\phi)^\alpha\, d\phi} \right)^{-(2\wedge|\beta|)/\alpha}. \qquad (2.4.14)
\end{aligned}$$

Note that, by Taylor expansion,

$$\left(\frac{V(tx)}{V(t)}\right)^{-\alpha} = \frac{1}{x} - \frac{\alpha}{x^{1+1/\alpha}} \left(\frac{V(tx)}{V(t)} - x^{1/\alpha} \right) + o\left(1/V(t) + |A(V(t))|\right). \qquad (2.4.15)$$

Therefore, replacing t and x in (2.2.3) by $V(t)$ and $V(tx)/V(t)$, respectively, and using (2.4.14) and (2.4.15), we obtain (2.2.4). Let $\mu_X = \mu_Y = 0$, then $\mathcal{J}_3(t) = 0$ and we obtain (2.2.5). $\square$

Proof of Theorem 2.2.3: In order to prove Theorem 2.2.3, we can assume $\mu_X = \mu_Y = 0$ since $\lambda(x,y)$ is independent of margins. We also set $\upsilon = 1$ and give the correction at the end of the proof. Using an upper-triangle decomposition of $\mathbf{\Sigma}$ yields $Y \stackrel{\mathrm{d}}{=} G \sin\Phi$, where $\Phi \sim \mathrm{unif}(-\pi,\pi)$ and is independent of G. Then, write

$$\begin{aligned}
\frac{P(Y \ge tx)}{P(Y \ge t)} - x^{-\alpha} &= \frac{\int_0^\pi P\left(G \ge tx/\sin\phi\right) \mathrm{d}\phi}{\int_0^\pi P\left(G \ge t/\sin\phi\right) \mathrm{d}\phi} - x^{-\alpha} \\
&= \left(\int_0^\pi \frac{P\left(G \ge t/\sin\phi\right)}{P(G \ge t)} \mathrm{d}\phi \right)^{-1} \left\{ \int_0^\pi \left[\frac{P\left(G \ge tx/\sin\phi\right)}{P(G\ge t)} - \left(\frac{x}{\sin\phi}\right)^{-\alpha} \right] \right. \\
&\qquad \left. - x^{-\alpha} \left[\frac{P\left(G \ge t/\sin\phi\right)}{P(G \ge t)} - \left(\frac{1}{\sin\phi}\right)^{-\alpha} \right] \mathrm{d}\phi \right\}.
\end{aligned}$$

Then, by (2.2.1), we have for $x > 0$

$$\begin{aligned}
&\lim_{t\to\infty} \left(\frac{P(Y \ge tx)}{P(Y \ge t)} - x^{-\alpha} \right) / A(t) \\
&= x^{-\alpha} \frac{x^\beta - 1}{\beta} \left(\int_0^\pi (\sin\phi)^\alpha\, \mathrm{d}\phi \right)^{-1} \left(\int_0^\pi (\sin\phi)^{\alpha-\beta}\, \mathrm{d}\phi \right).
\end{aligned}$$

Replacing t and x in the latter equation by $F_Y^{\leftarrow}(1-s)$ and $F_Y^{\leftarrow}(1-sy)/F_Y^{\leftarrow}(1-s)$, respectively, we obtain, for $y > 0$,

$$\lim_{s\to 0} \left(\left(\frac{F_Y^{\leftarrow}(1-sy)}{F_Y^{\leftarrow}(1-s)} \right)^{-\alpha} - y \right) / A\left(F_Y^{\leftarrow}(1-s)\right) = \mathcal{B}_{(2.2.6)}(y). \qquad (2.4.16)$$

Denote $f(t) := F_Y^{\leftarrow}(1-t)$. Then, by (2.4.6), we can write

$$
\begin{aligned}
& t^{-1} P\left(F_X(X) \geq 1-tx,\ F_Y(Y) \geq 1-ty\right) \\
= \quad & \frac{1}{\mathcal{B}_{(2.4.3)}(t)} \Bigg\{ \int_{h(x,y,t)}^{\pi/2} \Big[\frac{P\left(G \geq \frac{f(t)}{\cos\phi}\frac{f(tx)}{f(t)}\right)}{P\left(G \geq f(t)\right)} - \left(\frac{f(tx)}{f(t)\cos\phi}\right)^{-\alpha} \Big] \mathrm{d}\phi \\
& + \int_{h(x,y,t)}^{\pi/2} \Big[\left(\frac{f(tx)}{f(t)\cos\phi}\right)^{-\alpha} - x(\cos\phi)^{\alpha} \Big] \mathrm{d}\phi \ + \int_{h(x,y,t)}^{h(x,y,0)} x(\cos\phi)^{\alpha}\,\mathrm{d}\phi \\
& + \int_{-\arcsin\rho}^{h(x,y,t)} \Big[\frac{P\left(G \geq \frac{f(t)}{\sin(\arcsin\rho+\phi)}\frac{f(ty)}{f(t)}\right)}{P\left(G \geq f(t)\right)} \\
& \qquad\qquad - \left(\frac{f(ty)}{f(t)\sin\left(\arcsin\rho+\phi\right)}\right)^{-\alpha} \Big] \mathrm{d}\phi \\
& + \int_{-\arcsin\rho}^{h(x,y,t)} \Big[\left(\frac{f(ty)}{f(t)\sin\left(\arcsin\rho+\phi\right)}\right)^{-\alpha} - y\left(\sin\left(\arcsin\rho+\phi\right)\right)^{\alpha} \Big] \mathrm{d}\phi \\
& + \int_{h(x,y,0)}^{h(x,y,t)} y\left(\sin\left(\arcsin\rho+\phi\right)\right)^{\alpha}\,\mathrm{d}\phi \ + \int_{h(x,y,0)}^{\pi/2} x(\cos\phi)^{\alpha}\,\mathrm{d}\phi \\
& + \int_{-\arcsin\rho}^{h(x,y,0)} y\left(\sin\left(\arcsin\rho+\phi\right)\right)^{\alpha}\,\mathrm{d}\phi \Bigg\} \\
=: \quad & \frac{1}{\mathcal{B}_{(2.4.3)}(t)} \left(\sum_{i=1}^{6} \mathcal{J}_i(t) + \mathcal{J}_7 + \mathcal{J}_8 \right).
\end{aligned}
\tag{2.4.17}
$$

Note that $1/|\cos\phi| \geq 1$ and υ is given, using Potter's bound and similar arguments as in the proof of Draisma et al. (1999, Lemma 5.2), for any $\varepsilon > 0$, there exists some small $t_0 > 0$ such that for all $f(t) \geq f(t_0)$, $f(tx) \geq f(t_0)$ and $\phi \in [-\pi/2, \pi/2]$

$$
\begin{aligned}
& \Bigg| \frac{P\left(G \geq \frac{f(tx)}{\cos\phi}\right) / P\left(G \geq f(t)\right) - \left(\frac{f(tx)}{f(t)\cos\phi}\right)^{-\alpha}}{A(f(t))} \\
& \qquad\qquad - \left(\frac{f(tx)}{f(t)\cos\phi}\right)^{-\alpha} \frac{\left(\frac{f(tx)}{f(t)\cos\phi}\right)^{\beta} - 1}{\beta} \Bigg| \\
\leq \ & \varepsilon \left(1 + \left(\frac{f(tx)}{f(t)\cos\phi}\right)^{-\alpha} + \left(\frac{f(tx)}{f(t)\cos\phi}\right)^{-\alpha+\beta} \exp\left\{ \varepsilon \left| \ln \frac{f(tx)}{f(t)\cos\phi} \right| \right\} \right),
\end{aligned}
\tag{2.4.18}
$$

and for all $t \le t_0$ and $tx \le t_0$,

$$(1-\varepsilon)x^{-1/\alpha}\exp(-\varepsilon|\log x|) \;\le\; \frac{f(tx)}{f(t)} \;\le\; (1+\epsilon)x^{-1/\alpha}\exp(\varepsilon|\log x|). \tag{2.4.19}$$

Since $f(t) \ge t_0$ and $t \le t_0$ imply that $f(tx) \ge t_0$ and $tx \le t_0$ for all $0 \le x \le 1$, respectively, by (2.4.18), (2.4.19), (2.4.5) and dominated convergence, we have

$$\lim_{t\to 0}\frac{\mathcal{J}_1(t)}{A(f(t))} \;=\; \frac{x}{\beta}\int_{h(x,y,0)}^{\pi/2}\left[x^{-\beta/\alpha}(\cos\phi)^{\alpha-\beta}-(\cos\phi)^{\alpha}\right]\mathrm{d}\phi \tag{2.4.20}$$

holds for all $x, y \ge 0$ and uniformly on $\mathcal{S}_2^+$. Similarly,

$$\lim_{t\to 0}\frac{\mathcal{J}_4(t)}{A(f(t))} \;=\; \frac{y}{\beta}\int_{-\arcsin\rho}^{h(x,y,0)}\Big[y^{-\beta/\alpha}(\sin(\phi+\arcsin\rho))^{\alpha-\beta} -(\sin(\phi+\arcsin\rho))^{\alpha}\Big]\,\mathrm{d}\phi \tag{2.4.21}$$

holds for all $x, y \ge 0$ and uniformly on $\mathcal{S}_2^+$.

Using (2.4.16) and a way similar to the proof of Lemma 5.2 of Draisma et al. (1999), for any $\varepsilon > 0$, there exists $t_0 > 0$ such that for all $t \le t_0$ and $tx \le t_0$

$$\begin{aligned}&\left|\frac{(F_Y^{\leftarrow}(1-tx)/F_Y^{\leftarrow}(1-t))^{-\alpha}-x}{A(F_Y^{\leftarrow}(1-s))}-\mathcal{B}_{(2.2.6)}(x)\right|\\ &\quad\le\; \varepsilon\left(C_1+C_2x+C_3x^{1-\beta/\alpha}\exp(\varepsilon|\ln x|)\right),\end{aligned} \tag{2.4.22}$$

where the constants $C_1 > 0, C_2 > 0, C_3 > 0$ are independent of x and t. Hence, it follows from (2.4.5) and (2.4.22) that

$$\lim_{t\to 0}\frac{\mathcal{J}_2(t)}{A\left(F_Y^{\leftarrow}(1-t)\right)} \;=\; \mathcal{B}_{(2.2.6)}(x)\int_{h(x,y,0)}^{\pi/2}(\cos\phi)^{\alpha}\,d\phi \quad \text{and} \tag{2.4.23}$$

$$\lim_{t\to 0}\frac{\mathcal{J}_5(t)}{A\left(F_Y^{\leftarrow}(1-t)\right)} \;=\; \mathcal{B}_{(2.2.6)}(y)\int_{-\arcsin\rho}^{h(x,y,0)}(\sin(\phi+\arcsin\rho))^{\alpha}\,d\phi \tag{2.4.24}$$

holds for all $x, y \ge 0$ and uniformly on $\mathcal{S}_2^+$.

Note that

$$\begin{aligned}&\frac{1}{A(f(t))}\left(\left(\frac{f(ty)}{f(tx)}\right)^{-\alpha}-\frac{y}{x}\right)\\ &\quad=\; \frac{1}{A(f(t))}\left[\frac{1}{x}\left(\left(\frac{f(ty)}{f(t)}\right)^{-\alpha}-y\right)-\frac{1}{x}\left(\frac{f(ty)}{f(tx)}\right)^{-\alpha}\left(\left(\frac{f(tx)}{f(t)}\right)^{-\alpha}-x\right)\right]\\ &\quad\xrightarrow{t\to 0}\; \frac{1}{x}\mathcal{B}_{(2.2.6)}(y)-\frac{y}{x^2}\mathcal{B}_{(2.2.6)}(x).\end{aligned}$$

Similar to the proof of Draisma et al. (1999, Lemma 5.2), for any $\varepsilon > 0$, there exists $t_0 > 0$ such that for all $t \le t_0, tx \le t_0, ty \le t_0$

$$\begin{aligned}
&\left|\frac{1}{A(f(t))}\left(\left(\frac{f(ty)}{f(tx)}\right)^{-\alpha} - \frac{y}{x}\right) - \frac{1}{x}\mathcal{B}_{(2.2.6)}(y) + \frac{y}{x^2}\mathcal{B}_{(2.2.6)}(x)\right| \\
\le\ & \frac{1}{x}\varepsilon\left(C_1 + C_2 y + C_3 y^{1-\beta/\alpha} e^{\varepsilon|\log y|}\right) \\
&+\frac{1}{x}\left(\frac{y}{x}\right)\left(C_1 + C_2 x + C_3 x^{1-\beta/\alpha}\exp(\varepsilon|\log x|)\right) \\
&+\frac{1}{x}\left(\frac{y}{x}\right)\exp(\varepsilon|\log(y/x)|)\left(C_1 + C_2 x + C_3 x^{1-\beta/\alpha}\exp(\varepsilon|\log x|)\right), \quad (2.4.25)
\end{aligned}$$

where constants $C_1 > 0, C_2 > 0, C_3 > 0$ are independent of t, x, y. Using (2.4.25),

$$\begin{cases}
\limsup_{z\to 0} |g'(z^{-1/\alpha})z^{2/\alpha}| < \infty \\
\limsup_{z\to\infty} |g'(z^{-1/\alpha})| < \infty \\
\limsup_{z\to\infty} [\sin(g(z^{-1/\alpha}) + \arcsin\rho)]^\alpha z < \infty
\end{cases}$$

and applying a Taylor expansion to $g(z^{-1/\alpha})$, we can show that

$$\begin{aligned}
\lim_{t\to 0}\frac{\mathcal{J}_3(t)}{A(f(t))} &= \lim_{t\to 0}\frac{1}{A(f(t))}\int_{g(f(ty)/f(tx))}^{g\left((x/y)^{1/\alpha}\right)} x(\cos\phi)^\alpha \,\mathrm{d}\phi \\
&= \frac{x}{\alpha}\left[\cos\left(g\left((x/y)^{1/\alpha}\right)\right)\right]^\alpha g'\left((x/y)^{1/\alpha}\right)\left(\frac{\mathcal{B}_{(2.2.6)}(y)}{y} - \frac{\mathcal{B}_{(2.2.6)}(x)}{x}\right)\left(\frac{x}{y}\right)^{1/\alpha}
\end{aligned} \quad (2.4.26)$$

and

$$\begin{aligned}
\lim_{t\to 0}\frac{\mathcal{J}_6(t)}{A(f(t))} &= \lim_{t\to 0}\frac{1}{A(f(t))}\int_{g\left((x/y)^{1/\alpha}\right)}^{g(f(ty)/f(tx))} y(\sin(\phi + \arcsin\rho))^\alpha \,\mathrm{d}\phi \\
&= -\frac{y}{\alpha}\left[\sin\left(g\left((x/y)^{1/\alpha}\right) + \arcsin\rho\right)\right]^\alpha g'\left((x/y)^{1/\alpha}\right) \times \\
&\qquad \times \left(\frac{\mathcal{B}_{(2.2.6)}(y)}{y} - \frac{\mathcal{B}_{(2.2.6)}(x)}{x}\right)\left(\frac{x}{y}\right)^{1/\alpha}
\end{aligned} \quad (2.4.27)$$

holds for all $x, y \ge 0$ and uniformly on $\mathcal{S}_2^+$. Since

$$x\left[\cos\left(g\left((x/y)^{1/\alpha}\right)\right)\right]^\alpha = y\left[\sin\left(g\left((x/y)^{1/\alpha}\right) + \arcsin\rho\right)\right]^\alpha, \quad (2.4.28)$$

we obtain $\lim_{t\to 0}(\mathcal{J}_3(t) + \mathcal{J}_6(t))/A(f(t)) = 0$.

By Theorem 2.2.1, $\lambda(x,y) = (\mathcal{J}_7 + \mathcal{J}_8)/\mathcal{B}_{(2.4.3)}$, hence

$$\begin{aligned}
&\lim_{t\to 0} \frac{1}{A(f(t))} \left(\frac{1}{\mathcal{B}_{(2.4.3)}(t)} (\mathcal{J}_7 + \mathcal{J}_8) - \lambda(x,y) \right) \\
&= \lim_{t\to 0} \frac{1}{A(f(t))} \left(-\frac{\lambda(x,y)}{\mathcal{B}_{(2.4.3)}(t)} \left(\mathcal{B}_{(2.4.3)}(t) - \mathcal{B}_{(2.4.3)} \right) \right) \\
&= -\lambda(x,y) \frac{1}{\beta} \left(\int_{-\pi/2}^{\pi/2} (\cos\phi)^{\alpha} \left((\cos\phi)^{-\beta} - 1 \right) \mathrm{d}\phi \right) \left(\int_{-\pi/2}^{\pi/2} (\cos\phi)^{\alpha} \, \mathrm{d}\phi \right)^{-1},
\end{aligned} \tag{2.4.29}$$

which obviously holds uniformly on $\mathcal{S}_2^+$ since $\sup_{\mathcal{S}_2^+} \lambda(x,y) < \infty$. Note that

$$A\left(F_Y^{\leftarrow}(1-t)\right) / A\left(F_{\upsilon Y}^{\leftarrow}(1-t)\right) \xrightarrow{t\to 0} \upsilon^{-\beta}. \tag{2.4.30}$$

Hence the theorem follows from (2.4.20), (2.4.21), (2.4.23), (2.4.24), (2.4.26), (2.4.27), (2.4.29) and (2.4.30). □

Proof of Theorem 2.2.4: Similar to Huang (1992) or Einmahl, de Haan, and Li (2006), we have, as $n \to \infty$,

$$\sup_{0\le x,y\le T} \left| \sqrt{k_{\mathrm{Hu}}} \left\{ x + y - \widehat{\lambda}^{Hu}_{k_{\mathrm{Hu}},n}(x,y) - l(x,y) \right\} - \mathcal{K}_{\mathrm{Hu}} \mathcal{B}_{(2.2.7)}(x,y) - B(x,y) \right| = o_p(1),$$

where

$$B(x,y) = W(x,y) - \left(1 - \frac{\partial \lambda(x,y)}{\partial x} \right) W(x,0) - \left(1 - \frac{\partial \lambda(x,y)}{\partial y} \right) W(0,y),$$

and $W(x,y)$ is a Wiener process with zero mean and covariance structure

$$\begin{aligned}
E\left(W(x_1,y_1) W(x_2,y_2) \right) &= l(x_1 \wedge x_2, y_1) l(x_1 \wedge x_2, y_2) - l(x_1, y_1 \wedge y_2) \\
&\quad + l(x_2, y_1 \wedge y_2) - l(x_1, y_2) - l(x_2, y_1) - l(x_1 \wedge x_2, y_1 \wedge y_2).
\end{aligned}$$

Hence (2.2.9) follows from $\lambda(x,y) = x + y - l(x,y)$. It is straightforward to check that (2.2.10), (2.2.11) and (2.2.12) hold. Note that the result can also be obtained from Schmidt and Stadtmüller (2006) by taking the bias into account. □

Proof of Theorem 2.2.5: The result follows directly from

$$\sqrt{k_{\mathrm{El}}} \left(\widehat{\alpha}^{\mathrm{H}}_{k_{\mathrm{El}},n} - \alpha \right) \xrightarrow{d} \mathcal{N}\left(-\alpha^2 \mathcal{K}_{\mathrm{El}} \mathcal{B}_{(2.2.14)}, \alpha^2 \right)$$

(see de Haan and Peng (1998)), $\widehat{\tau}_n - \tau = o_p\left(k_{\mathrm{El}}^{-1/2} \right)$ and the delta method for the expression of $\lambda(x,y)$ given in Theorem 2.2.1. □

Chapter 3

Multivariate tail copula: modeling and estimation

Summary

In general, risk of an extreme outcome in financial markets can be expressed as a function of the tail copula of a high-dimensional vector after standardizing margins. Hence it is of importance to model and estimate tail copulas. Even for moderate dimension, nonparametrically estimating a tail copula is very inefficient and fitting a parametric model to tail copulas is not robust. In this chapter we propose a semi-parametric model for tail copulas via an elliptical copula. Based on this model assumption, we propose a novel estimator for the tail copula, which proves favorable compared to the empirical tail copula, both theoretically and empirically.

3.1 Introduction

Risk management is a discipline for living with the possibility that future events may cause adverse effects. An important issue for risk managers is how to quantify different types of risk such as market risk, credit risk, operational risk, etc. Due to the multivariate nature of risk, i.e., risk depending on high dimensional vectors of some underlying risk factors, a particular concern for a risk manager is how to model the dependence between extreme outcomes although those extreme outcomes occur rarely. A mathematical formulation of this question is as follows.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ be a random vector with distribution function F and continuous marginals $F_1, \ldots, F_d$. Then the dependence is completely determined by the *copula* C of $\boldsymbol{X}$ given by Sklar's representation (cf. Nelsen (1999) or Joe (1997))

$$F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d)), \quad \boldsymbol{x} = (x_1, \cdots, x_d)^T \in \mathbb{R}^d .$$

Moreover, the copula alone allows us to describe dependence on extreme outcomes. As C is a multivariate uniform distribution on $[0,1]^d$, extreme values are near the boundaries and extreme dependence happens around the points $(0, \ldots, 0)$ and $(1, \ldots, 1)$. This motivates the definition of the *tail copula* of $\boldsymbol{X}$ as

$$\lambda^{\boldsymbol{X}}(x_1, \ldots, x_d) \quad = \quad \lim_{t \to 0} t^{-1} P\left(1 - F_1(X_1) \le tx_1, \ldots, \; 1 - F_d(X_d) \le tx_d\right), \quad (3.1.1)$$

where $x_1, \ldots, x_d \ge 0$, if the limit exists. The bivariate case, when $d = 2$, has been thoroughly investigated and $\lambda^{\boldsymbol{X}}(1,1)$ is called the *upper tail dependence coefficient* of X_1 and X_2, see Joe (1997). It models dependence along the 45 degree line, where the bivariate dependence effects are mostly concentrated. For $x, y \in [0,1]^2$ the function $x + y - \lambda^{\boldsymbol{X}}(x,y)$ is called the *tail dependence function* of X_1 and X_2 by Huang (1992); such notions go back to Gumbel (1960), Pickands (1981) and Galambos (1987), and they represent the full dependence structure of the model.

The approach via a dependence function yields that the risk of an extreme outcome in financial markets can be expressed as a function of the tail copula $\lambda^{\boldsymbol{X}}(x_1, \ldots, x_d)$ after standardizing marginals. When $d = 2$, the tail copula $\lambda^{\boldsymbol{X}}(x,y)$ or the tail dependence function $x + y - \lambda^{\boldsymbol{X}}(x,y)$ can be estimated nonparametrically via bivariate extreme value theory; see Einmahl, de Haan, and Piterbarg (2001) and references therein. Also parametric models for the tail dependence function have been suggested and estimated, see Tawn (1988), Ledford and Tawn (1997) and Coles (2001) for examples and further references. The application of both, nonparametric and parametric estimation of tail dependence functions has almost only been investigated for the case $d = 2$ although theoretically both methods are applicable to the case $d > 2$. For an approach to nonparametric estimation of tail dependence in higher dimensions see Hsing, Klüppelberg, and Kuhn (2004). Recently, Heffernan and Tawn (2004) proposes a conditional approach to model multivariate extremes via investigating the limits of normalized conditional distributions. Obviously, nonparametric estimation severely suffers from the curse of dimensionality, when d becomes large, and fitting parametric models for large d is not robust in general.

In this chapter, we concentrate on the dependence structure only, which means we work in the tradition of estimating a dependence function. However, we neither work with purely nonparametric estimates nor do we specify a parametric model. Instead we propose to model the tail copula via an elliptical copula, a novel approach, which may be viewed as a semi-parametric approach. For the applications of copulas and elliptical copulae to risk management, we refer to Frey, McNeil, and Nyfeler (2001) and Embrechts, Lindskog, and McNeil (2003). Recently, Demarta and McNeil (2005) study some parameterized elliptical copulas. One of the advantages in employing elliptical copulae is simplicity of simulating multivariate extremes.

Recall that the random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^T$ has an elliptical distribution,

$$\boldsymbol{Z} \quad \stackrel{\mathrm{d}}{=} \quad G\mathbf{A}\boldsymbol{U}^{(d)}, \tag{3.1.2}$$

where $G > 0$ is a random variable, $\mathbf{A}$ is a deterministic $d \times d$ matrix with $\mathbf{A}\mathbf{A}^T := \boldsymbol{\Sigma} = (\sigma_{ij})_{1 \le i,j \le d}$ and $\operatorname{rank}(\boldsymbol{\Sigma}) = d$, $\boldsymbol{U}^{(d)}$ is a d-dimensional random vector uniformly distributed on the unit hyper-sphere $\mathcal{S}_d := \{\boldsymbol{z} \in \mathbb{R}^d : \boldsymbol{z}^T\boldsymbol{z} = 1\}$, and $\boldsymbol{U}^{(d)}$ is independent of G. Representation (3.1.2) implies that the elliptical distribution is uniquely defined by the matrix $\boldsymbol{\Sigma}$ and the random variable G. For a detailed description of elliptical distributions, we refer to Fang, Kotz, and Ng (1990). Then, an *elliptical copula* is defined as the copula of an elliptical distribution.

Define the *linear correlation* between Z_i and Z_j as $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ and denote by $\mathbf{R} := (\rho_{ij})_{1 \le i,j \le d}$ the correlation matrix. Note that ρ_{ij} exists for any elliptical distribution; if finite second moments exist it coincides with the usual correlation. Hult and Lindskog (2002) showed in their Theorem 4.3 under weak regularity conditions and $d = 2$ that regular variation of $P(G > \cdot)$ with index $\alpha > 0$ (notation: $P(G > \cdot) \in RV_{-\alpha}$) is equivalent to multivariate regular variation of $\boldsymbol{Z}$ with the same index α. We refer to Resnick (1987) for the definition and properties of multivariate regular variation. This implies, in particular, that the correlation matrix and the index α of regular variation are copula parameters.

Further, we denote the upper tail dependence coefficient between Z_i and Z_j as

$$\lambda_{ij}^{\boldsymbol{Z}}(1,1) \quad = \quad \left(\int_{(\pi/2 - \arcsin \rho_{ij})/2}^{\pi/2} (\cos\phi)^\alpha \, d\phi \right) \Big/ \left(\int_0^{\pi/2} (\cos\phi)^\alpha \, d\phi \right) \tag{3.1.3}$$

when $P(G > \cdot) \in RV_{-\alpha}$; in this case it is positive (cf. Hult and Lindskog (2002, Theorem 4.3)).

For illustration of our methodology, we focus on the case $d = 2$ from now on and the extension to $d > 2$ is given in Section 3.5. Klüppelberg, Kuhn, and Peng (2005a) studied two estimators for estimating the tail copula $\lambda^{\boldsymbol{X}}(x, y)$ as defined in (3.1.1), when observations have an elliptical distribution; i.e., $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{Z}$ with $\boldsymbol{Z}$ defined in (3.1.2) and $P(G > \cdot) \in RV_{-\alpha}$ for some $\alpha > 0$. One estimator is based on extreme value theory, another one on an extended version of (3.1.3); i.e., denoting $\lambda_{12}^{\boldsymbol{Z}} = \lambda^{\boldsymbol{Z}}$ and $\rho_{12} = \rho$,

$$\begin{aligned} \lambda^{\boldsymbol{Z}}(x,y) &= \left(\int_{g\left((x/y)^{1/\alpha}\right)}^{\pi/2} x(\cos\phi)^{\alpha}\, \mathrm{d}\phi + \int_{-\arcsin\rho}^{g\left((x/y)^{1/\alpha}\right)} y\left(\sin(\phi + \arcsin\rho)\right)^{\alpha}\, \mathrm{d}\phi \right) \\ &\quad \times \left(\int_{-\pi/2}^{\pi/2} (\cos\phi)^{\alpha}\, \mathrm{d}\phi \right)^{-1} := \lambda(\alpha; x, y, \rho), \end{aligned} \tag{3.1.4}$$

where $g(t) := \arctan\left((t-\rho)/\sqrt{1-\rho^2}\right) \in [-\arcsin\rho, \pi/2]$ for $t > 0$. Note that in this setup α can be estimated from observations.

Here we propose to model only the copula C (not the full distribution) of $\boldsymbol{X}$ by the copula of $\boldsymbol{Z}$ with $P(G > \cdot) \in RV_{-\alpha}$, i.e.,

$$P\left(F_1(X_1) \le x, F_2(X_2) \le y\right) = P\left(F_1^Z(Z_1) \le x, F_2^Z(Z_2) \le y\right), \tag{3.1.5}$$

where $F_1^{\boldsymbol{Z}}$ and $F_2^{\boldsymbol{Z}}$ denote the marginal distributions of $\boldsymbol{Z}$.

In our approach, the copula C is not completely determined, since we only work with the tail information (the regular variation) of G. Without doubt, how to test the above model assumptions is important, and will be investigated in a separate paper. In the present chapter, we focus on the estimation issue, i.e., seeking a way to improve the *empirical tail copula estimator*. For iid data $\boldsymbol{X}_i = (X_{i1}, X_{i2})$ for $i = 1, \ldots, n$, with unknown distribution function F and tail copula as in (3.1.1) the empirical tail copula estimator is defined as

$$\widehat{\lambda}^{\mathrm{emp}}(x,y;k) = \frac{1}{k}\sum_{i=1}^{n} I\left(1 - \widehat{F}_1(X_{i1}) \le \frac{k}{n}x,\ 1 - \widehat{F}_2(X_{i2}) \le \frac{k}{n}y\right), \tag{3.1.6}$$

where $\widehat{F}_j$ denotes the empirical distribution of $\{X_{ij}\}_{i=1}^{n}$ for $j = 1, 2$ and we consider $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$.

A natural way to improve the empirical tail copula estimator is to employ (3.1.4) like Klüppelberg et al. (2005a). However, α can not be estimated directly from the observations under the model assumptions. Hence, we propose to estimate α first by using (3.1.4) with the empirical tail copula and an estimator for ρ. Then we estimate the tail copula λ by plugging in the estimators for α and ρ; see section 3.2 for details. Some theoretical comparisons are provided in section 3.3. We present a simulation study in section 3.4. The generalization to higher dimension is discussed in section 3.5. Finally, all proofs are summarized in section 3.6.

3.2 Methodologies and Main Results

Throughout this section we assume that $d = 2$. Because of (3.1.5), we can estimate $\lambda^{\boldsymbol{Z}}(x, y)$ by $\widehat{\lambda}^{\text{emp}}(x, y; k)$. It follows from Lindskog, McNeil, and Schmock () that condition $|\rho| < 1$ implies $\tau = \frac{2}{\pi} \arcsin \rho$, where τ is *Kendall's tau*, i.e.

$$\tau \;=\; P\left(\left(X_{11} - X_{21}\right)\left(X_{12} - X_{22}\right) > 0\right) - P\left(\left(X_{11} - X_{21}\right)\left(X_{12} - X_{22}\right) < 0\right).$$

Hence we can estimate ρ by $\widehat{\rho} = \sin\left(\frac{\pi}{2}\widehat{\tau}\right)$, where

$$\widehat{\tau} \;=\; \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \operatorname{sign}\left((X_{i1} - X_{j1})(X_{i2} - X_{j2})\right). \tag{3.2.1}$$

In order to estimate α via (3.1.4), we need to solve this equation as a function of α.

Theorem 3.2.1. *For any fixed $x, y > 0$ and $|\rho| < 1$, define $\alpha^* := |\ln(x/y)/\ln(\rho \vee 0)|$. Then, $\lambda(\alpha; x, y, \rho)$ is strictly decreasing in α for all $\alpha > \alpha^*$.*

Based on the above theorem, we are able to define an estimator for α as follows. Let $\lambda^{\leftarrow}(\,\cdot\,; x, y, \rho)$ denote the inverse of $\lambda(\alpha; x, y, \rho)$ with respect to α, if it exists. By Theorem 3.2.1, we know that $\lambda^{\leftarrow}(\,\cdot\,; 1, 1, \rho)$ exists for all $\alpha > 0$. Hence, an obvious estimator for α is $\widetilde{\alpha}(1, 1, k) := \lambda^{\leftarrow}(\widehat{\lambda}^{\text{emp}}(1, 1; k); 1, 1, \widehat{\rho})$ for any estimator $\widehat{\rho}$ of ρ. Since this estimator only employs information at $x = y = 1$, it may not be efficient.

Next we extend the estimator $\widetilde{\alpha}(1, 1, k)$ to $\widetilde{\alpha}(x, y, k)$ for other values $(x, y) \in \mathbb{R}_+^2$. Based on Theorem 3.2.1 we define corresponding ranges for $y/x = \tan\theta$. To ensure that $(x, y) = (1, 1)$ is taken into account, we look at $(x, y) = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta)$ for different angles θ. Note that $\widehat{\lambda}^{\text{emp}}(x, y;\, k) = \widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta;\, k^*)$ for $\theta = \arctan(y/x)$ and some k^*, hence it is sufficient not to consider all $(x, y) \in \mathbb{R}_+^2$ but only $(x, y) = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta)$. Define

$$\begin{aligned}
\widehat{Q} \;&:=\; \left\{\theta \in \left(0, \frac{\pi}{2}\right) : \widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k) < \right. \\
&\qquad\qquad \left. < \lambda\left(\left|\frac{\ln(\tan\theta)}{\ln(\widehat{\rho} \vee 0)}\right|;\, \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\rho}\right)\right\}, \\
\widehat{Q}^* \;&:=\; \left\{\theta \in \left(0, \frac{\pi}{2}\right) : |\ln(\tan\theta)| < \widetilde{\alpha}(1, 1; k)\left(1 - k^{-1/4}\right)|\ln(\widehat{\rho} \vee 0)|\right\} \quad \text{and} \\
Q^* \;&:=\; \left\{\theta \in \left(0, \frac{\pi}{2}\right) : |\ln(\tan\theta)| < \alpha\,|\ln(\rho \vee 0)|\right\}.
\end{aligned}$$

It follows from Theorem 3.2.1 that there exists a unique $\alpha_1 > |\ln(\tan\theta)/\ln(\widehat{\rho} \vee 0)|$ such that

$$\lambda(\alpha_1; \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\rho}) = \widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k), \quad \theta \in \widehat{Q}.$$

Therefore, for $\theta \in \widehat{Q}$ we can define the inverse function of $\lambda(\,\cdot\,; \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\rho})$ giving

$$\widetilde{\alpha}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k) \;=\; \lambda^{\leftarrow}\left(\widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k); \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\rho}\right). \quad (3.2.2)$$

Next we have to ensure consistency of this estimator. This can be done by further requiring $\theta \in \widehat{Q}^{*}$, which implies that the true value of α is larger than $|\ln(\tan\theta)/\ln(\widehat{\rho} \vee 0)|$ with probability tending to one. Thus, our estimator for α is defined as a smoothed version of $\widetilde{\alpha}$. That is, for an arbitrary nonnegative weight function w we define

$$\widehat{\alpha}(k, w) = \frac{1}{W\left(\widehat{Q} \cap \widehat{Q}^{*}\right)} \int_{\theta \in \widehat{Q} \cap \widehat{Q}^{*}} \widetilde{\alpha}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k)\, W(d\theta) \quad (3.2.3)$$

where W is the measure defined by w.

Before we give the asymptotic normality of $\widehat{\alpha}$, we list the following regularity conditions:

(C1) X satisfies relation (3.1.5) and Z has tail dependence function (3.1.4) and $P(G > \cdot\,) \in RV_{-\alpha}$ for some $\alpha > 0$ and $|\rho| < 1$.

(C2) There exists $A(t) \to 0$ such that

$$\lim_{t \to 0} \frac{t^{-1}P(1 - F_1(X_1) \le tx, 1 - F_2(X_2) \le ty) - \lambda^{X}(x, y)}{A(t)} \;=\; b_{(C2)}(x, y)$$

uniformly on $\mathcal{S}_2$, where $b_{(C2)}(x, y)$ is not a multiple of $\lambda^{X}(x, y)$.

(C3) $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A(k/n) \to b_{(C3)} \in (-\infty, \infty)$ as $n \to \infty$.

The following theorem gives the asymptotic normality of $\widehat{\alpha}$.

Theorem 3.2.2. *Suppose that (C1)-(C3) hold, and that w is a positive weight funtion satisfying $\sup_{\theta \in Q^{*}} w(\theta) < \infty$. Then, denoting by W the measure defined by w, as $n \to \infty$,*

$$\begin{aligned}
&\sqrt{k}\left(\widehat{\alpha}(k, w) - \alpha\right) \\
&\xrightarrow{d} \frac{1}{W(Q^{*})} \int_{\theta \in Q^{*}} \frac{b_{(C3)} b_{(C2)}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta) + \widetilde{B}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta)}{\lambda'\left(\alpha; \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \rho\right)} W(d\theta),
\end{aligned}$$

where $\lambda'(\alpha; x, y, \rho) := \frac{\partial}{\partial\alpha}\lambda(\alpha; x, y, \rho)$,

$$\widetilde{B}(x, y) \;=\; B(x, y) - B(x, 0)\left(1 - \frac{\partial}{\partial x}\lambda(x, y)\right) - B(0, y)\left(1 - \frac{\partial}{\partial y}\lambda(x, y)\right)$$

and $B(x, y)$ is a Brownian motion with zero mean and covariance structure

$$\begin{aligned}
E\left(B(x_1, y_1)B(x_2, y_2)\right) \;=\;& x_1 \wedge x_2 + y_1 \wedge y_2 - \lambda(x_1 \wedge x_2, y_1) - \lambda(x_1 \wedge x_2, y_2) \\
&- \lambda(x_1, y_1 \wedge y_2) - \lambda(x_2, y_1 \wedge y_2) + \lambda(x_1, y_2) + \lambda(x_2, y_1) + \lambda(x_1 \wedge x_2, y_1 \wedge y_2).
\end{aligned}$$

Next, like in Klüppelberg et al. (2005a), we estimate $\widehat{\rho}$ via the identity $\tau = \frac{2}{\pi}\arcsin\rho$ and the estimator (3.2.1) and obtain an estimator for $\lambda(x,y)$ by

$$\widehat{\lambda}(x,y;k,w) \;=\; \lambda\left(\widehat{\alpha}(k,w);x,y,\widehat{\rho}\right). \tag{3.2.4}$$

We derive the asymptotic normality of this new estimator $\widehat{\lambda}(x,y;k,w)$ as follows.

Theorem 3.2.3. *Suppose that the conditions of Theorem 3.2.2 hold. Then, for $T>0$, we have as $n\to\infty$,*

$$\sup_{0\le x,y\le T}\Bigg|\sqrt{k}\left(\widehat{\lambda}(x,y;k,w)-\lambda^X(x,y)\right)-\lambda'(\alpha;x,y,\rho)\frac{1}{W(Q^*)}$$
$$\times\int_{\theta\in Q^*}\frac{b_{(C3)}b_{(C2)}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta)+\widetilde{B}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta,t)}{\lambda'(\alpha;\sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\rho)}W(d\theta)\Bigg| \;=\; o_p(1).$$

3.3 Theoretical Comparisons

The following corollary gives the optimal choice of the sample fraction k for $\widehat{\alpha}$ in terms of the *asymptotic mean squared error*. First, denote

$$\text{abias}_\alpha(w) \;=\; \frac{1}{W(Q^*)}\int_{\theta\in Q^*}\frac{b_{(C2)}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta)}{\lambda'(\alpha;\sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\rho)}W(d\theta)$$

and

$$\text{avar}_\alpha(w)=\frac{1}{(W(Q^*))^2}\times$$
$$\int_{\theta_1\in Q^*}\int_{\theta_2\in Q^*}\frac{E\left(\widetilde{B}(\sqrt{2}\cos\theta_1,\sqrt{2}\sin\theta_1)\widetilde{B}(\sqrt{2}\cos\theta_2,\sqrt{2}\sin\theta_2)\right)}{\lambda'(\alpha;\sqrt{2}\cos\theta_1,\sqrt{2}\sin\theta_1,\rho)\lambda'(\alpha;\sqrt{2}\cos\theta_2,\sqrt{2}\sin\theta_2,\rho)}W(d\theta_2)W(d\theta_1).$$

Corollary 3.3.1. *Assume that (C1)-(C3) hold and $A(t)\sim ct^\beta$ as $t\to 0$ for some $c\neq 0$ and $\beta>0$. Then the asymptotic mean squared error of $\widehat{\alpha}(k,w)$ is*

$$\text{amse}_\alpha(k,w) \;=\; c^2(k/n)^{2\beta}\left(\text{abias}_\alpha(w)\right)^2+k^{-1}\text{avar}_\alpha(w).$$

By minimizing the above asymptotic mean squared error, we obtain the optimal choice of k as

$$k_0(w) \;=\; \left(\frac{\text{avar}_\alpha(w)}{2\beta c^2(\text{abias}_\alpha(w))^2}\right)^{1/(2\beta+1)} n^{2\beta/(2\beta+1)}.$$

Hence the optimal asymptotic mean squared error of $\widehat{\alpha}$ is

$$\text{amse}_\alpha(k_0(w),w) \;=\; \left(\left(\frac{\text{avar}_\alpha(w)}{n}\right)^\beta \text{abias}_\alpha(w)c\sqrt{2\beta}\right)^{2/(2\beta+2)}\left(1+\frac{1}{2\beta}\right).$$

Firstly, we compare $\widehat{\alpha}(k,w)$ with $\widetilde{\alpha}(1,1;k)$. As a first weight function we choose $w_0(\theta)$ equal to one if $\theta = \pi/4$, and equal to zero otherwise. Since $\widetilde{\alpha}(1,1;k) = \widehat{\alpha}(k,w_0)$, the asymptotic variance and optimal asymptotic mean squared error of $\widetilde{\alpha}(1,1;k)$ are

$$(k_0(w_0))^{-1}\,\mathrm{avar}_\alpha(w_0) \quad \text{and} \quad \mathrm{amse}_\alpha(k_0(w_0),w_0),$$

respectively. For simplicity, we only compare $\widehat{\alpha}(k,w_0)$ and $\widehat{\alpha}(k,w_1)$ with the weight function

$$w_1(\theta) \;=\; 1-\left(\frac{\theta}{\pi/4}-1\right)^2, \quad 0\le\theta\le\frac{\pi}{2}. \tag{3.3.1}$$

In Figure 3.1, we plot the ratio $\mathrm{ratio}_{\mathrm{var},\alpha} = \mathrm{avar}_\alpha(w_1)/\mathrm{avar}_\alpha(w_0)$ against α for $\rho\in\{0.3,0.7\}$, which shows that $\widehat{\alpha}(k,w_1)$ has a smaller variance than $\widetilde{\alpha}(1,1;k)$ in many cases, especially when α is large or ρ is small. Hence $\widehat{\alpha}(k,w_1)$ is better than $\widetilde{\alpha}(1,1;k)$ in terms of asymptotic variance. Without doubt, the weight function w_1 is not an optimal one. Seeking an optimal weight function is important, but difficult.

Secondly, we compare $\widehat{\lambda}(x,y;k,w)$ with $\widehat{\lambda}^{\mathrm{emp}}(x,y;k)$. It follows from Theorem 3.2.3 that the asymptotic variance and the asymptotic mean squared error of $\widehat{\lambda}(x,y;k,w)$ are

$$(\lambda'(\alpha;x,y,\rho))^2\,\mathrm{avar}_\alpha(k,w) \quad \text{and} \quad (\lambda'(\alpha;x,y,\rho))^2\,\mathrm{amse}_\alpha(k,w),$$

respectively. As before, we obtain the optimal asymptotic mean squared error of $\widehat{\lambda}(x,y;k,w)$ as $(\lambda'(\alpha;x,y,\rho))^2\,\mathrm{amse}_\alpha(k_0(w),w)$. Put

$$\begin{aligned} k_{\mathrm{emp}} &= \left(\frac{E(B^2(x,y))}{2\beta c^2(b_{(C2)}(x,y))^2}\right)^{1/(2\beta+1)} n^{2\beta/(2\beta+1)} \quad \text{and} \\ \mathrm{amse}_{\mathrm{emp}}(k) &= c^2(k/n)^{2\beta}(b_{(C2)}(x,y))^2 + k^{-1}E(B^2(x,y)). \end{aligned}$$

Then the asymptotic variance and the optimal asymptotic mean squared error of $\widehat{\lambda}^{\mathrm{emp}}(x,y;k)$ are

$$\mathrm{avar}_{\lambda^{\mathrm{emp}}}(k,w) = k^{-1}(E\widetilde{B}(x,y))^2 \quad \text{and} \quad \mathrm{amse}_{\lambda^{\mathrm{emp}}}(k,w) = \mathrm{amse}_{\mathrm{emp}}(k_{\mathrm{emp}})\,.$$

In Figure 3.2, we plot the ratio of the variances of $\widehat{\lambda}(x,y;w_1)$ and $\widehat{\lambda}^{\mathrm{emp}}(x,y;k)$ given by

$$\mathrm{ratio}_{\mathrm{var},\lambda} \;=\; \frac{E(B^2(x,y))}{(\lambda'(\alpha;x,y,\rho))^2\,\mathrm{avar}_\alpha(w_1)},$$

for $(x,y) = (\sqrt{2}\cos\phi, \sqrt{2}\sin\phi)$ against $\phi\in(0,\pi/2)$ for different pairs $(\alpha,\rho)\in\{1,5\}\times\{0.3,0.7\}$, which shows that the new estimator for $\lambda^X(x,y)$ has a smaller variance than the empirical estimator $\widehat{\lambda}^{\mathrm{emp}}(x,y;k)$.

3.4 Simulation Study

In this section we conduct a simulation study to compare $\widehat{\alpha}(k, w_1)$ with $\widehat{\alpha}(k, w_0) = \widetilde{\alpha}(1, 1, k)$, and to compare $\widehat{\lambda}(x, y; k, w_1)$ with $\widehat{\lambda}^{\text{emp}}(x, y; k)$ by drawing 1000 random samples with sample size $n = 3000$ from an elliptical copula with $P(G > x) = \exp\{-x^{-\alpha}\}$, $x > 0$.

For comparison of $\widehat{\alpha}(k, w_1)$ and $\widetilde{\alpha}(1, 1, k)$, we plot the averages of $\widetilde{\alpha}(1, 1, k)$, $\widehat{\alpha}(k, w_1)$ and corresponding mean squared errors in Figures 3.3 and 3.4. We observe that $\widehat{\alpha}(k, w_1)$ has a smaller mean squared error than $\widetilde{\alpha}(1, 1, k)$ in most cases. Further, we plot $\widetilde{\alpha}(1, 1, k)$ and $\widehat{\alpha}(k, w_1)$ based on a particular sample in Figure 3.7, which shows that $\widehat{\alpha}(k, w_1)$ is much smoother than $\widetilde{\alpha}(1, 1, k)$ with respect to k. This is because $\widehat{\alpha}(k, w_1)$ employs more $\widehat{\lambda}^{\text{emp}}(x, y; k)'s$ and $\widetilde{\alpha}(1, 1, k)$ only uses $\widehat{\lambda}^{\text{emp}}(1, 1; k)$. In summary, one may prefer $\widehat{\alpha}(k, w_1)$ to $\widetilde{\alpha}(1, 1, k)$.

Next we compare the empirical estimator $\widehat{\lambda}^{\text{emp}}(x, y; k)$ with the new $\widehat{\lambda}(x, y; k, w_1)$. We plot the averages of $\widehat{\lambda}^{\text{emp}}(1, 1; k)$, $\widehat{\lambda}(1, 1, k, w_1)$ and corresponding mean squared errors in Figures 3.5 and 3.6. We also plot estimators $\widehat{\lambda}^{\text{emp}}(1, 1; k)$ and $\widehat{\lambda}(1, 1; k, w_1)$ based on a particular sample in Figure 3.8. Like the comparisons for estimators of α, we observe that $\widehat{\lambda}(1, 1; k, w_1)$ has a slightly smaller mean squared error than $\widehat{\lambda}^{\text{emp}}(1, 1; k)$, but $\widehat{\lambda}(1, 1; k, w_1)$ is much smoother than $\widehat{\lambda}^{\text{emp}}(1, 1; k)$ with respect to k. More improvement of $\widehat{\lambda}(x, y; k, w_1)$ over $\widehat{\lambda}^{\text{emp}}(x, y; k, w_0)$ are found when x/y is away from one; see Figures 3.9 and 3.10.

Finally, we compare $\widehat{\lambda}(x, y; 50, w_1)$ and $\widehat{\lambda}^{\text{emp}}(x, y; 50, w_0)$ for different x and y. It follows from Figure 3.5 that $k = 50$ is a reasonable choice. Again, we plot the averages of $\widehat{\lambda}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50, w_1)$, $\widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50)$ for $0 \le \phi \le \pi/2$ and corresponding mean squared errors in Figures 3.11 and 3.12. Based on a particular sample, we also plot estimators $\widehat{\lambda}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50, w_1)$ and $\widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50)$ in Figure 3.13. From these figures, we observe that, when ϕ is away from $\pi/4$, $\widehat{\lambda}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50, w_1)$ becomes much better than $\widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50)$.

In conclusion, with the help of an elliptical copula, we are able to estimate the tail dependence function more efficiently.

3.5 Elliptical Copula of Arbitrary Dimension

In this section we generalize our results in section 2 to the case, where the dimension $d \ge 2$ is arbitrary.

Theorem 3.5.1. *Assume that $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ has the same copula as the elliptical vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^T$, whose distribution is given in (3.1.2). W.l.o.g. assume that $\mathbf{A}\mathbf{A}^T = R$ is the correlation matrix of $\boldsymbol{Z}$. Let $\mathbf{A}_{i\cdot}$ denote the i-th row of $\mathbf{A}$ and and let $F_{\boldsymbol{U}^{(d)}}$ denote the uniform distribution on $\mathcal{S}_d$. Then the tail copula of $\boldsymbol{X}$ is given by*

$$\begin{aligned}\lambda^{\boldsymbol{X}}(x_1, \ldots, x_d) &:= \lim_{t\to 0} t^{-1} P\left(1 - F_1(X_1) < tx_1, \ldots, 1 - F_d(X_d) < tx_d\right) \\ &= \int\limits_{\boldsymbol{u}\in\mathcal{S}_d, \mathbf{A}_{1\cdot}\boldsymbol{u}>0,\ldots,\mathbf{A}_{d\cdot}\boldsymbol{u}>0} \bigwedge_{i=1}^{d} x_i (\mathbf{A}_{i\cdot}\boldsymbol{u})^\alpha \, dF_{\boldsymbol{U}^{(d)}}(\boldsymbol{u}) \Bigg(\int\limits_{\boldsymbol{u}\in\mathcal{S}_d, \mathbf{A}_{1\cdot} u>0} (\mathbf{A}_{1\cdot}\boldsymbol{u})^\alpha \, dF_{\boldsymbol{U}^{(d)}}(\boldsymbol{u}) \Bigg)^{-1}. \end{aligned} \tag{3.5.1}$$

Remark 3.5.2. (a) For $d = 2$ representation (3.5.1) coincides with (3.1.4). To see this write $\boldsymbol{u} \in \mathcal{S}_2$ as $\boldsymbol{u} = (\cos\phi, \sin\phi)^T$ for some $\phi \in (-\pi, \pi)$, $\mathbf{A}_{1\cdot} = (1, 0)$ and $\mathbf{A}_{2\cdot} = (\rho, \sqrt{1-\rho^2})$. Then, $\mathbf{A}\boldsymbol{u} = (\cos\phi, \rho\cos\phi + \sqrt{1-\rho^2}\sin\phi)^T = (\cos\phi, \sin(\phi + \arcsin\rho))^T$, giving the equivalence of (3.5.1) and (3.1.4).
(b) For $d \geq 3$ one can also use multivariate polar coordinates and obtain analogous representations. The expression, however, becomes much more complicated.

The estimation procedure in d dimensions is a simple extension of the two-dimensional case. Assume iid observations $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})^T$, $i = 1, \ldots, n$, with an elliptical copula. Then we can estimate ρ_{pq} via Kendall's τ and α_{pq} based on bivariate subvectors (X_{ip}, X_{iq}) for $1 \leq p, q \leq d$. Denote these estimators by $\widehat{\rho}_{pq}$ and (for any positive weight function w) $\widehat{\alpha}_{pq}(k, w)$, respectively. Then we estimate α and $\mathbf{R}$ by

$$\widehat{\alpha}(k, w) = \frac{1}{d(d-1)} \sum_{p\neq q} \widehat{\alpha}_{pq}(k, w) \quad \text{and} \quad \widehat{\mathbf{R}} = (\widehat{\rho}_{pq})_{1\leq p,q\leq d}.$$

For any decomposition $\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T = \widehat{\mathbf{R}}$, we obtain an estimator for $\mathbf{A}$. This yields an estimator for $\lambda(x_1, \ldots, x_d)$ by replacing α and $\mathbf{A}_{i\cdot}$ in (3.5.1) by $\widehat{\alpha}(k, w)$ and $\widehat{\mathbf{A}}_{i\cdot}$, respectively. The asymptotic normality of this new estimator can be derived similarly as in Theorems 3.2.2 and 3.2.3.

In Figure 3.14 we give a three-dimensional example. We simulate a sample of length $n = 3\,000$ from an elliptical copula with $P(G > x) = \exp\{-x^{-\alpha}\}$, $x > 0$, and parameters $\rho_{12} = 0.3$, $\rho_{13} = 0.5$, $\rho_{23} = 0.7$ and $\alpha = 5$. In the upper row we plot the true tail copula $\lambda^{\boldsymbol{X}}\left(\sqrt{3}\cos\phi_1, \sqrt{3}\sin\phi_1\cos\phi_2, \sqrt{3}\sin\phi_1\sin\phi_2\right)$, $\phi_1, \phi_2 \in (0, \pi/2)$, and each column corresponds to perspective, contour and grey-scale image plot of $\lambda^{\boldsymbol{X}}$, respectively. In the middle and lower row, we plot the corresponding estimators $\widehat{\lambda}(\ldots; 100, w_1)$ and $\widehat{\lambda}^{\text{emp}}(\ldots; 100)$, respectively. From this figure, we also observe that $\widehat{\lambda}$ becomes much better than $\widehat{\lambda}^{\text{emp}}$ in the three-dimensional case.

Next we apply our estimators to a three-dimensional real data set which consists of $n = 4\,903$ daily log returns of currency exchange rates of GBP, USD and CHF with respect to EURO between May 1985 and June 2004. As in Figure 3.14, we plot the perspective, contour and grey-scale image of $\widehat{\lambda}\left(\sqrt{3}\cos\phi_1, \sqrt{3}\sin\phi_1\cos\phi_2, \sqrt{3}\sin\phi_1\sin\phi_2;\ k, w_1\right)$ and $\widehat{\lambda}^{\text{emp}}(\ldots; k)$; see Figures 3.15, 3.16 and 3.17 for $k = 100$, $k = 150$ and $k = 200$, respectively. Comparing the contour plots (middle columns) of $\widehat{\lambda}$ and $\widehat{\lambda}^{\text{emp}}$, one may conclude that the assumption of having an elliptical tail copula ist not restrictive.

3.6 Proofs

Proof of Theorem 3.2.1. Define

$$c_0 = \int_{-\pi/2}^{\pi/2} (\cos\phi)^\alpha \, d\phi, \qquad c_1 = \int_{-\pi/2}^{\pi/2} (\cos\phi)^\alpha \ln(\cos\phi)\, d\phi,$$

$$\begin{aligned} D(\alpha, z) &= c_0 \int_z^{\pi/2} (\cos\phi)^\alpha \ln(\cos\phi)\, d\phi - c_1 \int_z^{\pi/2} (\cos\phi)^\alpha \, d\phi \quad \text{and} \\ C(\alpha, z) &= D(\alpha, z) + \left(\rho + \sqrt{1-\rho^2}\tan z\right)^{-\alpha} D(\alpha, -z + \arccos\rho). \end{aligned}$$

Then, by variable transformation, we obtain

$$\lambda(\alpha; x, y, \rho) = c_0^{-1}\left(x \int_{g\left((x/y)^{1/\alpha}\right)}^{\pi/2} (\cos\phi)^\alpha \, \mathrm{d}\phi + y \int_{g\left((x/y)^{-1/\alpha}\right)}^{\pi/2} (\cos\phi)^\alpha \, \mathrm{d}\phi \right)$$

and

$$\begin{aligned} \lambda'(\alpha; x, y, \rho) &:= \frac{\partial}{\partial\alpha}\lambda(\alpha; x, y, \rho) = c_0^{-2}\left[x D\left(\alpha, g\left((x/y)^{1/\alpha}\right)\right) + y D\left(\alpha, g\left((x/y)^{-1/\alpha}\right)\right)\right] \\ &= c_0^{-2} x C\left(\alpha, g\left((x/y)^{1/\alpha}\right)\right). \end{aligned}$$

Since $D_{0,1}(\alpha, z) := \frac{\partial}{\partial z} D(\alpha, z) = (\cos z)^\alpha \left(c_1 - c_0 \ln(\cos z)\right)$, we can show that there exists $0 < z_0 < \pi/2$ such that

$$\begin{cases} D_{0,1}(\alpha, z) > 0, & \text{if } z \in (-\pi/2, -z_0), \\ D_{0,1}(\alpha, z) = 0, & \text{if } z = -z_0, \\ D_{0,1}(\alpha, z) < 0, & \text{if } z \in (-z_0, z_0), \\ D_{0,1}(\alpha, z) = 0, & \text{if } z = z_0, \\ D_{0,1}(\alpha, z) > 0, & \text{if } z \in (z_0, \pi/2). \end{cases}$$

Note that z_0 depends on α. Since $D(\alpha, 0) = \lim_{z \to \pm\pi/2} D(\alpha, z) = 0$, we have

$$\begin{cases} D(\alpha, z) > 0, & \text{if } z \in (-\pi/2, 0), \\ D(\alpha, z) < 0, & \text{if } z \in (0, \pi/2). \end{cases}$$

Hence, if $x/y \in \left[(\rho \vee 0)^{\alpha^*}, (\rho \vee 0)^{-\alpha^*}\right]$ for some $\alpha^* \in (0, \infty)$, then $C\left(\alpha, g\left((x/y)^{1/\alpha^*}\right)\right) < 0$ for all $\alpha > \alpha^*$. Since also $x/y \in [(\rho \vee 0)^{\alpha}, (\rho \vee 0)^{-\alpha}]$ holds for all $\alpha > \alpha^*$, we have $C\left(\alpha, g\left((x/y)^{1/\alpha}\right)\right) < 0$ for all $\alpha > \alpha^*$. Hence the theorem follows by choosing $\alpha^* = |\ln(x/y)/\ln(\rho \vee 0)|$.

Proof of Theorem 3.2.2. Using the same arguments as in of Huang (1992, Lemma 1, page 30)) or Einmahl (1997, Corollary 3.8), we can show that

$$\sup_{0<x,y<T} \left| \sqrt{k}\left(\widehat{\lambda}^{\text{emp}}(x, y) - \lambda^X(x, y)\right) - b_{(C3)} b_{(C2)}(x, y) - \widetilde{B}(x, y) \right| = o_p(1) \qquad (3.6.1)$$

as $n \to \infty$. Note that the above equation can also be shown in a way similar to Schmidt and Stadtmüller (2006) by taking the bias term into account. Since $\lambda(\alpha; x, y, \rho)$ in (3.1.4) is a continuous function of α, by invoking the delta method, the theorem follows from (3.6.1), $\widehat{\tau} - \tau = o_p(1/\sqrt{k})$ (see e.g. Hoeffding (1948)), $\sup_{\theta \in Q^*} |\lambda'(\alpha; \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \rho)| < \infty$ and a Taylor expansion.

Proof of Theorem 3.2.3. It easily follows from (3.1.4) and Theorem 3.2.2.

Proof of Theorem 3.5.1. Since copulae are invariant under strictly increasing transformations, we can assume w.l.o.g that $\mathbf{A}\mathbf{A}^T = \mathbf{R}$ is the correlation matrix. Therefore, the $Z_i \overset{\text{d}}{=} G\mathbf{A}_{i\,.}\boldsymbol{U}^{(d)}$, $1 \le i \le d$, have the same distribution, say F_Z. Hence

$$\begin{aligned} & P\left(1 - F_Z(Z_1) < tx_1, \ldots, 1 - F_Z(Z_d) < tx_d\right) \\ = & \int_{\boldsymbol{u} \in \mathcal{S}_d, \mathbf{A}_{1\,.}\boldsymbol{u} > 0, \ldots, \mathbf{A}_{d\,.}\boldsymbol{u} > 0} P\Big(G > \bigvee_{i=1}^{d} \frac{F_Z^{\leftarrow}(1 - tx_i)}{\mathbf{A}_{i\,.}\boldsymbol{u}}\Big)\, dF_{\boldsymbol{U}^{(d)}}(\boldsymbol{u}), \end{aligned} \qquad (3.6.2)$$

where $F_Z^{\leftarrow}$ denotes the inverse function of F_Z. Since $P(G > \cdot) \in RV_{-\alpha}$ implies that $1 - F_Z \in RV_{-\alpha}$, the inverse function $F_Z^{\leftarrow}$ is regularly varying in 0 with index $-1/\alpha$ (e.g. Resnick (1987, Proposition 0.8(v)). This implies

$$\lim_{t \to 0} \frac{P\left(G > F_Z^{\leftarrow}(1 - tx_i)/(\mathbf{A}_{i\,.}\boldsymbol{u})\right)}{P\left(G > F_Z^{\leftarrow}(1 - t)\right)} = x_i(\mathbf{A}_{i\,.}\boldsymbol{u})^{\alpha}, \quad i = 1, \ldots, d.$$

Now note that, for all $i = 1, \ldots, d$,

$$\begin{aligned} t &= P(Z_i > F_Z^{\leftarrow}(1 - t)) = P(G\mathbf{A}_{i\,.}\boldsymbol{U}^{(d)} > F_Z^{\leftarrow}(1 - t)) \\ &= \int_{\boldsymbol{u} \in \mathcal{S}_d, \mathbf{A}_{i\,.}\boldsymbol{u} > 0} P\left(G > \frac{F_Z^{\leftarrow}(1 - t)}{\mathbf{A}_{i\,.}\boldsymbol{u}}\right) dF_{\boldsymbol{U}^{(d)}}(\boldsymbol{u}), \end{aligned}$$

giving by means of Potter's bounds (e.g. see Geluk and de Haan (1987, (1.20))),

$$\begin{aligned}
&\lim_{t\to 0} \frac{t}{P\left(G > F_Z^{\leftarrow}(1-t)\right)} \\
&= \lim_{t\to 0} \int\limits_{\boldsymbol{u}\in\mathcal{S}_d, \mathbf{A}_i\,.\,\boldsymbol{u}>0} \frac{P(G > F_Z^{\leftarrow}(1-t)/(\mathbf{A}_i\,.\,\boldsymbol{u}))}{P\left(G > F_Z^{\leftarrow}(1-t)\right)}\, dF_{\boldsymbol{U}^{(d)}}(\boldsymbol{u}) \\
&= \int\limits_{\boldsymbol{u}\in\mathcal{S}_d, \mathbf{A}_i\,.\,\boldsymbol{u}>0} (\mathbf{A}_i\,.\,\boldsymbol{u})^{\alpha}\, dF_{\boldsymbol{U}^{(d)}}(\boldsymbol{u}) \quad \forall i = 1, \ldots, d. \qquad (3.6.3)
\end{aligned}$$

Applying the same method to (3.6.2) yields the proof. □

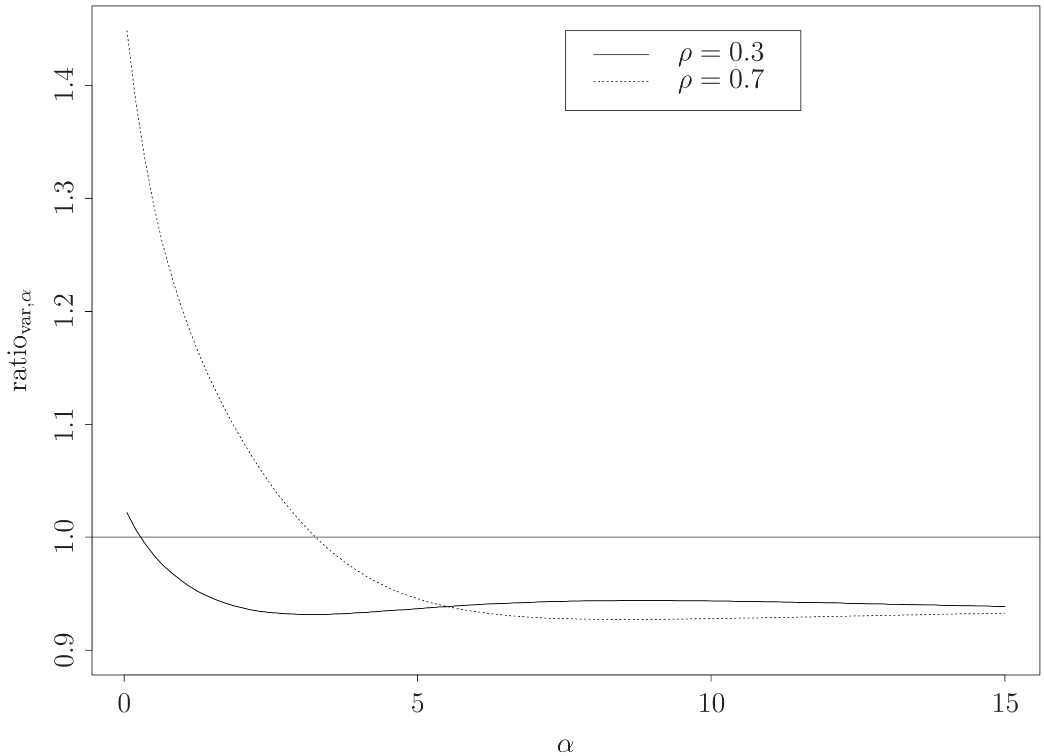

Figure 3.1: Theoretical ratios, $\text{ratio}_{\text{var},\alpha}$, are plotted against α for $\rho = 0.3$ and 0.7.

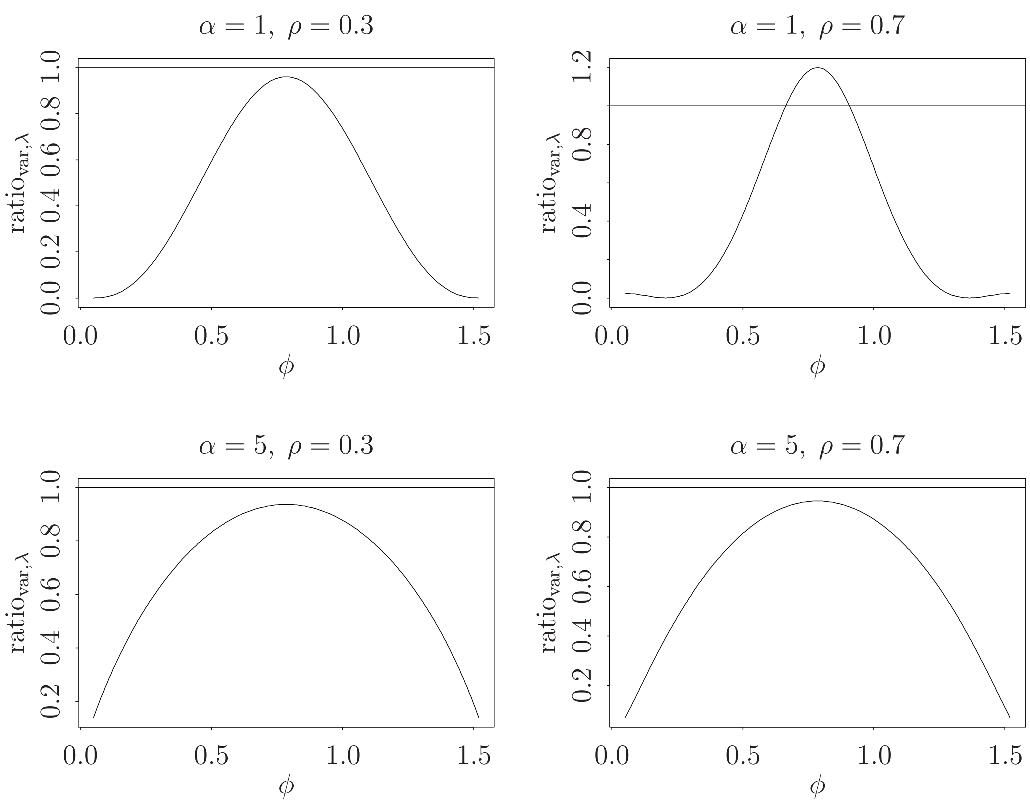

Figure 3.2: Theoretical ratios, $\text{ratio}_{\text{var},\lambda}$, are plotted against $\phi \in (0, \pi/2)$ for $(\alpha, \rho) \in \{1, 5\} \times \{0.3, 0.7\}$.

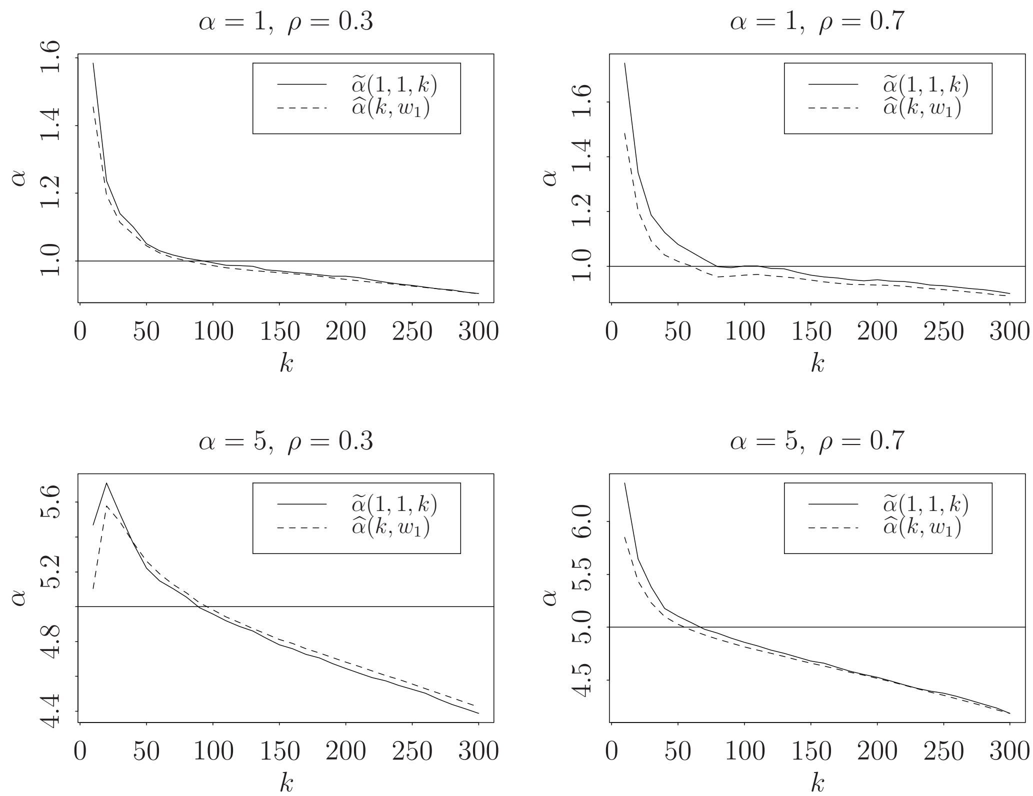

Figure 3.3: Averages of $\widetilde{\alpha}(1,1,k)$ and $\widehat{\alpha}(k,w_1)$ are plotted against $k = 10, 20, \ldots, 300$.

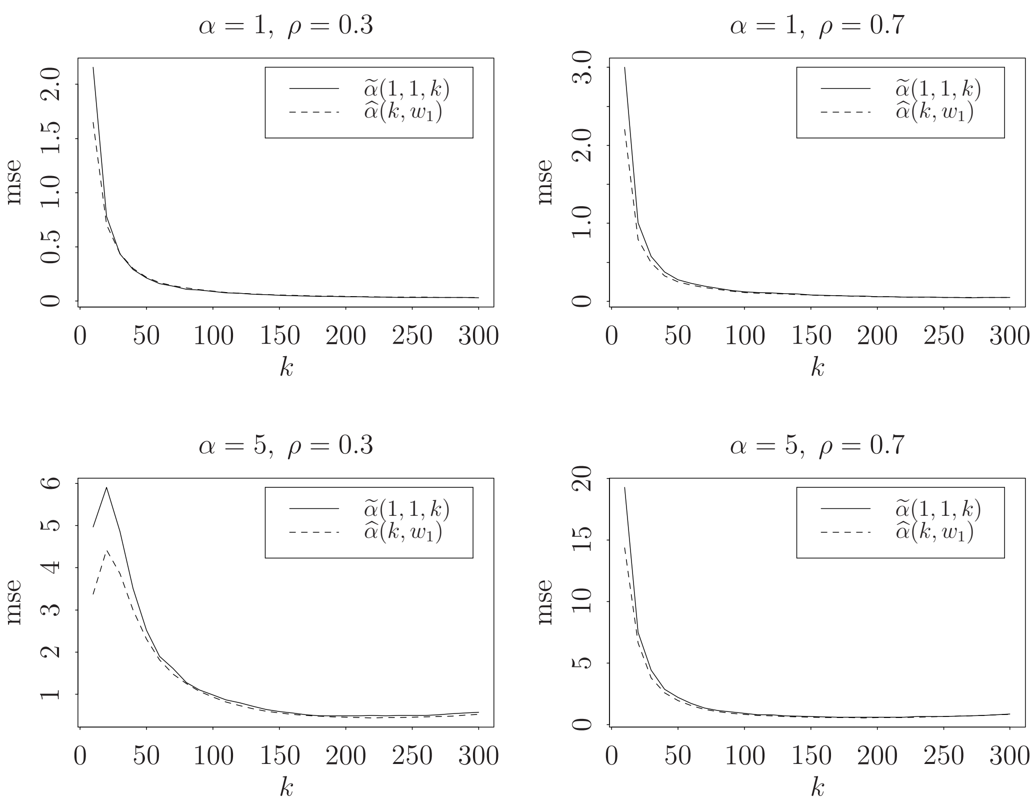

Figure 3.4: Estimated mean squared errors of estimators in Figure 3.3.

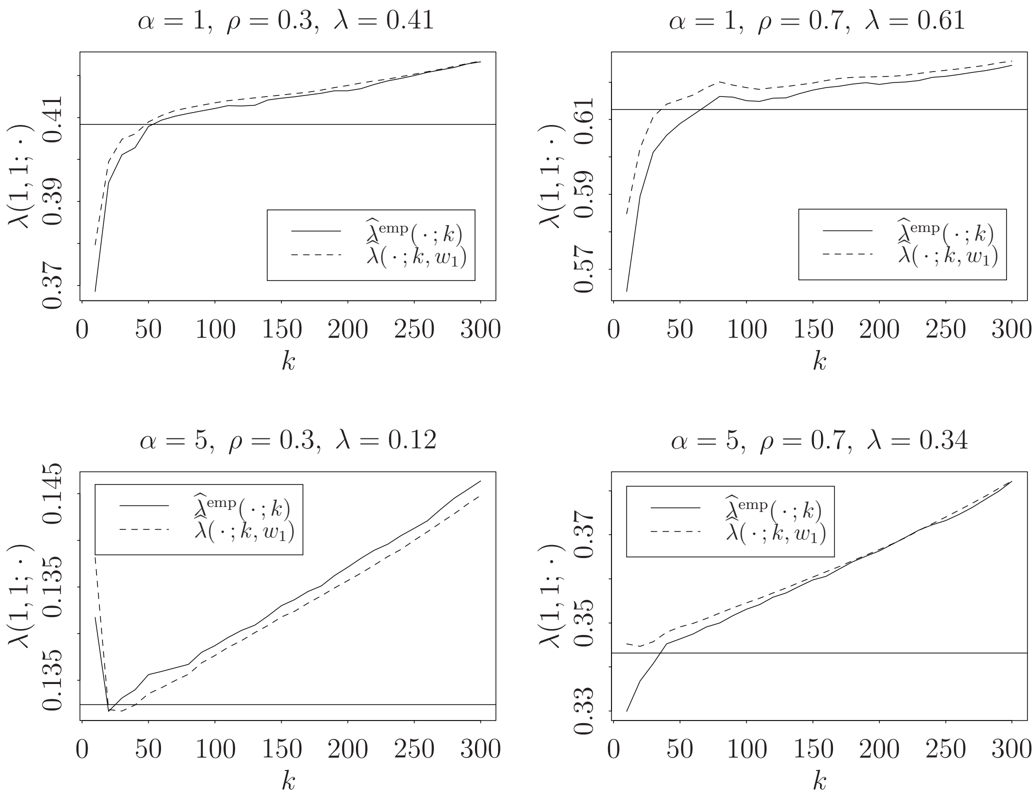

Figure 3.5: Averages of $\widehat{\lambda}^{\text{emp}}(1,1;k)$ and $\widehat{\lambda}(1,1;k,w_1)$ are plotted against $k = 10, 20, \ldots, 300$.

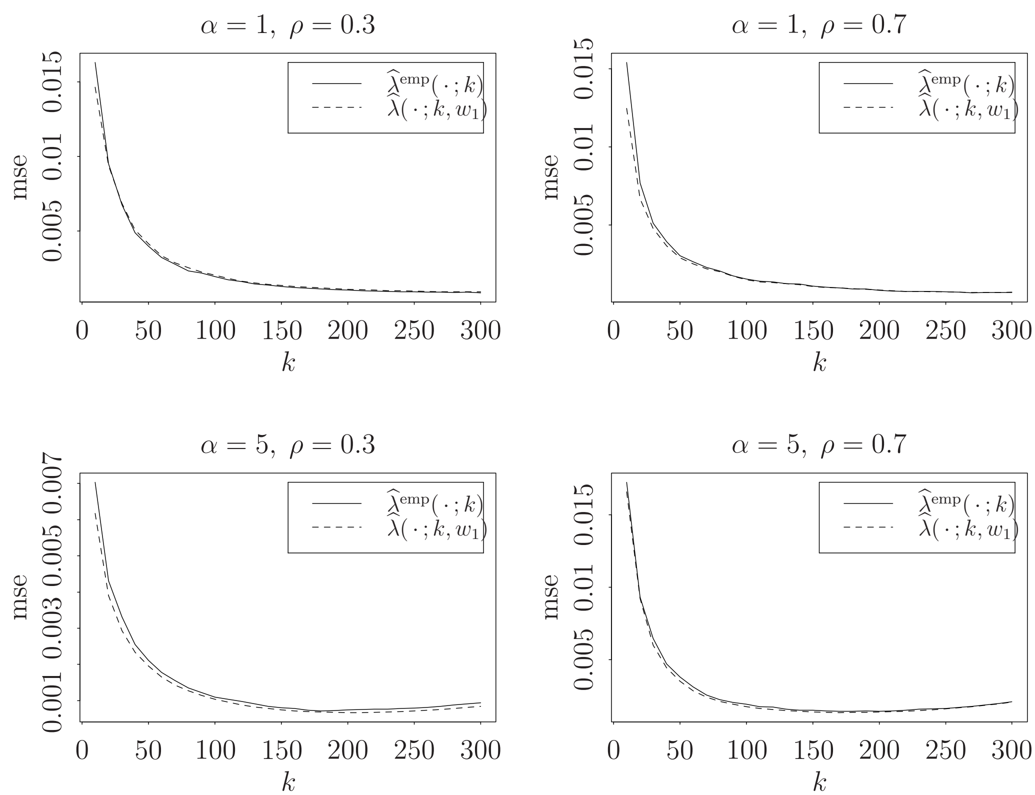

Figure 3.6: Estimated mean squared errors of estimators in Figure 3.5.

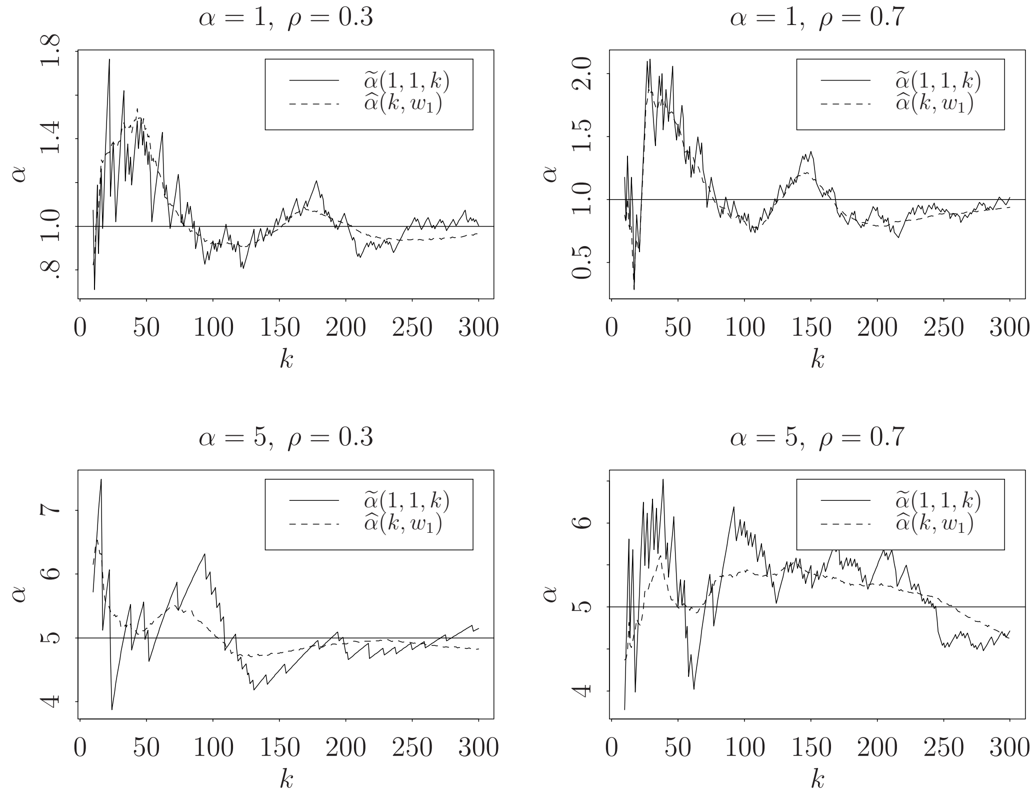

Figure 3.7: Estimators $\widetilde{\alpha}(1,1,k)$ and $\widehat{\alpha}(k,w_1)$ based on a particular sample are plotted against $k = 10, 11, \ldots, 300$.

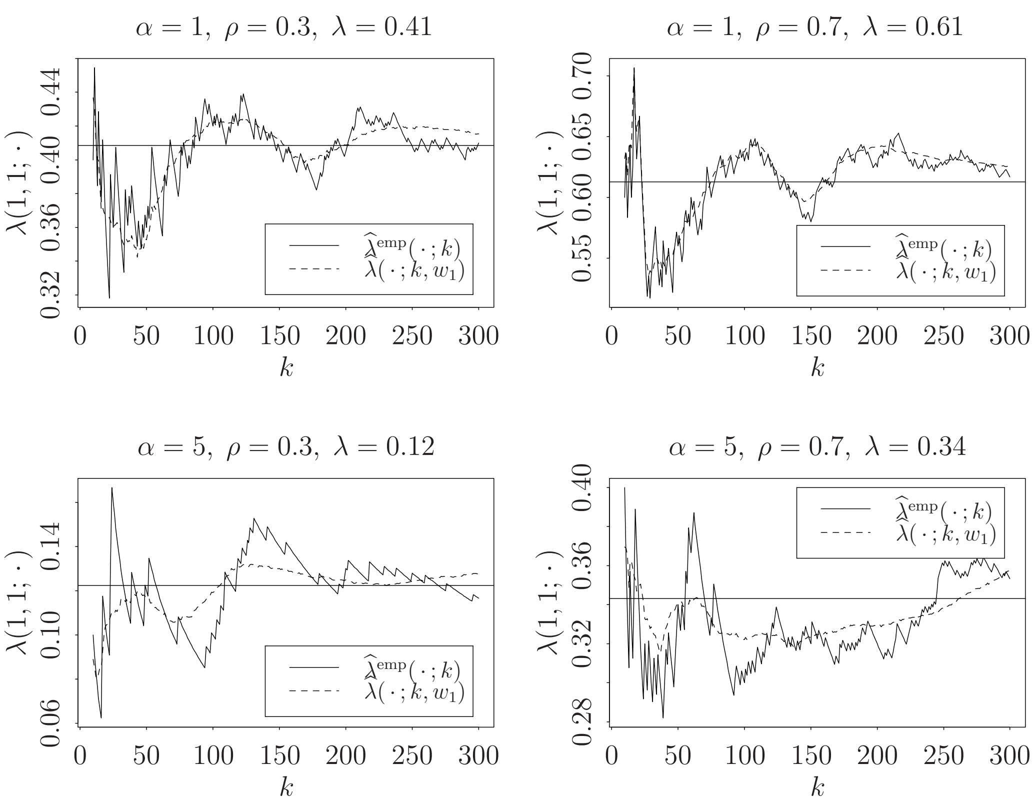

Figure 3.8: Estimators $\widehat{\lambda}^{\text{emp}}(1,1;k)$ and $\widehat{\lambda}(1,1;k,w_1)$ based on a particular sample are plotted against $k = 10, 11, \ldots, 300$.

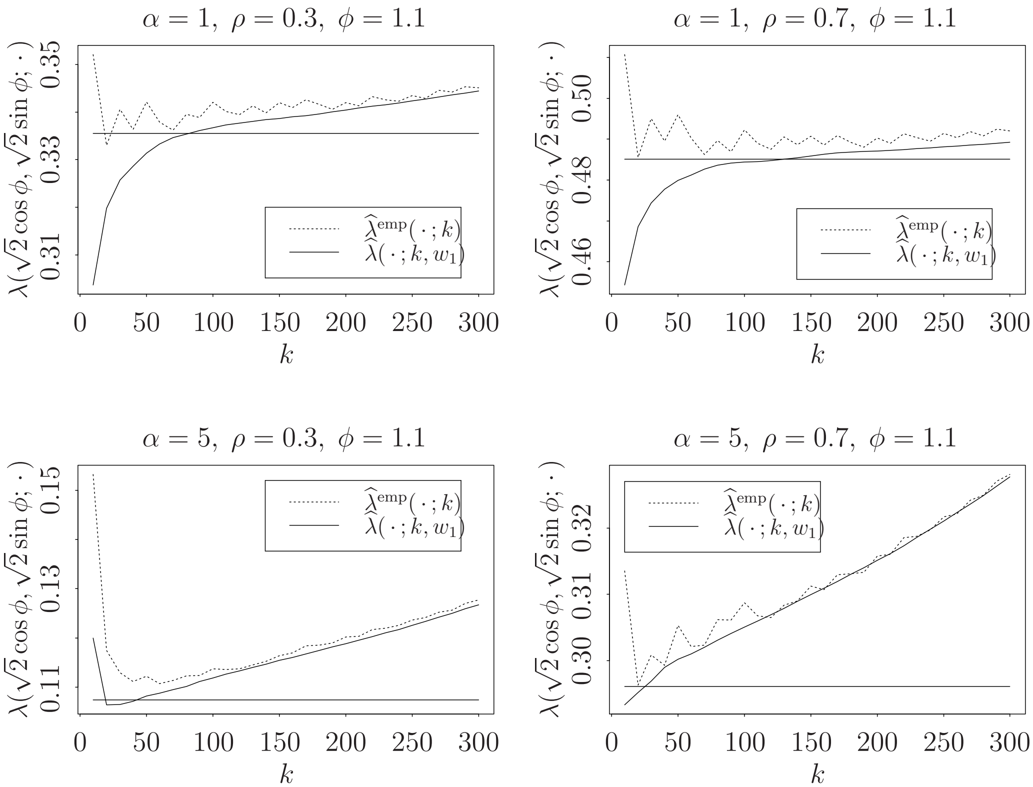

Figure 3.9: Averages of $\widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; k)$ and $\widehat{\lambda}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; k, w_1)$ with $\phi = 1.1$ are plotted against $k = 10, 20, \ldots, 300$.

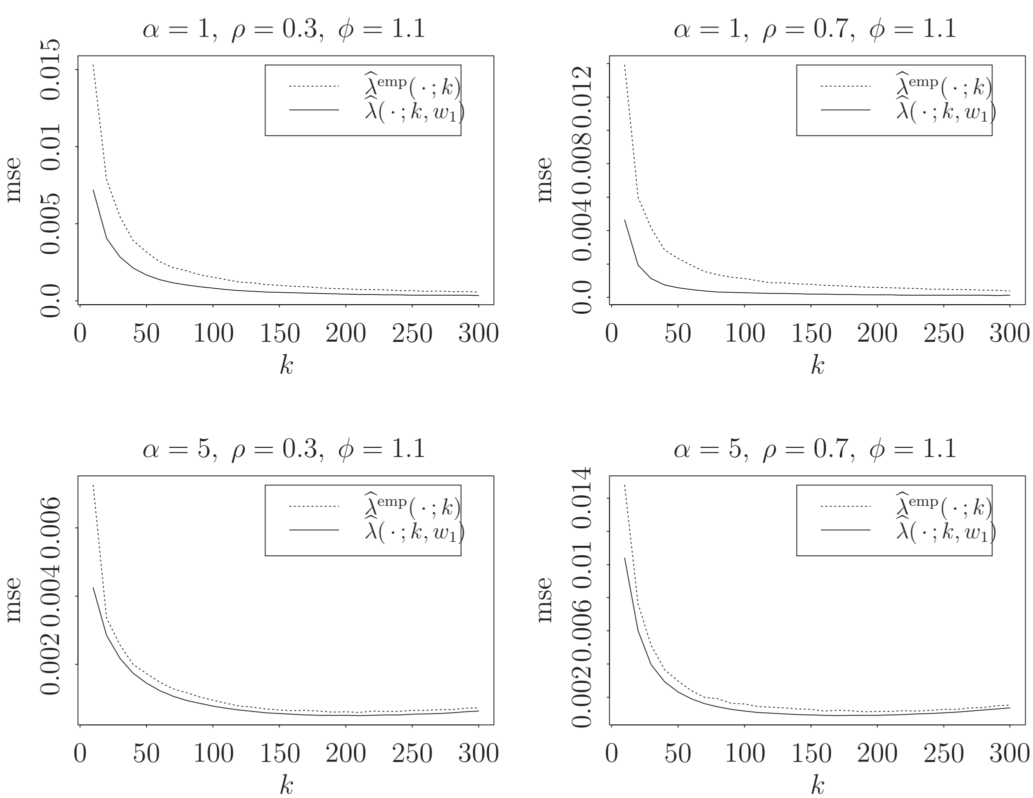

Figure 3.10: Estimated mean squared errors of estimators in Figure 3.9.

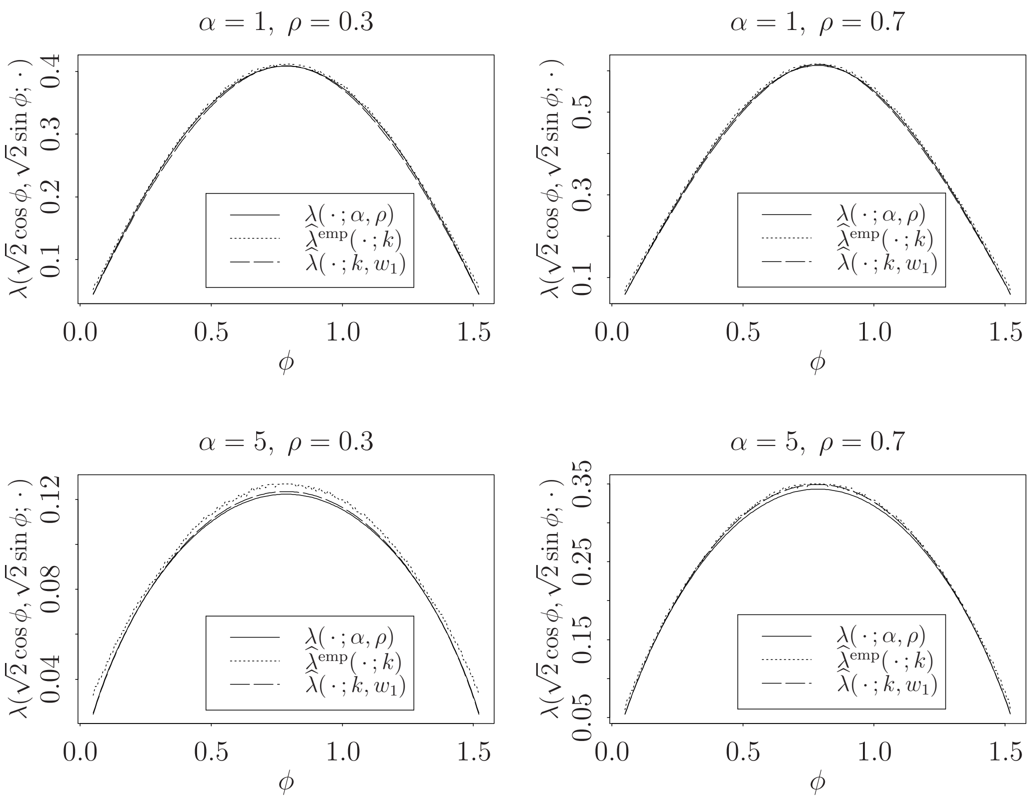

Figure 3.11: Averages of $\widehat{\lambda}^{\mathrm{emp}}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50)$ and $\widehat{\lambda}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50, w_1)$ are plotted against $\phi \in (0, \pi/2)$.

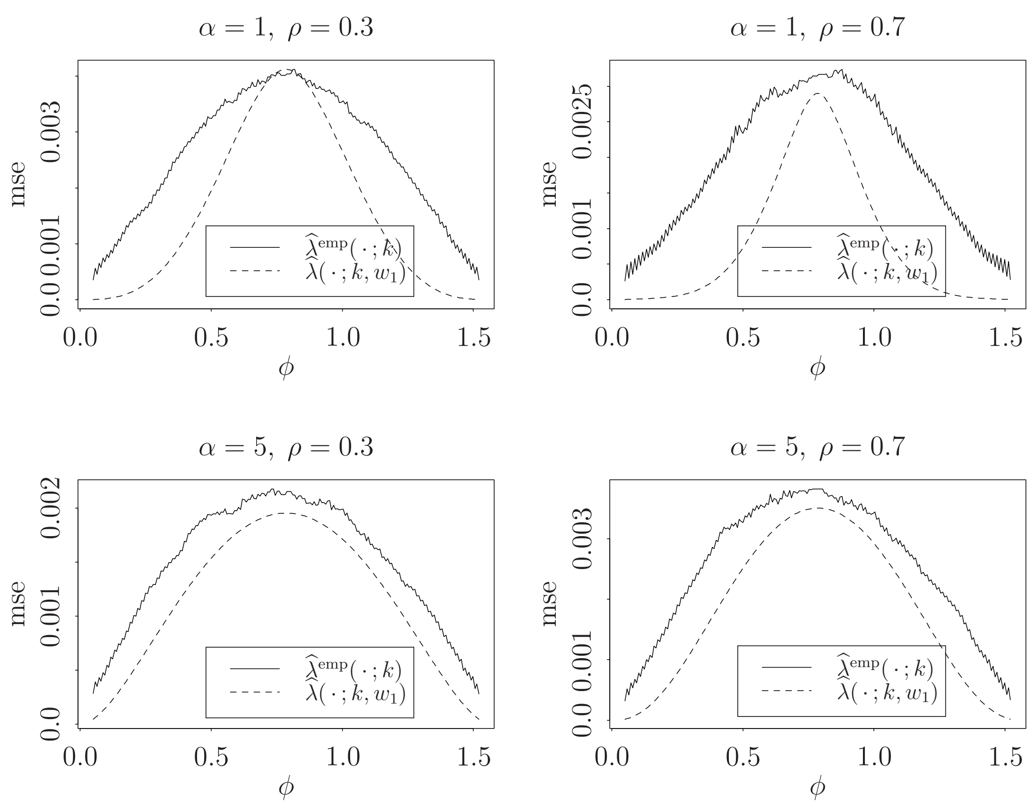

Figure 3.12: Estimated mean squared errors of estimators in Figure 3.11.

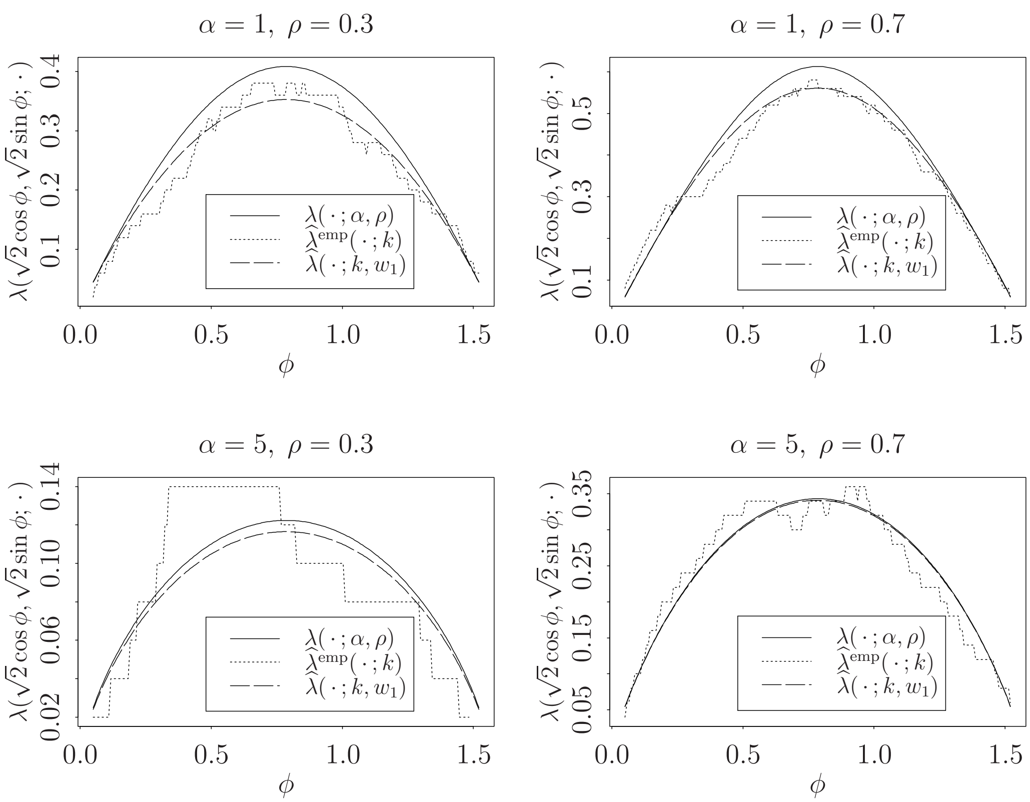

Figure 3.13: Estimators $\widehat{\lambda}^{\text{emp}}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50)$ and $\widehat{\lambda}(\sqrt{2}\cos\phi, \sqrt{2}\sin\phi; 50, w_1)$ based on a particular sample are plotted against $\phi \in (0, \pi/2)$.

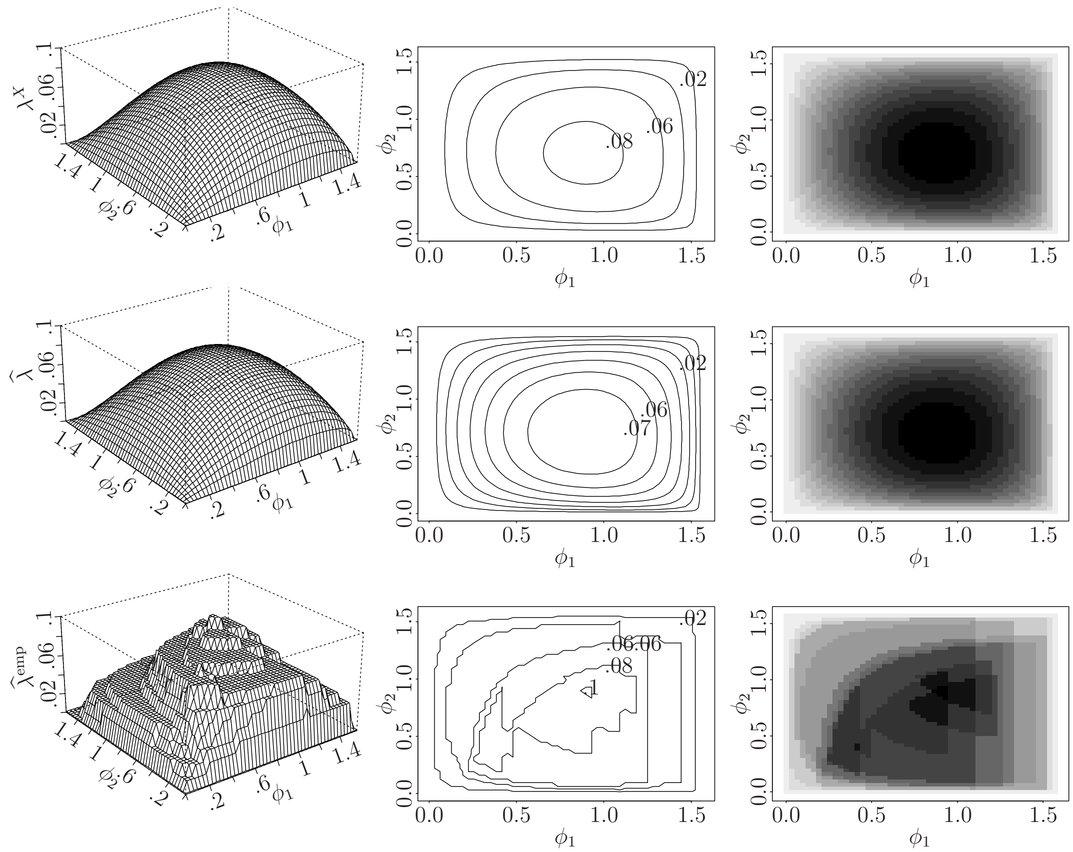

Figure 3.14: From left to right column: perspective, contour and grey-scale image plot of true $\lambda^X\left(\sqrt{3}\cos\phi_1, \sqrt{3}\sin\phi_1\cos\phi_2, \sqrt{3}\sin\phi_1\sin\phi_2\right)$ with parameters $\rho_{12} = 0.3$, $\rho_{13} = 0.5$, $\rho_{23} = 0.7$ and $\alpha = 5$ (first row) and corresponding estimators based on a particular sample, $\widehat{\lambda}(...; 100, w_1)$ (middle row) and $\widehat{\lambda}^{\text{emp}}(...; 100)$ (lower row).

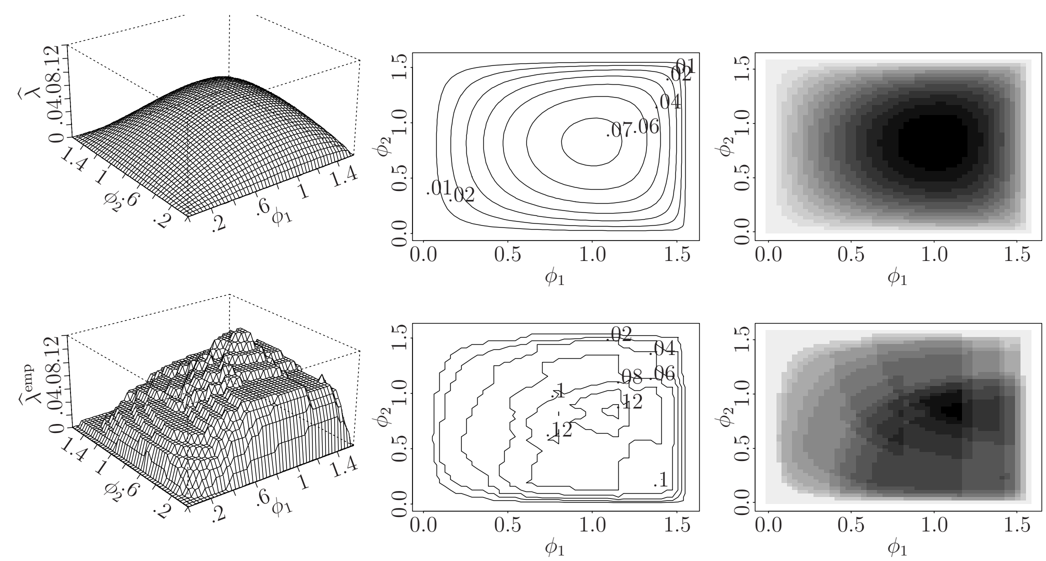

Figure 3.15: From left to right column: perspective, contour and grey-scale image plot of estimators $\widehat{\lambda}(...; 100, w_1)$ (upper row) and $\widehat{\lambda}^{\text{emp}}(...; 100)$ (lower row) of currencies (GBP, USD, CHF).

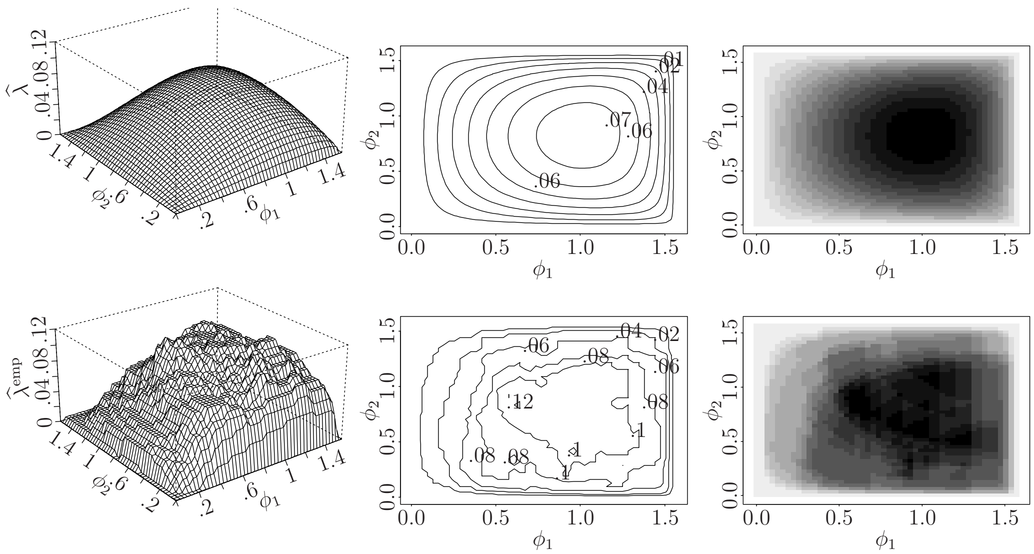

Figure 3.16: Same as Figure 3.15 but for $k = 150$.

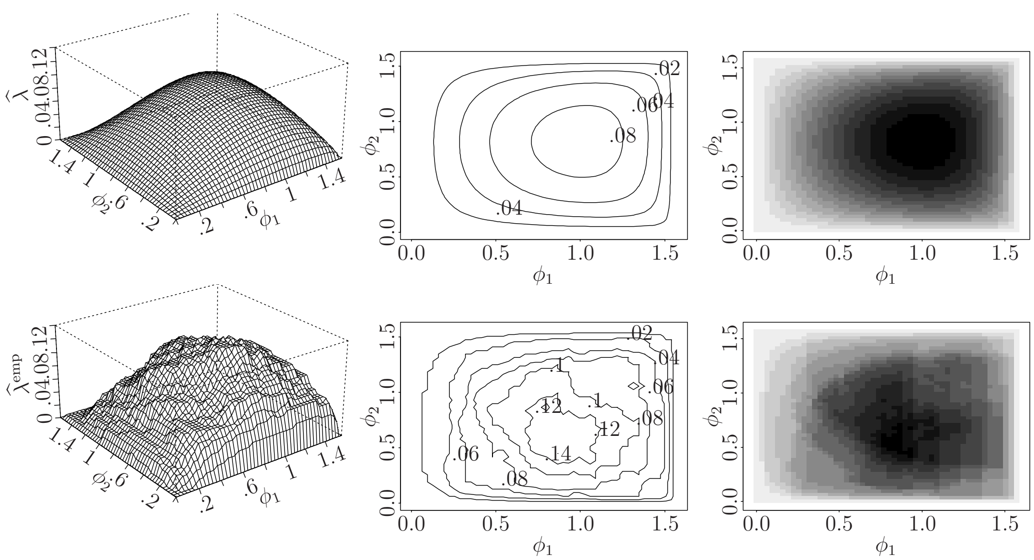

Figure 3.17: Same as Figure 3.15 but for $k = 200$.

Chapter 4

Copula Structure Analysis Based on Robust and Extreme Dependence Measures

Summary

In this paper we extend the standard approach of correlation structure analysis in order to reduce the dimension of highdimensional statistical data. The classical assumption of a linear model for the distribution of a random vector is replaced by the weaker assumption of a model for the copula. For elliptical copulae a 'correlation-like' structure remains but different margins and non-existence of moments are possible. Moreover, elliptical copulae allow also for a 'copula structure analysis' of dependence in extremes. After introducing the new concepts and deriving some theoretical results we observe in a simulation study the performance of the estimators: the theoretical asymptotic behavior of the statistics can be observed even for a sample of only 100 observations. Finally, we test our method on real financial data and explain differences between our copula based approach and the classical approach. Our new method yields a considerable dimension reduction also in non-linear models.

4.1 Introduction

When analyzing high-dimensional data one is often interested in understanding the dependence structure aiming at a dimension reduction. In the framework of correlation representing linear dependence, *correlation structure analysis* is a classical tool; see Steiger

(1994) or Bentler and Dudgeon (1996). Correlation structure analysis is based on the assumption that the correlation matrix of the data satisfies the equation $\mathbf{R} = \mathbf{R}(\boldsymbol{\vartheta})$ for some function $\mathbf{R}(\boldsymbol{\vartheta})$ and a parameter vector $\boldsymbol{\vartheta}$. Typically, a *general linear structure model* is then considered for a d-dimensional random vector $\boldsymbol{X}$, i.e. $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \mathbf{A}\boldsymbol{\xi}$, where $\mathbf{A} = \mathbf{A}(\boldsymbol{\vartheta})$ is a function of a parameter vector $\boldsymbol{\vartheta}$, and $\boldsymbol{\xi}$ represents some (latent) random vector.

The typical goal of correlation structure analysis is to reduce dimension, i.e. to explain the whole dependence structure through lower dimensional parameters summarized in $\boldsymbol{\vartheta}$. One particularly popular method is *factor analysis*, where the data $\boldsymbol{X}$ are assumed to satisfy the linear model $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{L}\boldsymbol{f} + \mathbf{V}\boldsymbol{e}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T$, $\boldsymbol{f} = (f_1, \ldots, f_m)^T$ $(m < d)$ are non-observable and (usually) uncorrelated *factors* and $\boldsymbol{e} = (e_1, \ldots, e_d)^T$ is some *noise variables*. Further, $\mathbf{L} \in \mathbb{R}^{d \times m}$ is called *loading matrix* and $\mathbf{V}$ is a diagonal matrix with nonnegative entries. An often used additional assumption is that $(\boldsymbol{f}^T, \boldsymbol{e}^T)$ has mean zero and covariance matrix $\mathbf{I}$, the identity matrix. Then, describing the dependence structure of $\boldsymbol{X}$ through its covariance matrix yields $\mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{V}^2$, i.e., the dependence of $\boldsymbol{X}$ is described through the entries of $\mathbf{L}$.

Provided that the data are normally distributed this approach of decomposing the correlation structure is justified, since dependence in normal data is uniquely determined by correlation. However, many data sets exhibit properties contradicting the assumption of normality, see e.g. Cont (2001) for a study of financial data. Further, several covariance structure studies based on the normal assumption exhibit problems for nonnormal data, see e.g. Browne (1982, 1984). A modified approach is to assume an elliptical model, and the corresponding methods can be found for instance in Muirhead and Waternaux (1980) and Browne and Shapiro (1987). Browne (1982, 1984) also developed a method being asymptotically free of any distributional assumption, but it was found that acceptable performance of this procedure requires very large sample sizes; see Hu, Bentler, and Kano (1992).

Relaxing more and more the assumptions of classical correlation structure analysis as indicated above, one assumption still remains, namely that $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \mathbf{A}(\boldsymbol{\vartheta})\boldsymbol{\xi}$, i.e. $\boldsymbol{X}$ can be described as a linear combination of some (latent) random variables $\boldsymbol{\xi}$ with existing second moments (and existing fourth moments to ensure asymptotic distributional limits of sample covariance estimators). For real multivariate data it may happen that some margins are well modeled as being normal and some are more heavy-tailed (or leptokurtic). Moreover, nonlinear dependence can occur, e.g. in financial portfolios of assets and derivatives. If this happens, it is hard to believe that some linear model is appropriate. Since the primary aim of correlation or covariance structure analysis is to decompose and describe dependence we present a simple method to avoid problems of non-existing moments or different marginal distributions by using *copulae*. A copula is a d-dimensional distribution function with unif$(0,1)$ margins and, by Sklar's theorem, each distribution function can be described through its margins and its copula separately. We will focus on *elliptical copulae* being the copulae of elliptical distributions, which are very flexible and easy to handle also in high dimensions. As a correlation matrix is a parameter of an elliptical copula, correlation structure analysis can be easily extended to such copulae and we will call this method *copula structure analysis*.

In many applications dependence in extremes is an important issue. For example, financial risk management is confronted with problems concerning joint extreme losses, and one of its prominent questions is how to measure or understand dependence in the extremes; see e.g. McNeil, Frey, and Embrechts (2005). This requires a different approach and is one of the major issue of this paper. We assess extreme dependence by a concept called *tail copula*. For such elliptical copulae, which model extreme dependence, we present a new structure analysis based on the tail copula. This focusses on dependence structure in the extremes.

Our paper is organized as follows. We start with definitions and preliminary results on copulae and elliptical distributions in Section 4.2. In Section 4.3 we introduce the new copula structure model and show which (classical) methods can be used for a structure analysis and model selection. In Section 4.4 we show two copula dependence concepts, one based on Kendall's tau, one on the tail copula, and develop estimators, which can then be used for the copula structure analysis. These concepts lead to different estimates of the copula structure model, and we derive asymptotic results for our estimates.

In Section 4.5 a simulation study shows that the derived asymptotic results hold already for a rather small simulated sample. Finally, we fit a copula factor model to real data based on both our dependence concepts and give an interpretation of the results. Proofs are summarized in Section 4.6.

4.2 Preliminaries

First, we introduce the copula concept. For more technical background information we refer to Nelsen (1999).

Definition 4.2.1. *A* copula $C : [0,1]^d \to [0,1]$ *is a d-dimensional distribution function with standard uniform margins, i.e.* $C(1,\ldots,1,u_j,1,\ldots,1) = u_j$, $1 \le j \le d$.

The following theorem shows that each multivariate distribution function can be separated in its dependence structure, i.e. the copula, and its margins. This important result is used in essentially all applications of copulae. We shall need the notion of a generalized inverse function. For a distribution function F the *generalized inverse* is defined as $F^{\leftarrow}(y) = \inf\{x \in \mathbb{R} \mid F(x) \geq y\}$, $y \in (0,1)$, and $\mathrm{Ran}F := F(\mathbb{R})$ denotes the *range* of F.

Theorem 4.2.2 (Sklar's Theorem (1996))**.** *Let F be a d-dimensional distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula C such that for all $\boldsymbol{x} \in \mathbb{R}^d$*

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

The copula C is unique on $\mathrm{Ran}F_1 \times \cdots \times \mathrm{Ran}F_d$.
If F is a continuous d-dimensional distribution function with margins $F_1, \ldots, F_d$, and generalized inverse functions $F_1^{\leftarrow}, \ldots, F_d^{\leftarrow}$, then the copula C of F is $C(u_1, \ldots, u_d) = F\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right)$.

We will focus on copulae of elliptical distributions, and we first give some definitions and state some properties. For a general treatment of elliptical distributions we refer to Fang, Kotz, and Ng (1990) and to Cambanis, Huang, and Simons (1981). Elliptical copulae and their properties have also been investigated with respect to financial application by Embrechts, Lindskog, and McNeil (2003) or Frahm, Junker, and Szimayer (2003).

Definition 4.2.3. *A d-dimensional random vector $\boldsymbol{X}$ has an elliptical distribution, if, for some $\boldsymbol{\mu} \in \mathbb{R}^d$, some positive (semi-)definite matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{1 \leq i,j \leq d} \in \mathbb{R}^{d \times d}$, a positive random variable G and a random vector $\boldsymbol{U}^{(m)} \sim \mathrm{unif}\{\boldsymbol{s} \in \mathbb{R}^m : \boldsymbol{s}^T\boldsymbol{s} = 1\}$ independent of G it holds that $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + G\mathbf{A}\boldsymbol{U}^{(m)}$, $\mathbf{A} \in \mathbb{R}^{d \times m}$, $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$ and some $m \in \mathbb{N}$. We write $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, G)$. The random variable G is called* generating variable. *Further, if the first moment exists, then $E\boldsymbol{X} = \boldsymbol{\mu}$, and if the second moment exists, then G can be chosen such that* $\mathrm{Cov}\boldsymbol{X} = \boldsymbol{\Sigma}$.

Definition 4.2.4. *Let $\boldsymbol{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \Phi)$ with $\boldsymbol{\Sigma} = (\sigma_{ij})_{1 \leq i,j \leq d}$. We define the* correlation matrix $\mathbf{R}$ *by* $\mathbf{R} := \left(\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}\right)_{1 \leq i,j \leq d}$. *If $\boldsymbol{X}$ has finite second moment, then* $\mathrm{Corr}\boldsymbol{X} = \mathbf{R}$.

Definition 4.2.5. *We define an* elliptical copula *as the copula of an elliptical random vector. Let $\mathbf{R}$ be the correlation matrix corresponding to $\boldsymbol{\Sigma}$. We denote the copula of $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, G)$ by $\mathcal{EC}_d(\mathbf{R}, G)$ and call $\mathbf{R}$ the* copula correlation matrix.

The following corollary shows that the notation $\mathcal{EC}_d(\mathbf{R}, G)$ of elliptical copulae is reasonable. It is a simple consequence of the definition and the fact that copulae are invariant under strictly increasing transformations; see Embrechts et al. (2003, Theorem 2.6).

Corollary 4.2.6. *An elliptical copula is characterized by the generating variable G and the copula correlation matrix $\mathbf{R}$. The generating variable G is uniquely determined up to some positive constant.*

Based on elliptical copulae, we can now formulate the copula structure model.

4.3 Copula structure models

First, we give some notations: let $\boldsymbol{\vartheta} \in \Theta \subset \mathbb{R}^p$ be a p-dimensional parameter vector in some parameter space Θ with $\dim(\Theta) \leq p$. A *correlation structure model* is then a function

$$\mathbf{R} : \Theta \mapsto \mathbb{R}^{d \times d}, \quad \boldsymbol{\vartheta} \to \mathbf{R}(\boldsymbol{\vartheta}), \tag{4.3.1}$$

such that $\mathbf{R}(\boldsymbol{\vartheta})$ is a correlation matrix, i.e. $\mathbf{R}(\boldsymbol{\vartheta})$ is positive definite with diagonal $\mathbf{1}$. As we will later also use vector notation, we denote by vec[$\cdot$] the column vector formed from the non-duplicated and non-fixed elements of a symmetric matrix. If a matrix $\mathbf{A}$ is not symmetric, then vec[$\mathbf{A}$] denotes the column vector formed from all non-fixed elements of the columns of $\mathbf{A}$. In case of a correlation matrix

$$\boldsymbol{r} \ := \ \mathrm{vec}[\mathbf{R}] \in \mathbb{R}^{d(d-1)/2}. \tag{4.3.2}$$

For a general linear correlation structure model, (4.3.1) corresponds to the following situation: let $\boldsymbol{\xi} \in \mathcal{E}_q(\mathbf{0}, \mathbf{I}, G)$ and let $\mathbf{A} : \Theta \mapsto \mathbb{R}^{d \times q}$, $\boldsymbol{\vartheta} \to \mathbf{A}(\boldsymbol{\vartheta})$, be some matrix valued function and define

$$\boldsymbol{\Sigma} : \Theta \mapsto \mathbb{R}^{d \times d}, \quad \boldsymbol{\vartheta} \to \boldsymbol{\Sigma}(\boldsymbol{\vartheta}) := \mathbf{A}(\boldsymbol{\vartheta})\mathbf{A}(\boldsymbol{\vartheta})^T.$$

Then, (4.3.1) can be written as $\mathbf{R}(\boldsymbol{\vartheta}) = \mathrm{diag}[\boldsymbol{\Sigma}(\boldsymbol{\vartheta})]^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{\vartheta})\mathrm{diag}[\boldsymbol{\Sigma}(\boldsymbol{\vartheta})]^{-1/2}$.

4.3.1 The model

As by Definition 4.2.5 a correlation matrix is a parameter of an elliptical copula, we can extend the usual correlation structure model to elliptical copulae.

Definition 4.3.1. *Let $\boldsymbol{\vartheta} \in \Theta \subset \mathbb{R}^p$ be a p-dimensional parameter vector, $\mathbf{A} : \Theta \mapsto \mathbb{R}^{d \times q}$ a matrix valued function and $\boldsymbol{\xi} \in \mathcal{E}_q(\mathbf{0}, \mathbf{I}, G)$ a q-dimensional elliptical random vector with continuous generating variable $G > 0$. Further, denote by $C_{\mathbf{A}(\vartheta)\boldsymbol{\xi}}$ the copula of $\mathbf{A}(\boldsymbol{\vartheta})\boldsymbol{\xi} \in \mathbb{R}^d$.*

We say that the random vector $\boldsymbol{X} \in \mathbb{R}^d$ *with copula* $C_{\boldsymbol{X}}$ *satisfies a* copula structure model*, if*

$$C_{\boldsymbol{X}} = C_{\mathbf{A}(\boldsymbol{\vartheta})\boldsymbol{\xi}} \in \mathcal{EC}_d(\mathbf{R}(\boldsymbol{\vartheta}), G), \tag{4.3.3}$$

where $\mathbf{R}(\boldsymbol{\vartheta}) := \text{diag}[\boldsymbol{\Sigma}(\boldsymbol{\vartheta})]^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{\vartheta})\text{diag}[\boldsymbol{\Sigma}(\boldsymbol{\vartheta})]^{-1/2}$ *and* $\boldsymbol{\Sigma}(\boldsymbol{\vartheta}) := \mathbf{A}(\boldsymbol{\vartheta})\mathbf{A}(\boldsymbol{\vartheta})^T$.

Remark 4.3.2. (i) Define by $\boldsymbol{F}^{\leftarrow}(\boldsymbol{u}) := (F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$ the vector of the inverses of the marginal distribution functions of $\boldsymbol{X}$ and by $\boldsymbol{H}(\boldsymbol{x}) := (H_1(x_1), \ldots, H_d(x_d))$ the vector of the marginal distribution functions of $\mathbf{A}(\boldsymbol{\vartheta})\boldsymbol{\xi}$. Then, (4.3.3) is equivalent to $\boldsymbol{X} \stackrel{\text{d}}{=} \boldsymbol{F}^{\leftarrow}(\boldsymbol{H}(\mathbf{A}(\boldsymbol{\vartheta})\boldsymbol{\xi}))$, where all operations are componentwise. Hence, the copula model can also be seen as an extension of a correlation structure model for elliptical data: if not only $C_{\boldsymbol{X}} = C_{\mathbf{A}(\boldsymbol{\vartheta})\boldsymbol{\xi}}$ holds but also $\boldsymbol{H} = \boldsymbol{F}$ with existing second moment, then this would be a classical correlation or covariance structure model. For normal $\boldsymbol{\xi}$ it gives back the standard normal model and for elliptical $\boldsymbol{\xi}$ the elliptical model of Browne (1984).

(ii) The classical correlation structure model assumes some (functional) structure for the correlation matrix of the observed data. In the copula structure model this functional structure prevails. The only difference lies in the interpretation of the 'correlation' matrix. In the classical model it represents the linear correlation between the data, in the copula model it represents a dependence parameter which can be interpreted as a 'correlation-like' measure; see Lemma 4.2.6.

Example 4.3.3. For classical factor analysis, (4.3.3) translates to $\boldsymbol{\vartheta} = \text{vec}[\mathbf{L}, \mathbf{V}]$, $\mathbf{R}(\boldsymbol{\vartheta}) = \mathbf{L}\mathbf{L}^T + \mathbf{V}^2$ for some $m < d$, $\mathbf{L} \in \mathbb{R}^{d\times m}$ and a diagonal matrix (with nonnegative entries) $\mathbf{V} \in \mathbb{R}^{d\times d}$. The corresponding copula structure model assumes that there exists $\boldsymbol{\xi} \in \mathcal{E}_{m+d}(\mathbf{0}, \mathbf{I}, G)$ such that

$$C_{\boldsymbol{X}} \quad = \quad C_{(\mathbf{L},\mathbf{V})\boldsymbol{\xi}}. \tag{4.3.4}$$

We call this identity a *copula factor model.* An example of this copula factor model is the *Credit Metrics* model in the framework of credit risk, see e.g. Bluhm, Overbeck, and Wagner (2003, Section 2.4). There, a factor model $\boldsymbol{X} = (X_1, \ldots, X_d)^T = \mathbf{L}\boldsymbol{f} + \mathbf{V}\boldsymbol{e}$ is assumed for the underlying (latent) variables of a set of credit default indicators $\left(I_{\{X_i<s_i\}}\right)_{1\leq i\leq d}$ and $\boldsymbol{X}$ is assumed to be normal. By Frey, McNeil, and Nyfeler (2001, Proposition 2), the distribution of $\left(I_{\{X_i<s_i\}}\right)_{1\leq i\leq d}$ is uniquely determined by the single default probabilities $P(I_{\{X_i<s_i\}} = 1)$ and the copula of $\boldsymbol{X}$. Therefore, in this case the assumption of $\boldsymbol{X} = \mathbf{L}\boldsymbol{f} + \mathbf{V}\boldsymbol{e}$ is equivalent to $C_{\boldsymbol{X}} = C_{(\mathbf{L},\mathbf{V})\boldsymbol{\xi}}$ with $\boldsymbol{\xi} \sim \mathcal{N}_{m+d}(\mathbf{0}, \mathbf{I})$. The model extends easily to non-normal $\boldsymbol{X}$.

4.3.2 Estimation of $\boldsymbol{\vartheta}$

The next step is to estimate a structure model. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be an iid sequence of d-dimensional random vectors and denote by $\widehat{\mathbf{R}} := \widehat{\mathbf{R}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ an estimator of $\mathbf{R}$, a correlation matrix. This estimator can be the empirical correlation or a copula correlation estimator or some other correlation estimator. We review some results from the literature, which we will need for the estimation of the copula structure model later.

Given this estimator $\widehat{\mathbf{R}}$ we want to find some parameter vector $\boldsymbol{\vartheta}$ which fits the assumed structure $\mathbf{R}(\boldsymbol{\vartheta})$ to $\widehat{\mathbf{R}}$ as good as possible. Similarly to (4.3.2), we define $\widehat{\boldsymbol{r}} := \mathrm{vec}[\widehat{\mathbf{R}}]$ and $\boldsymbol{r}(\boldsymbol{\vartheta}) := \mathrm{vec}[\mathbf{R}(\boldsymbol{\vartheta})]$.

Estimation of $\boldsymbol{\vartheta}$ is based on the minimization of a *discrepancy function* $D = D(\widehat{\boldsymbol{r}}, \boldsymbol{r}(\boldsymbol{\vartheta}))$ which measures the discrepancy between the estimated correlation matrix represented by $\widehat{\boldsymbol{r}}$ and $\boldsymbol{r}(\boldsymbol{\vartheta})$. A discrepancy function D has to satisfy

(i) $D \geq 0$,

(ii) $D(\widehat{\boldsymbol{r}}, \boldsymbol{r}) = 0$ if and only if $\widehat{\boldsymbol{r}} = \boldsymbol{r}$ and

(iii) D is twice differentiable with respect to both $\widehat{\boldsymbol{r}}$ and $\boldsymbol{r}$.

Note that the concept of a discrepancy function (without condition (iii)) is weaker than the concept of a metric, as a discrepancy function D does not have to be symmetric or translation invariant in its arguments, nor does it have to satisfy the triangular inequality.

In the following example we present two classical discrepancy functions, for more details about discrepancy functions, their properties, advantages and drawbacks, we refer to Bentler and Dudgeon (1996) and Steiger (1994). For more details about the quadratic form discrepancy function below see Steiger, Shapiro, and Browne (1985).

Example 4.3.4. (i) The *normal theory maximum likelihood discrepancy function* is

$$D_{\mathrm{ML}}(\widehat{\boldsymbol{r}}, \boldsymbol{r}(\boldsymbol{\vartheta})) = \ln |\mathbf{R}(\boldsymbol{\vartheta})| + \mathrm{tr}\left(\widehat{\mathbf{R}}\,(\mathbf{R}(\boldsymbol{\vartheta}))^{-1}\right) - \ln |\widehat{\mathbf{R}}| - d. \tag{4.3.5}$$

This function is the log-likelihood term of $\mathbf{R}(\boldsymbol{\vartheta})$ in case of normal data.

(ii) The *quadratic form* (or *weighted least squares*) *discrepancy function* is

$$D_{\mathrm{QD}}(\widehat{\boldsymbol{r}}, \boldsymbol{r}(\boldsymbol{\vartheta}) | \boldsymbol{\Upsilon}) = (\widehat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{\vartheta}))^T \boldsymbol{\Upsilon}^{-1} (\widehat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{\vartheta}))\,, \tag{4.3.6}$$

where $\boldsymbol{\Upsilon}$ is a positive definite matrix or a consistent estimator of some positive definite matrix $\boldsymbol{\Upsilon}^*$. Note that $D_{\mathrm{QD}}(\cdot, \cdot | \boldsymbol{\Upsilon})$ is a metric.

Given some discrepancy function D and some estimator $\widehat{\mathbf{R}}$ of the correlation matrix $\mathbf{R}$, we can define a consistent estimator of $\boldsymbol{\vartheta}$.

Proposition 4.3.5 (Browne (1984), Proposition 1)**.** *Let* $\mathbf{R}_0$ *be some correlation matrix,* $\boldsymbol{r}_0 := \text{vec}[\mathbf{R}_0] \in \mathbb{R}^{d(d-1)/2}$ *and* $\Theta \subset \mathbb{R}^p$ *a closed and bounded parameter space. Further assume that* $\widehat{\boldsymbol{r}}$ *is an estimator based on an iid sample* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ *of* d*-dimensional random vectors and let* D *be a discrepancy function. Assume that* $\widehat{\boldsymbol{r}} \xrightarrow{P} \boldsymbol{r}_0$ *as* $n \to \infty$ *and that* $\boldsymbol{\vartheta}_0 \in \Theta$ *is the unique minimizer of* $D(\boldsymbol{r}_0, \boldsymbol{r}(\boldsymbol{\vartheta}))$ *in* Θ*. Assume also that the Jacobian matrix* $\partial \boldsymbol{r}(\boldsymbol{\vartheta})/\partial \boldsymbol{\vartheta}^T$ *is continuous in* $\boldsymbol{\vartheta}$*. Define the estimator*

$$\widehat{\boldsymbol{\vartheta}} \;:=\; \arg\min_{\boldsymbol{\vartheta}\in\Theta} D(\widehat{\boldsymbol{r}}, \boldsymbol{r}(\boldsymbol{\vartheta})). \tag{4.3.7}$$

Then $\widehat{\boldsymbol{\vartheta}} \xrightarrow{P} \boldsymbol{\vartheta}_0$ *as* $n \to \infty$*.*

Of course, if we know the true correlation vector $\boldsymbol{r}_0$ satisfying the structure model $\boldsymbol{r}_0 = \boldsymbol{r}(\boldsymbol{\vartheta}_0)$ for some parameter vector $\boldsymbol{\vartheta}_0$, then $\widehat{\boldsymbol{\vartheta}}$ will always be such that $\boldsymbol{r}_0 = \boldsymbol{r}(\boldsymbol{\vartheta}_0) = \boldsymbol{r}(\widehat{\boldsymbol{\vartheta}})$, independent of the choice of the discrepancy function. We also have $D(\boldsymbol{r}_0, \boldsymbol{r}(\widehat{\boldsymbol{\vartheta}})) = 0$ in this case. Since in practice we neither know the true $\boldsymbol{r}_0$ nor the true structure model, we need a method to find an appropriate model.

4.3.3 Model selection

First, we show the asymptotic distribution of a certain test statistic, which will later be used for model selection.

Definition 4.3.6. *Under the settings of Proposition 4.3.5, we define the test statistic*

$$T \;:=\; n\widehat{D} \;=\; nD(\widehat{\boldsymbol{r}}, \boldsymbol{r}(\widehat{\boldsymbol{\vartheta}})) \;=\; n \min_{\boldsymbol{\vartheta}\in\Theta} D(\widehat{\boldsymbol{r}}, \boldsymbol{r}(\vartheta)). \tag{4.3.8}$$

The null hypothesis is that the true correlation vector $\boldsymbol{r}_0$ *satisfies a structure model, i.e.*

$$H_0: \quad \boldsymbol{r}_0 \;=\; \boldsymbol{r}(\boldsymbol{\vartheta}_0) \quad \textit{for some } \boldsymbol{\vartheta}_0 \in \Theta. \tag{4.3.9}$$

To obtain the limit distribution of T we use a version of Steiger et al. (1985, Theorem 1), adapted to our situation. We replace the regularity condition (R7) of that article by the stronger assumption that the null hypothesis (4.3.9) holds. The equivalent statement in case of the quadratic form discrepancy function $D_{\text{QD}}(\cdot,\cdot|\boldsymbol{\Upsilon})$ is given in Browne (1984, Corollary 4.1), where it is additionally required that $\boldsymbol{\Upsilon}$ is a consistent estimator of $\boldsymbol{\Gamma}$, the asymptotic covariance matrix of $\widehat{\boldsymbol{r}}$.

Theorem 4.3.7. *Assume that the conditions of Proposition 4.3.5 hold and* $\boldsymbol{\vartheta}_0$ *is an interior point of* Θ. *Further assume that* $\sqrt{n}(\widehat{\boldsymbol{r}} - \boldsymbol{r}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$ *as* $n \to \infty$ *and that the Hessian matrix*

$$2\boldsymbol{\Psi}_0 \;=\; \left.\frac{\partial^2 D(\boldsymbol{r}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T}\right|_{\boldsymbol{r}=\boldsymbol{\xi}=\boldsymbol{r}_0} \tag{4.3.10}$$

is nonsingular and satisfies $\boldsymbol{\Psi}_0 = \boldsymbol{\Gamma}^{-1}$. *In case of the quadratic form discrepancy function* $D_{\mathrm{QD}}(\cdot, \cdot \,|\, \boldsymbol{\Upsilon})$ *defined in* (4.3.6), *the assumption* (4.3.10) *is replaced by assuming that* $\boldsymbol{\Upsilon}$ *is a consistent estimator of* $\boldsymbol{\Gamma}$. *Also assume that the* $p \times d$ *Jacobian matrix*

$$\triangle \;=\; \left.\frac{\partial \boldsymbol{r}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \tag{4.3.11}$$

is of full column rank p. *Then, under the null hypothesis* (4.3.9),

$$T \;=\; n\widehat{D} \;\xrightarrow{d}\; \chi^2_{df}\,, \quad n \to \infty\,, \tag{4.3.12}$$

where $df = d(d-1)/2 - p^*$ *with* $p^* \le p$ *is the number of free parameters of* $\boldsymbol{\vartheta} \in \Theta \subset \mathbb{R}^p$.

Remark 4.3.8. Under the conditions of Proposition 4.3.5, if $\boldsymbol{\Psi}_0 = \boldsymbol{\Gamma}^{-1}$ does not hold, the limiting distribution of T in (4.3.8) under the null hypothesis (4.3.9) will not be χ^2_{df}, see Satorra and Bentler (2001) or van Praag, Dijkstra, and van Velzen (1985). In this case,

$$T \;\xrightarrow{d}\; \sum_{j=1}^{df} \kappa_j \zeta_j\,, \quad n \to \infty\,,$$

where the ζ_j are iid χ^2_1 distributed and κ_j are the non-null eigenvalues of the matrix $\mathbf{U}\boldsymbol{\Gamma}$ with

$$\mathbf{U} \;=\; \boldsymbol{\Psi}_0 - \boldsymbol{\Psi}_0 \triangle (\triangle^T \boldsymbol{\Psi}_0 \triangle)^{-1} \triangle^T \boldsymbol{\Psi}_0,$$

where $\triangle$ is given in (4.3.11). An example for this situation is $D_{\mathrm{ML}}(\widehat{\boldsymbol{\sigma}}, \boldsymbol{\sigma}(\boldsymbol{\vartheta}))$ given in (4.3.5), where $\widehat{\boldsymbol{\sigma}}$ is the vector of a covariance matrix estimator, $\boldsymbol{\sigma}(\boldsymbol{\vartheta})$ is the vector of a covariance structure model and $\widehat{\boldsymbol{\sigma}}$ has an asymptotic covariance matrix different from the asymptotic covariance matrix of the empirical covariance estimator under a normal population.

From now on we will use the quadratic form discrepancy function $D := D_{\mathrm{QD}}$ from Example 4.3.4(ii), where $\boldsymbol{\Upsilon} = \widehat{\boldsymbol{\Gamma}}$ is an estimator of $\boldsymbol{\Gamma}$. If $\widehat{\boldsymbol{\Gamma}}$ is consistent, Theorem 4.3.7 applies and by Browne (1984, Corollary 2.1), $\widehat{\boldsymbol{\vartheta}}$ is asymptotically normal with covariance matrix $\left(\triangle^T \boldsymbol{\Gamma}^{-1} \triangle\right)^{-1}$, where $\triangle$ is given in (4.3.11). Note that, if $\widehat{\boldsymbol{\Gamma}}$ is only consistent and does not have a finite second moment, large sample sizes may be necessary to observe the limiting χ^2-distribution of the test statistic T or the asymptotic normality of $\widehat{\boldsymbol{\vartheta}}$.

To select an appropriate structural model, we consider a set of g models (which all have to satisfy the assumptions of Theorem 4.3.7)

$$\boldsymbol{r}^{(i)} : \Theta^{(i)} \to \mathbb{R}^{d(d-1)/2}, \quad \boldsymbol{\vartheta}^{(i)} \mapsto \boldsymbol{r}^{(i)}(\boldsymbol{\vartheta}^{(i)}), \quad \text{and } \Theta^{(i)} \subset \mathbb{R}^{p^{(i)}}, \quad 1 \le i \le g. \qquad (4.3.13)$$

Further, we require that the g models are *nested*, i.e. for every $1 \le i \le g-1$ and $\boldsymbol{\vartheta}^{(i)} \in \Theta^{(i)}$ there exists some $\boldsymbol{\vartheta}^{(i+1)} \in \Theta^{(i+1)}$ such that $\boldsymbol{r}^{(i+1)}(\boldsymbol{\vartheta}^{(i+1)}) = \boldsymbol{r}^{(i)}(\boldsymbol{\vartheta}^{(i)})$. Next, define the null hypotheses

$$H_0^{(i)} : \quad \boldsymbol{r}_0 \;=\; \boldsymbol{r}^{(i)}(\boldsymbol{\vartheta}_0) \quad \text{for some } \boldsymbol{\vartheta}_0^{(i)} \in \Theta^{(i)}, \quad 1 \le i \le g,$$

and assume that some of these null hypotheses are true. Then there exists some j such that $H_0^{(i)}$ does not hold for $1 \le i < j$ and does hold for $j \le i \le g$. As we are interested in a structure model, which is likely to explain the observed dependence structure, and is as simple as possible, hence, since the models are nested, we have to estimate j, the smallest index where the null hypothesis holds. By Theorem 4.3.7, the corresponding test statistics $T^{(i)} := nD(\widehat{\boldsymbol{r}}, \boldsymbol{r}^{(i)}(\vartheta_0^{(i)})) := n \min_{\boldsymbol{\vartheta} \in \Theta^{(i)}} D(\widehat{\boldsymbol{r}}, \boldsymbol{r}^{(i)}(\vartheta))$ are not χ^2 distributed for $1 \le i < j$ and are χ^2_{df}-distributed for $j \le i \le g$ with df given in Theorem 4.3.7. Consequently, we reject a null hypothesis $H_0^{(i)}$, if the corresponding test statistic $T^{(i)}$ is larger than some χ^2_{df}-quantile. Hence, j is the smallest number, where $H_0^{(j)}$ cannot be rejected.

Remark 4.3.9. (i) Note that classical estimates of $\boldsymbol{\Gamma}$ rely on the estimation of second and fourth moments of $\boldsymbol{X}$. For non-normal or, especially, for heavy-tailed data these estimates often have large sampling variability and in simulation studies it turned out that large samples are necessary for acceptable performance of the test statistics, see e.g. Hu, Bentler, and Kano (1992).

(ii) In general, a unique *true* parameter $\boldsymbol{\vartheta}_0$ need not exist: in the classical factor model (see Example 4.3.3, where $\mathbf{R} = \mathbf{L}\mathbf{L}^T + \mathbf{V}^2$), $\mathbf{L}$ can always be replaced by $\mathbf{L}^* = \mathbf{L}\mathbf{P}$ for any orthogonal matrix $\mathbf{P}$. By a minor adaption of the parameter space Θ (i.e. $\mathbf{L}^T\mathbf{V}^{-2}\mathbf{L}$ has to be diagonal), $\widehat{\boldsymbol{\vartheta}}$ can be forced to be unique and Proposition 4.3.5 applies, see Lawley and Maxwell (1971, Section 2.3). By Lee and Bentler (1980) the degrees of freedom in (4.3.12) are then increased by the number of additional constraints. For better interpretation, the factors can be rotated after estimation, e.g. with the *varimax* method, for details see Anderson (2003, chapter 14).

(iii) With the correction for uniqueness in (ii) above, the factor model of Example 4.3.3 satisfies the regularity conditions of Proposition 4.3.5 and Theorem 4.3.7, see Steiger et al. (1985, Section 4) and Browne (1984, Section 5).

(iv) In case of the copula factor model (see Remark 4.3.2(iii)) we only need to estimate the loading matrix $\mathbf{L} \in \mathbb{R}^{d\times m}$, since $\mathrm{diag}(\mathbf{V}^2) = \mathbf{1} - \mathrm{diag}(\mathbf{L}\mathbf{L}^T)$. Therefore the number of free parameters are dm minus the number of the additional constraints to ensure that $\mathbf{L}^T\mathbf{V}^{-2}\mathbf{L}$ is diagonal, i.e. the degrees of freedom of the limiting χ^2 distribution are $df = d(d-1)/2 - dm + m(m-1)/2$.

(v) For the quadratic form discrepancy function $D(\cdot,\cdot|\widehat{\mathbf{\Gamma}})$, where $\widehat{\mathbf{\Gamma}}$ is a consistent estimator of $\mathbf{\Gamma}$, it can be shown that $T^{(i)}$, $1 \le i < j$, has an approximate noncentral χ^2_{df}-distribution with non-centrality parameter $nD(\boldsymbol{r}_0, \boldsymbol{r}^{(i)}(\boldsymbol{\vartheta}_0^{(i)})|\mathbf{\Gamma})$, see Browne (1984, Corollary 4.1).

4.4 Methodology

As we consider a copula structure model, we need an estimator $\widehat{\mathbf{R}}$ of the copula correlation matrix $\mathbf{R}$, whose limit distribution is $\mathcal{N}(\mathbf{0},\mathbf{\Gamma})$ for some non-degenerate covariance matrix $\mathbf{\Gamma}$ and a consistent estimator of $\mathbf{\Gamma}$. In the following we will introduce two copula based dependence concepts and their corresponding correlation and asymptotic covariance estimators (which are also consistent and asymptotically normal).

4.4.1 Dependence Concepts

A well known dependence concept is (linear) correlation or covariance, which is limited by the fact that it measures only linear dependence. Further, since correlation is not invariant under non-linear (strictly increasing) transformations, it is not a copula property. As we want for our copula structure analysis a dependence concept which is at least related to correlation we use the following one known as *Kendall's tau.*

This copula-based dependence concept provides a good alternative to the linear correlation as a measure also for non-elliptical distributions, for which linear correlation is an inappropriate measure of dependence and often misleading. Originally, it has been suggested as a robust dependence measure, which makes it also appropriate for heavy-tailed data; for more details see Kendall and Gibbons (1990).

Definition 4.4.1. *Kendall's tau τ_{ij} between two components (X_i, X_j) of a random vector $\boldsymbol{X}$ is defined as*

$$\tau_{ij} := P\left((X_i - \widetilde{X}_i)(X_j - \widetilde{X}_j) > 0\right) - P\left((X_i - \widetilde{X}_i)(X_j - \widetilde{X}_j) < 0\right),$$

where $(\widetilde{X}_i, \widetilde{X}_j)$ is an independent copy of (X_i, X_j). Moreover, we call $\mathbf{T} := (\tau_{ij})_{1\le i,j\le d}$ the Kendall's tau matrix.

Concerning elliptical copulae the following result is given in Lindskog, McNeil, and Schmock (, Theorem 2).

Theorem 4.4.2. *Let $\boldsymbol{X}$ be a vector of random variables with elliptical copula $C \sim \mathcal{EC}_d(\mathbf{R}, G)$ and continuous generating variable $G > 0$, then $\tau_{ij} = 2\arcsin(\rho_{ij})/\pi$.*

Considering extreme observations, we need the concept of regular variation. A textbook treatment of this topic is to be found in Bingham, Goldie, and Teugels (1989), for a multivariate extension we refer to Resnick (1987, 2004) or Basrak, Davis, and Mikosch (2002).

Definition 4.4.3. *A random variable G is called* regularly varying at infinity with index $-\alpha$*, $0 < \alpha < \infty$, if $\lim_{x\to\infty} P(G > tx)/P(G > x) = t^{-\alpha}$, for all $t > 0$. We write $G \in RV_{-\alpha}$.*

In financial risk management, one is often interested only in the dependence of extreme observations. By Sklar's theorem, the copula is sufficient to describe dependence in extremes. As C is a uniform distribution on $[0,1]^d$, extreme values happen near the boundaries and extreme dependence happens around the points $(0,\dots,0)$ and $(1,\dots,1)$. This can be captured by the following concept.

Definition 4.4.4. *(i) We define the* upper tail copula *of $\boldsymbol{X}$ as*

$$\begin{aligned}\lambda^{\boldsymbol{X}}_{\text{upper}}(\boldsymbol{x}) &= \lambda^{\boldsymbol{X}}_{\text{upper}}(x_1,\dots,x_d) \\ &= \lim_{t\to 0} t^{-1} P\left(1 - F_1(X_1) \le tx_1, \dots,\ 1 - F_d(X_d) \le tx_d\right),\end{aligned} \tag{4.4.1}$$

for $x_1,\dots,x_d \ge 0$ if the limit exists.

(ii) We define the lower tail copula *of $\boldsymbol{X}$ as*

$$\lambda^{\boldsymbol{X}}_{\text{lower}}(\boldsymbol{x}) := \lim_{t\to 0} t^{-1} P\left(F_1(X_1) \le tx_1, \dots, F_d(X_d) \le tx_d\right). \tag{4.4.2}$$

for $x_1,\dots,x_d \ge 0$ if the limit exists.

Remark 4.4.5. Since by symmetry $\lambda^{\boldsymbol{X}}_{\text{lower}}(\boldsymbol{x}) = \lambda^{\boldsymbol{X}}_{\text{upper}}(\boldsymbol{x}) =: \lambda^{\boldsymbol{X}}(\boldsymbol{x})$ holds for elliptical copulae (see Definitions 4.2.3 and 4.2.5), we concentrate only on the upper tail copula and call it *tail copula*. Of course, by definition, the tail copula is a copula property. For more details about the tail copula, see Schmidt and Stadtmüller (2006).

Notions like tail copula or tail dependence function go back to Gumbel (1960), Pickands (1981) and Galambos (1987), and they represent the full dependence structure of the model in the extremes. If $\lambda^{\boldsymbol{X}}(\boldsymbol{x}) > 0$ for some $\boldsymbol{x} > \mathbf{0}$, $\boldsymbol{X}$ is called *asymptotically dependent* and *asymptotically independent*, otherwise. Assuming elliptical copulae, Hult and Lindskog (2002, Theorem 4.3) show that $\boldsymbol{X}$ is asymptotically dependent if $\boldsymbol{X}$ has an elliptical copula with regularly varying generating variable $G \in RV_{-\alpha}$, $\alpha > 0$. For a textbook treatment of multivariate extremes, see Resnick (1987).

By definition, $\lambda^{\boldsymbol{X}}(\boldsymbol{x}) = 0$ if $\lambda^{(X_i,X_j)}(x_i, x_j) = 0$ for some i, j and $\boldsymbol{x} > \mathbf{0}$, i.e. $\boldsymbol{X}$ is asymptotically independent if some bivariate margin (X_i, X_j) of $\boldsymbol{X}$ is asymptotically independent. Concerning asymptotic independence we refer to Ledford and Tawn (1996, 1997), and for a conditional modeling and estimation approach allowing for asymptotic independence in some components and asymptotic dependence in others, see Heffernan and Tawn (2004). We will use the assumption of asymptotic dependence for modeling and estimation and therefore we omit further discussions about asymptotic independence.

For estimation of $\mathbf{R}$ we only need a representation of the bivariate marginal tail copula (4.4.1) for elliptical copulae. It follows from Hult and Lindskog (2002, Corollary 3.1), Klüppelberg, Kuhn, and Peng (2005a, Theorem 2.1) and transformation of variable. A representation of the full multivariate version is given in Klüppelberg, Kuhn, and Peng (2005b, Theorem 5.1).

Theorem 4.4.6. *Suppose $\boldsymbol{X}$ has copula $C_{\boldsymbol{X}} \in \mathcal{EC}_d(\mathbf{R}, G)$ with generating variable $G \in RV_{-\alpha}$, $\alpha > 0$, and copula correlation matrix $\mathbf{R} = (\rho_{ij})_{1\le i,j\le d}$ with $\max|\rho_{ij}| < 1$. Then the* bivariate marginal tail copula *of $\boldsymbol{X}$ is given by*

$$\begin{aligned}
\lambda^{\boldsymbol{X}}_{ij}(x, y) &:= \lambda^{\boldsymbol{X}}(\infty, \ldots, \infty, x, \infty, \ldots, \infty, y, \infty, \ldots, \infty) \\
&= \left(x \int_{g_{ij}\left((x/y)^{1/\alpha}\right)}^{\pi/2} (\cos\phi)^{\alpha}\, \mathrm{d}\phi + y \int_{g_{ij}\left((x/y)^{-1/\alpha}\right)}^{\pi/2} (\cos\phi)^{\alpha}\, \mathrm{d}\phi \right) \left(\int_{-\pi/2}^{\pi/2} (\cos\phi)^{\alpha}\, d\phi \right)^{-1} \\
&=: \lambda(x, y, \alpha, \rho_{ij}),
\end{aligned} \tag{4.4.3}$$

where x is the i-th, y the j-th component and $g_{ij}(t) := \arctan\left((t - \rho_{ij})/\sqrt{1 - \rho_{ij}^2}\right)$.

Remark 4.4.7. The case of $\rho := \rho_{ij} = 1$ can be interpreted as a limit, i.e.

$$\lambda(x, y, \alpha, 1) \quad := \quad \lim_{\rho \to 1} \lambda(x, y, \alpha, \rho).$$

Then

$$g_{ij}(t) = \lim_{\rho \to 1} \arctan\left((t - \rho_{ij})/\sqrt{1 - \rho_{ij}^2}\right) \quad = \quad \begin{cases} +\pi/2, & t > 1, \\ 0, & t = 1, \\ -\pi/2, & t < 1, \end{cases}$$

and we obtain $\lambda(x, y, \alpha, 1) = x \wedge y$. Similarly, $\lambda(x, y, \alpha, -1) = 0$.

This bivariate marginal tail copula $\lambda_{ij}^{\boldsymbol{X}}$ given in (4.4.3) measures the amount of dependence in the upper right quadrant of (X_i, X_j). Note that by Klüppelberg et al. (2005b, Theorem 5.1), $\lambda^{\boldsymbol{X}}$ is completely characterized by the copula correlation matrix $\mathbf{R}$ and the index α of regular variation of G.

By Theorems 4.4.2 and 4.4.6 we see that for an elliptical copula the correlation matrix $\mathbf{R}$ is a function of Kendall's tau or of the tail copula with the index α of regular variation of G. In Sections 4.4.2 and 4.4.3 we will invoke this functional relationship for the estimation of $\mathbf{R}$. The two approaches differ in their interpretation: estimating $\mathbf{R}$ via Kendall's tau fits a robust dependence structure of the data to an elliptical copula. Using the tail copula for estimation of $\mathbf{R}$ fits only the dependence structure in the upper extremes to an elliptical copula and does not necessarily fit the dependence of the data in other regions. Of course, copula structure analysis can be applied to any copula correlation estimator with a certain limiting behavior as given by Theorem 4.3.7. Using Kendall's tau for estimation can then be seen as a robust extension of the usual correlation structure analysis, whereas using the tail copula provides a structure analysis of dependence in the extremes. The next two sections explain the estimation procedures and give asymptotic results.

4.4.2 Copula correlation estimator based on Kendall's tau

The first method is based on Kendall's tau, which can be used for estimating the correlation matrix $\mathbf{R}$ by Theorem 4.4.2. For a general treatment of U-statistics see Lee (1990); the results we use go back to Hoeffding (1948).

Definition 4.4.8. *Given an iid sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^T$, we define the estimator $\widehat{\mathbf{T}} = (\widehat{\tau}_{ij})_{1 \le i,j \le d}$ of Kendall's tau matrix $\mathbf{T}$ by $\widehat{\tau}_{ii} = 1$ for $i = 1, \ldots, d$ and*

$$\widehat{\tau}_{ij} \quad = \quad \binom{n}{2}^{-1} \sum_{1 \le l < k \le n} \operatorname{sign}\left((X_{k,i} - X_{l,i})(X_{k,j} - X_{l,j})\right), \quad 1 \le i \ne j \le d.$$

Estimating the copula correlation matrix via Kendall's tau yields the following result. Its proof can be found in Section 4.6.

Theorem 4.4.9. *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be an iid sequence of d-dimensional random vectors with elliptical copula $C_{\boldsymbol{X}} \in \mathcal{EC}_d(\mathbf{R}, G)$ with continuous G. Further, define*

$$\widehat{\mathbf{R}}_\tau = (\widehat{\rho}^\tau_{ij})_{1\le i,j\le d} \quad := \quad \sin\left(\frac{\pi}{2}\widehat{\mathbf{T}}\right), \tag{4.4.4}$$

where the 'sin' *is used componentwise and define $\widehat{\boldsymbol{r}}_\tau := \mathrm{vec}[\widehat{\mathbf{R}}_\tau]$ and $\boldsymbol{r} := \mathrm{vec}[\mathbf{R}]$. Then,*

$$\sqrt{n}\,(\widehat{\boldsymbol{r}}_\tau - \boldsymbol{r}) \quad \xrightarrow{d} \quad \mathcal{N}_{d(d-1)/2}(\mathbf{0}, \boldsymbol{\Gamma}_\tau)\,, \quad n \to \infty\,,$$

holds, where $\boldsymbol{\Gamma}_\tau = (\gamma^\tau_{ij,kl})_{1\le i\ne j, k\ne l\le d}$ with

$$\begin{aligned} \gamma^\tau_{ij,kl} &= \pi^2 \cos\left(\frac{\pi}{2}\tau_{ij}\right)\cos\left(\frac{\pi}{2}\tau_{kl}\right)(\tau_{ij,kl} - \tau_{ij}\tau_{kl}) \quad \text{and} \qquad (4.4.5)\\ \tau_{ij,kl} &= E\Big(E\left(\mathrm{sign}\left[(X_{1,i}-X_{2,i})(X_{1,j}-X_{2,j})\right] \,\big|\boldsymbol{X}_1\right) E\left(\mathrm{sign}\left[(X_{1,k}-X_{2,k})(X_{1,l}-X_{2,l})\right] \,\big|\boldsymbol{X}_1\right)\Big). \end{aligned}$$

By (4.4.5), an estimator of $\boldsymbol{\Gamma}_\tau = (\gamma^\tau_{ij,kl})_{1\le i\ne j,k\ne l\le d}$ can be defined by its empirical version.

Definition 4.4.10. *Given an iid sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^T$, we define the estimator $\widehat{\boldsymbol{\Gamma}}_\tau = (\widehat{\gamma}^\tau_{ij,kl})_{1\le i\ne j,k\ne l\le d}$, where*

$$\begin{aligned} \widehat{\gamma}^\tau_{ij,kl} &:= \pi^2 \cos\left(\frac{\pi}{2}\widehat{\tau}_{ij}\right)\cos\left(\frac{\pi}{2}\widehat{\tau}_{kl}\right)(\widehat{\tau}_{ij,kl} - \widehat{\tau}_{ij}\widehat{\tau}_{kl}) \quad \text{and} \qquad (4.4.6)\\ \widehat{\tau}_{ij,kl} &:= \frac{1}{n(n-1)^2}\sum_{p=1}^n \left[\left(\sum_{q=1,\,q\ne p}^n \mathrm{sign}\left((X_{p,i}-X_{q,i})(X_{p,j}-X_{q,j})\right)\right) \times \right.\\ &\qquad \left. \times \left(\sum_{q=1,\,q\ne p}^n \mathrm{sign}\left((X_{p,k}-X_{q,k})(X_{p,l}-X_{q,l})\right)\right)\right]. \qquad (4.4.7) \end{aligned}$$

The following result is also proved in Section 4.6.

Theorem 4.4.11. *The estimator $\mathrm{vec}[\widehat{\boldsymbol{\Gamma}}_\tau]$ is consistent and asymptotically normal.*

4.4.3 Copula correlation estimator based on the tail copula

The second estimation method is based on the tail copula. If someone is interested in the dependence structure of the extreme data and assumes an elliptical copula, (4.4.1) shows how $\lambda^{\boldsymbol{X}}$ can be expressed as a function of $\mathbf{R}$ and α. By estimation of $\lambda^{\boldsymbol{X}}$ one can estimate

$\mathbf{R}$ and α (i.e. the elliptical structure), which is likely to generate the observed extreme dependence.

We use an approach closely related to Klüppelberg et al. (2005b); i.e. we use the tail copula for the estimation of $\mathbf{R}$ and α. By Theorem 4.4.6 we need an estimator of α and of all bivariate marginal tail copulae. We start with an empirical tail copula estimator, for details see Klüppelberg et al. (2005a, 2005b) (and references therein) and estimate $\mathbf{R}$ and α from this.

Definition 4.4.12. *Given an iid sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $\boldsymbol{X}_l = (X_{l,1}, \ldots, X_{l,d})^T$, we define the empirical tail copula estimator for $\boldsymbol{x} = (x_1, \ldots, x_d) > \mathbf{0}$ as*

$$\widehat{\lambda}^{\mathrm{emp}}(\boldsymbol{x}; k) \;=\; \frac{1}{k}\sum_{l=1}^{n} I\left(1 - \widehat{F}_j(X_{lj}) \leq \frac{k}{n}x_j,\; j = 1, \ldots, d\right), \qquad (4.4.8)$$

where $1 \leq k \leq n$ and $\widehat{F}_j$ denotes the empirical distribution function of $\{X_{l,j}\}_{l=1}^n$, $1 \leq j \leq d$. Further, we define the empirical estimator of the bivariate marginal tail copula as

$$\begin{aligned}\widehat{\lambda}_{ij}^{\mathrm{emp}}(x, y; k) \;&:=\; \widehat{\lambda}^{\mathrm{emp}}(\infty, \ldots, \infty, x, \infty, \ldots, \infty, y, \infty, \ldots, \infty)\\ &:=\; \frac{1}{k}\sum_{l=1}^{n} I\left(1 - \widehat{F}_i(X_{li}) \leq \frac{k}{n}x\,, 1 - \widehat{F}_j(X_{lj}) \leq \frac{k}{n}y\right), \qquad (4.4.9)\end{aligned}$$

where x is at the i-th and y at the j-th component, respectively.

Since $\widehat{\lambda}^{\mathrm{emp}}$ estimates the tail copula, the number k should be small in comparison to n. Setting $x_j = 1$, $1 \leq j \leq d$, only the k largest observations of $X_{l,j}$ satisfy $1 - \widehat{F}_j(X_{lj}) \leq k/n$, therefore k can be interpreted as the number of the largest order statistics which are used for the estimation as is typical in extreme value theory.

Immediately by definition (4.4.1), $\lambda^{\boldsymbol{X}}$ is homogenous of order 1, and, for large k and n, also $\widehat{\lambda}_{ij}^{\mathrm{emp}}$ is (see (4.4.8)). Consequently, setting $\theta = \arctan(y/x)$, i.e. $(x, y) = (c\cos\theta, c\sin\theta)$ for some constant $c > 0$, we have $\widehat{\lambda}_{ij}^{\mathrm{emp}}(x, y;\, k) = \widehat{\lambda}_{ij}^{\mathrm{emp}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k^*)$ for some appropriate k^*. Hence, for the estimation, we follow the convention and only consider points $(x, y) = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta)$, $\theta \in (0, \pi/2)$.

For estimation of α we use the approach of Klüppelberg et al. (2005b), which is based on inversion of the tail copula with respect to α.

Definition 4.4.13. *Define $\lambda^{\leftarrow\alpha}(\cdot; x, y, \rho)$ as the inverse of $\lambda(x, y, \alpha, \rho)$ (given in* (4.4.1)*)* *with respect to* α *and, using* $\widehat{\rho}_{ij}^{\tau}$ *given in* (4.4.4) *and* $\widehat{\lambda}^{\text{emp}}$ *given in* (4.4.9)*, define for* $i \neq j$

$$\begin{aligned}\widehat{Q}_{ij} &:= \Big\{\theta \in \Big(0, \frac{\pi}{2}\Big) : \widehat{\lambda}_{ij}^{\text{emp}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k) < \\ &\qquad < \lambda\left(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \left|\frac{\ln(\tan\theta)}{\ln(\widehat{\rho}_{ij}^{\tau} \vee 0)}\right|, \widehat{\rho}_{ij}^{\tau}\right)\Big\}, \\ \widehat{Q}_{ij}^{*} &:= \Big\{\theta \in \Big(0, \frac{\pi}{2}\Big) : |\ln(\tan\theta)| < \big(1 - k^{-1/4}\big)\, \widetilde{\alpha}_{ij}(1,1;k)\, \big|\ln(\widehat{\rho}_{ij}^{\tau} \vee 0)\big|\Big\} \quad \text{and} \\ Q_{ij}^{*} &:= \Big\{\theta \in \Big(0, \frac{\pi}{2}\Big) : |\ln(\tan\theta)| < \alpha\, |\ln(\rho_{ij} \vee 0)|\Big\},\end{aligned}$$

where for $\theta \in \widehat{Q}_{ij}$ *we define* $\widetilde{\alpha}_{ij}$ *as the estimator of* α *based on the empirical bivariate tail copula* (4.4.9)

$$\begin{aligned}&\widetilde{\alpha}_{ij}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k) \\ &\quad := \lambda^{\leftarrow\alpha}\left(\widehat{\lambda}_{ij}^{\text{emp}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k); \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\rho}_{ij}^{\tau}\right). \qquad (4.4.10)\end{aligned}$$

Further, let w *be a nonnegative weight function. Then we define the smoothed estimator* $\widehat{\alpha}$ *of* α *as*

$$\widehat{\alpha}(k,w) := \frac{2}{d(d-1)} \sum_{1 \le i < j \le d} \frac{1}{W\big(\widehat{Q}_{ij} \cap \widehat{Q}_{ij}^{*}\big)} \int_{\theta \in \widehat{Q}_{ij} \cap \widehat{Q}_{ij}^{*}} \widetilde{\alpha}_{ij}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k)\, W(d\theta), \qquad (4.4.11)$$

where W *is the measure induced by* w.

To define an estimator of $\mathbf{R}$ via extreme observations, we invert the bivariate tail copula with respect to ρ. Using (4.4.3) it is straightforward to show the following.

Lemma 4.4.14. *For fixed* x, y, $\alpha > 0$ *define* $\rho^* := ((x \wedge y)/(x \vee y))^{1/\alpha}$. *Then, for all* $\rho < \rho^*$, $\frac{\partial}{\partial\rho}\lambda(x,y,\alpha,\rho) > 0$ *holds and the inverse* $\lambda^{\leftarrow\rho}(\cdot;\, x, y, \alpha)$ *of* λ *with respect to* ρ *exists.*

By Remark 4.4.7, $\lambda(1,1,\alpha,1) = 1$ and $\lambda(1,1,\alpha,-1) = 0$ for $\alpha > 0$. Hence, we can define

$$\widetilde{\rho}_{ij}(1,1;\, k) \;:=\; \lambda^{\leftarrow\rho}\left(\widehat{\lambda}_{ij}^{\text{emp}}(1,1;\, k);\, 1, 1, \widehat{\alpha}(k,w)\right). \qquad (4.4.12)$$

Since this estimator only employs information at $(x,y) = (1,1)$, it may not be very efficient. Therefore, we define an estimator based on $\widehat{\lambda}_{ij}^{\text{emp}}(x,y;\, k)$ for other values $(x,y) = (\sqrt{2}\cos\theta\,,\ \sqrt{2}\sin\theta) \in \mathbb{R}_+^2$.

To ensure existence and consistency of the estimator, we define the following sets and give some explanations below:

$$
\begin{aligned}
\widehat{U}_{ij} &:= \Big\{\theta \in \Big(0, \frac{\pi}{2}\Big) : \widehat{\lambda}^{\mathrm{emp}}_{ij}\Big(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta;\, k\Big) < \\
&\qquad\qquad < \lambda\Big(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\alpha}(k,w), e^{-|\ln(\tan\theta)|/\widehat{\alpha}(k,w)}\Big)\Big\}, \\
\widehat{U}^*_{ij} &:= \Big\{\theta \in \Big(0, \frac{\pi}{2}\Big) : |\ln(\tan\theta)| < (1-k^{-1/4})\widehat{\alpha}(k,w)\, \Big|\ln\Big(\widetilde{\rho}_{ij}(1,1;\, k) \vee 0\Big)\Big|\Big\} \text{ and} \\
U^*_{ij} &:= \Big\{\theta \in \Big(0, \frac{\pi}{2}\Big) : |\ln(\tan\theta)| < \alpha\, |\ln(\rho_{ij} \vee 0)|\Big\}.
\end{aligned}
$$

By Lemma 4.4.14 there exists a unique ρ such that

$$
\lambda\Big(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\alpha}(k,w), \rho\Big) = \widehat{\lambda}^{\mathrm{emp}}_{ij}\Big(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta;\, k\Big), \quad \theta \in \widehat{U}_{ij}.
$$

Hence, we can define

$$
\begin{aligned}
&\widetilde{\rho}_{ij}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta;\, k) \qquad (4.4.13)\\
&:= \lambda^{\leftarrow\rho}_{ij}\Big(\widehat{\lambda}^{\mathrm{emp}}_{ij}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta;\, k);\, \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \widehat{\alpha}(k,w)\Big), \quad \theta \in \widehat{U}_{ij}.
\end{aligned}
$$

Note that, by the definition of $\widetilde{\rho}_{ij}(1,1;\, k)$ in (4.4.12), it always holds that $\pi/4 \in \widehat{U}_{ij}$ provided that $\widehat{\lambda}^{\mathrm{emp}}_{ij}(1,1;\, k) < 1$. Hence, if $\widehat{U}_{ij} = \emptyset$, we can replace it by $\widehat{U}_{ij} := \{\pi/4\}$ and also replace $\widehat{U}^*_{ij} := \{\pi/4\}$. To ensure consistency we further require $\theta \in \widehat{U}^*_{ij}$. This implies that the true ρ_{ij} is smaller than $e^{-|\ln(\tan\theta)|/\widehat{\alpha}(k,w)}$ with probability tending to one. The set U^*_{ij} is then the true set of $\theta \in (0, \pi/2)$, where Lemma 4.4.14 applies.

Now we can define an estimator for ρ_{ij} as a smooth version of $\widetilde{\rho}_{ij}$:

Definition 4.4.15. *Let w^* be a nonnegative weight function and W^* be the measure induced by w^*. Then we define for $i \neq j$ and with* (4.4.13)

$$
\widehat{\rho}^{\lambda}_{ij}(k, w^*) := \frac{1}{W^*\big(\widehat{U}_{ij} \cap \widehat{U}^*_{ij}\big)} \int_{\theta \in \widehat{U}_{ij} \cap \widehat{U}^*_{ij}} \widetilde{\rho}_{ij}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta; k)\, W^*(d\theta). \quad (4.4.14)
$$

Further, define $\widehat{\rho}^{\lambda}_{ii}(k, w^*) := 1$, $1 \leq i \leq d$, and $\widehat{\mathbf{R}}_{\lambda}(k, w^*) := \big(\widehat{\rho}^{\lambda}_{ij}(k, w^*)\big)_{1 \leq i,j \leq d}$.

The next theorem shows the asymptotic properties of $\widehat{\mathbf{R}}_{\lambda}(k, w^*)$. We use the theory developed in Schmidt and Stadtmüller (2006) and give a formal proof in Section 4.6.

Theorem 4.4.16. *Suppose the following regularity conditions hold:*

(C1) *$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are iid with copula $C_{\boldsymbol{X}} \in \mathcal{EC}_d(\mathbf{R}, G)$, $G \in RV_{-\alpha}$ for $\alpha > 0$ and $\max_{i \neq j} |\rho_{ij}| < 1$.*

(C2) *There exists $A(t) \to 0$ such that for $i \neq j$*

$$\lim_{t\to 0} \frac{t^{-1}P(1-F_i(X_i) \leq tx, 1-F_j(X_j) \leq ty) - \lambda(x,y,\alpha,\rho_{ij})}{A(t)} = b_{(C2)_{ij}}(x,y)$$

uniformly on $\mathcal{S}_2 := \{\boldsymbol{s} \in \mathbb{R}^2 : \boldsymbol{s}^T\boldsymbol{s} = 1\}$, where $b_{(C2)_{ij}}(x,y)$ is some non-constant function.

(C3) $k = k(n) \to \infty$, $k/n \to 0$ *and* $\sqrt{k}A(k/n) \to 0$ *as* $n \to \infty$.

Let w^* be a nonnegative weight function with $\sup_{\theta \in U^*_{ij}} w^*(\theta) < \infty$ for all $i \neq j$, λ'^{ρ} and λ'^{α} denotes the derivative of λ with respect to ρ and α, respectively, and $(\lambda^{\leftarrow\rho})'^{\alpha}$ denotes the derivative of $\lambda^{\leftarrow\rho}$ with respect to α. Define

$$\begin{aligned}\widetilde{B}_{ij}(x,y) &:= B_{ij}(x,y) - B_{ij}(x,\infty)\frac{\partial}{\partial x}\lambda_{ij}(x,y) - B_{ij}(\infty,y)\frac{\partial}{\partial y}\lambda_{ij}(x,y), \quad (4.4.15)\\ B_{ij}(x,y) &:= B(\infty,\dots,\infty,x,\infty,\dots,\infty,y,\infty,\dots,\infty),\end{aligned}$$

where x is the i-th, y the j-th component and B is a centered tight continuous Gaussian random field on $\overline{\mathbb{R}}^d$ with covariance structure

$$E(B(\boldsymbol{x})B(\boldsymbol{y})) = \lambda^{\boldsymbol{X}}(\boldsymbol{x} \wedge \boldsymbol{y}), \quad \boldsymbol{x},\boldsymbol{y} \in [0,\infty]^d, \qquad (4.4.16)$$

where $\boldsymbol{x} \wedge \boldsymbol{y}$ is taken componentwise. Set as before $\boldsymbol{r} := \mathrm{vec}[\mathbf{R}]$ and $\widehat{\boldsymbol{r}}_\lambda(k,w^*) := \mathrm{vec}[\widehat{\mathbf{R}}_\lambda(k,w^*)]$, then

$$\sqrt{k}\,(\widehat{\boldsymbol{r}}_\lambda(k,w^*) - \boldsymbol{r}) \xrightarrow{d} \mathcal{N}_{d(d-1)/2}(\boldsymbol{0},\boldsymbol{\Gamma}_\lambda)\,, \quad n \to \infty\,,$$

where $\Gamma_\lambda = (\gamma^{\lambda}_{ij,kl})_{1 \leq i \neq j, k \neq l \leq d}$ with

$$\gamma^{\lambda}_{ij,kl} = \sigma_\alpha + \sigma_{ij,\alpha} + \sigma_{kl,\alpha} + \sigma_{ij,kl}, \qquad (4.4.17)$$

and

$$\begin{aligned}\sigma_\alpha = {} & \frac{2}{d^2(d-1)^2W^*(U^*_{ij})W^*(U^*_{kl})} \qquad (4.4.18)\\ & \times \prod_{J\in\{ij,kl\}} \int_{\theta\in U^*_J} (\lambda^{\leftarrow\rho})'^{\alpha}\left(\lambda^{\boldsymbol{X}}_J(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta),\sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\alpha\right)W^*(d\theta)\\ & \times \Bigg(\sum_{1\leq p<q,r<s\leq d} \frac{1}{W(Q^*_{pq})W(Q^*_{rs})}\\ & \times \int_{\theta_1\in Q^*_{pq}}\int_{\theta_2\in Q^*_{rs}} \frac{E\left(\widetilde{B}_{pq}(\sqrt{2}\cos\theta_1,\sqrt{2}\sin\theta_1)\widetilde{B}_{rs}(\sqrt{2}\cos\theta_2,\sqrt{2}\sin\theta_2)\right)}{\lambda'^{\alpha}(\cos\theta_1,\sin\theta_1,\alpha,\rho_{pq})\lambda'^{\alpha}(\cos\theta_2,\sin\theta_2,\alpha,\rho_{rs})}W(d\theta_2)W(d\theta_1)\Bigg)\,,\end{aligned}$$

$$\sigma_{ij,\alpha} = \frac{1}{d(d-1)W^*(U^*_{ij})W^*(U^*_{kl})} \sum_{1\le p<q\le d} \frac{1}{W(Q^*_{pq})} \times \tag{4.4.19}$$
$$\times \Bigg(\int_{\theta_1\in U^*_{ij}} \int_{\theta_1\in U^*_{kl}} \int_{\theta_3\in Q^*_{pq}} (\lambda^{\leftarrow\rho})'^{\alpha} \left(\lambda^{\boldsymbol{X}}_{ij}(\sqrt{2}\cos\theta_1, \sqrt{2}\sin\theta_1), \sqrt{2}\cos\theta_1, \sqrt{2}\sin\theta_1, \alpha \right) \times$$
$$\times \frac{E\left(\widetilde{B}_{pq}(\sqrt{2}\cos\theta_3, \sqrt{2}\sin\theta_3)\widetilde{B}_{kl}(\sqrt{2}\cos\theta_2, \sqrt{2}\sin\theta_2)\right)}{\lambda'^{\alpha}(\cos\theta_3, \sin\theta_3, \alpha, \rho_{pq})\lambda'^{\rho}(\cos\theta_2, \sin\theta_2, \alpha, \rho_{kl})} W^*(d\theta_3)W^*(d\theta_2)W(d\theta_1) \Bigg),$$

similarly $\sigma_{kl,\alpha}$ *(by interchanging the indices 'ij' and 'kl'), and*

$$\sigma_{ij,kl} = \frac{1}{2W^*(U^*_{ij})W^*(U^*_{kl})} \tag{4.4.20}$$
$$\times \int_{\theta_1\in U^*_{ij}} \int_{\theta_2\in U^*_{kl}} \frac{E\left(\widetilde{B}_{ij}(\sqrt{2}\cos\theta_1, \sqrt{2}\sin\theta_1)\widetilde{B}_{kl}(\sqrt{2}\cos\theta_2, \sqrt{2}\sin\theta_2)\right)}{\lambda'^{\rho}(\cos\theta_1, \sin\theta_1, \alpha, \rho_{ij})\lambda'^{\rho}(\cos\theta_2, \sin\theta_2, \alpha, \rho_{kl})} W^*(d\theta_2)W^*(d\theta_1).$$

Remark 4.4.17. If condition (C3) in Theorem 4.4.16 is replaced by

(C3') $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A(k/n) \to b_{(C3)} \in (-\infty, \infty)$ as $n \to \infty$,

an asymptotic bias occurs in $\mathrm{vec}[\widehat{\mathbf{R}}_\lambda(k, w^*)]$. Using the delta method it immediately follows that

$$\sqrt{k}\,(\widehat{\boldsymbol{r}}_\lambda(k, w^*) - \boldsymbol{r}) \quad \xrightarrow{d} \quad \mathcal{N}_{d(d-1)/2}(\boldsymbol{b}_\rho + \boldsymbol{b}_\alpha, \boldsymbol{\Gamma}_\lambda),$$

where $\boldsymbol{\Gamma}_\lambda$ is given in (4.4.17), $\boldsymbol{b}_\rho = \mathrm{vec}[(b_{ij,\rho})_{1\le i,j\le d}]$, $\boldsymbol{b}_\alpha = \mathrm{vec}[(b_{ij,\alpha})_{1\le i,j\le d}]$,

$$b_{ij,\rho} = \frac{1}{W(U^*_{ij})} \int_{\theta\in U^*_{ij}} \frac{b_{(C3)}b_{(C2)_{ij}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta)}{\lambda'^{\rho}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \alpha, \rho_{ij})} W^*(d\theta), \ i \ne j, \quad \text{and}$$

$$b_{ij,\alpha} = \frac{1}{W(U^*_{ij})} \int_{\theta\in U^*_{ij}} (\lambda^{\leftarrow\rho})'^{\alpha} \left(\lambda^{\boldsymbol{X}}_{ij}\left(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta\right), \sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \alpha \right) W^*(d\theta) \times$$
$$\times \frac{2}{d(d-1)} \sum_{1\le p<q\le d} \frac{1}{W(Q^*_{pq})} \int_{\theta\in Q^*_{pq}} \frac{b_{(C3)}b_{(C2)_{pq}}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta)}{\lambda'^{\alpha}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta, \alpha, \rho_{pq})} W(d\theta).$$

Using (4.4.17), we can define an estimator of $\boldsymbol{\Gamma}_\lambda$.

Definition 4.4.18. *We define the estimator of* $\boldsymbol{\Gamma}_\lambda = (\gamma^\lambda_{ij,kl})_{1\le i\ne j,k\ne j}$ *by* $\widehat{\boldsymbol{\Gamma}}_\lambda = (\widehat{\gamma}^\lambda_{ij,kl})_{1\le i\ne j,k\ne j}$ *with*

$$\widehat{\gamma}^\lambda_{ij,kl} = \widehat{\sigma}_\alpha + \widehat{\sigma}_{ij,\alpha} + \widehat{\sigma}_{kl,\alpha} + \widehat{\sigma}_{ij,kl}, \tag{4.4.21}$$

the $\widehat{\sigma}$ *are defined in* (4.4.18)–(4.4.20)*, where* α*,* ρ_{ij} *and* ρ_{kl} *are replaced by their estimators* $\widehat{\alpha}(k, w)$*,* $\widehat{\rho}^\lambda_{ij}(k, w^*)$ *and* $\widehat{\rho}^\lambda_{kl}(k, w^*)$*, respectively, the sets* U^* *and* Q^* *are replaced by their* *estimators* $\widehat{U} \cap \widehat{U}^*$ *and* $\widehat{Q} \cap \widehat{Q}^*$*, respectively, and the covariances* $E\left(\widetilde{B}_{ij}(\cdot)\widetilde{B}_{kl}(\cdot)\right)$ *are replaced by their estimators* $\widehat{E}\left(\widetilde{B}_{ij}(\cdot)\widetilde{B}_{kl}(\cdot)\right)$ *using* (4.4.15) *and* (4.4.16) *and estimating* $\lambda^{\boldsymbol{X}}$ *by* $\widehat{\lambda}^{\text{emp}}$.

The asymptotic properties of $\widehat{\lambda}^{\text{emp}}$, $\widehat{\alpha}$, $\widehat{\rho}_{ij}^{\lambda}$ in combination with the delta method yield immediately the following result.

Theorem 4.4.19. *Under the regularity conditions (C1)–(C3), the estimator* $\text{vec}[\widehat{\boldsymbol{\Gamma}}_\lambda]$ *is consistent and asymptotically normal.*

Estimation of dependence in extremes is always a difficult topic, for some methods of estimation of $\lambda_{ij}^{\boldsymbol{X}}(1,1)$ and pitfalls we refer to Frahm, Junker, and Schmidt (2005). The problem of estimating tail dependence lies in its definition as a limit, see (4.4.1). Estimators of the tail dependence are based on a sub-sample using the largest (or smallest) observations. Concerning the optimal choice of the threshold, we refer to Danielsson, de Haan, Peng, and de Vries (2001), Drees and Kaufmann (1998) and to Klüppelberg et al. (2005a, 2005b).

Remark 4.4.20. It may happen that the correlation matrix estimators (4.4.4) or (4.4.14) are not positive definite. In this case we use the approach of Higham (2002), i.e. we replace $\widehat{\mathbf{R}}$ by the (positive definite) correlation matrix $\mathbf{R}^*$ solving

$$\left\|\widehat{\mathbf{R}} - \mathbf{R}^*\right\|_2 \quad = \quad \min\left\{\left\|\widehat{\mathbf{R}} - \mathbf{R}\right\|_2 : \mathbf{R} \text{ is a correlation matrix}\right\},$$

where $\|\mathbf{R}\|_2 = \sum_{i,j} \rho_{ij}^2$ is the Euclidean or *Frobenius* norm of a matrix $\mathbf{R} = (\rho_{ij})_{1\le i,j\le d}$. Let $\mathbf{R}$ have spectral decomposition $\mathbf{R} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$ with $\mathbf{Q}$ orthogonal and $\mathbf{D} = \text{diag}(\kappa_1, \dots, \kappa_d)$. By Higham (2002, Theorem 3.1 and 3.2), $P_U(\mathbf{R}) := \mathbf{R} - \text{diag}(\mathbf{R} - \mathbf{I})$ is the projection of $\mathbf{R}$ to the set of symmetric matrices with diagonal $\mathbf{1}$ and $P_S(\mathbf{R}) := \mathbf{Q}\,\text{diag}(\max(\kappa_i, 0))\,\mathbf{Q}^T$ is the projection of $\mathbf{R}$ to the set of positive definite matrices, respectively. Then, Higham (2002, Algorithm 3.3) calculates $\mathbf{Y}_i$ converging to $\mathbf{R}^*$ with respect to the Frobenius norm as $i \to \infty$:

$$
\begin{aligned}
&\triangle\mathbf{S}_0 - \mathbf{0},\ \mathbf{Y}_0 = \widehat{\mathbf{R}}\\
&\text{for } i = 1, 2, \dots\\
&\qquad \mathbf{Z}_i = \mathbf{Y}_{i-1} - \triangle\mathbf{S}_{i-1}\\
&\qquad \mathbf{X}_i = P_S(\mathbf{Z}_i)\\
&\qquad \triangle\mathbf{S}_i = \mathbf{X}_i - \mathbf{Z}_i\\
&\qquad \mathbf{Y}_i = P_U(\mathbf{X}_i)\\
&\text{end.}
\end{aligned}
$$

Considering covariance matrices, we do not need the projection P_U. Hence, if we observe not positive definite covariance estimators (4.4.6) or (4.4.21), we project them to the set of positive definite matrices by $P_S(\widehat{\mathbf{\Gamma}})$.

4.5 The new methods at work

Using the estimators (4.4.4) and (4.4.6) or (4.4.14) and (4.4.21) together with the quadratic form discrepancy function (4.3.6), we can now apply copula structure analysis. In the following, we consider the copula factor model, i.e. we choose the setting $C_{\boldsymbol{X}} = C_{(\mathbf{L},\mathbf{V})\boldsymbol{\xi}}$, where $\mathbf{L} \in \mathbb{R}^{d\times m}$, $\mathbf{V} \in \mathbb{R}^{d\times d}$ is a diagonal matrix with nonnegative entries and $\boldsymbol{\xi} \in \mathcal{E}_{m+d}(\mathbf{0}, \mathbf{I}, G)$; also see Remark 4.3.2(iii).

As for the test statistic T based on the quadratic form discrepancy function (4.3.6) we first compare in a simulation study T to its limiting χ^2-distribution. Therefore, we define by

$$T_{\text{QD}}^{\tau} \quad := \quad n \min_{\boldsymbol{\vartheta}\in\Theta} D_{\text{QD}}\left(\widehat{\boldsymbol{r}}_\tau, \boldsymbol{r}(\boldsymbol{\vartheta}) | \, \widehat{\mathbf{\Gamma}}_\tau\right)$$

the quadratic form test statistic obtained from the Kendall's tau based estimators $\widehat{\boldsymbol{r}}_\tau = \text{vec}[\widehat{\mathbf{R}}_\tau]$ and $\widehat{\mathbf{\Gamma}}_\tau$ given in (4.4.4) and (4.4.6), respectively.

Similarly,

$$T_{\text{QD}}^{\lambda} \quad := \quad k \min_{\boldsymbol{\vartheta}\in\Theta} D_{\text{QD}}\left(\widehat{\boldsymbol{r}}_\lambda, \boldsymbol{r}(\boldsymbol{\vartheta}) | \, \widehat{\mathbf{\Gamma}}_\lambda\right),$$

where k is the number of the largest order statistics used for estimation, $\widehat{\boldsymbol{r}}_\lambda = \text{vec}[\widehat{\mathbf{R}}_\lambda]$ and $\widehat{\mathbf{\Gamma}}_\lambda$ given in (4.4.14) and (4.4.21), respectively. As a weight function we choose a discrete version of

$$w(\theta) \quad = \quad 1 - \left(\frac{\theta}{\pi/4} - 1\right)^2, \quad 0 \le \theta \le \frac{\pi}{2}, \tag{4.5.1}$$

both for the estimation of α and $\mathbf{R}$ given a copula $C \in \mathcal{EC}(\mathbf{R}, G)$, $G \in RV_{-\alpha}$ and $\alpha > 0$.

We also compare the copula factor model to the classical factor model $\boldsymbol{X} = (\mathbf{L}, \mathbf{V})\boldsymbol{\xi}$, $\boldsymbol{\xi} \in \mathcal{E}_{m+d}(\mathbf{0}, \mathbf{I}, G)$. To this end we define

$$T_{\text{QD}}^{\text{emp}} \quad := \quad n \min_{\boldsymbol{\vartheta}\in\Theta} D_{\text{QD}}\left(\widehat{\boldsymbol{r}}_{\text{emp}}, \boldsymbol{r}(\boldsymbol{\vartheta}) | \, \widehat{\mathbf{\Gamma}}_{\text{emp}}\right),$$

where $\widehat{\boldsymbol{r}}_{\text{emp}} = \text{vec}[\widehat{\mathbf{R}}_{\text{emp}}]$ is the vector of the standard empirical correlation estimator with its asymptotic covariance matrix estimator $\widehat{\mathbf{\Gamma}}_{\text{emp}}$ under normal assumptions, for details see Browne and Shapiro (1986).

The parameter $\boldsymbol{\vartheta}$ is then estimated also in three different ways, denoted by $\widehat{\boldsymbol{\vartheta}}_\tau$, $\widehat{\boldsymbol{\vartheta}}_\lambda$ and $\widehat{\boldsymbol{\vartheta}}_{\text{emp}}$, by minimizing T_{QD}^τ, T_{QD}^λ and $T_{\text{QD}}^{\text{emp}}$, respectively.

Example 4.5.1. [Model selection by χ^2-tests]
To see the performance of the quadratic form test statistics T_{QD}^τ and T_{QD}^λ, we perform a simulation study. We choose a $d = 10$ dimensional setting with $m = 2$ factors and loadings as given in Table 4.1. Then $\mathbf{L}\mathbf{L}^T + \mathbf{V}^2 = \mathbf{R}$ is a correlation matrix.

component	1	2	3	4	5	6	7	8	9	10
$\mathbf{L}_{\cdot,1}$	.9	.9	.9	.9	.9	0	0	0	0	0
$\mathbf{L}_{\cdot,2}$	0	0	0	0	0	.9	.9	.9	.9	.9
$\text{diag}(\mathbf{V}^2)$	.19	.19	.19	.19	.19	.19	.19	.19	.19	.19

Table 4.1: Factor loadings of Example 4.5.1

Define a multivariate t_α-copula as the copula of the random vector $G\boldsymbol{N}$, where $G \sim \sqrt{\alpha/\chi^2_\alpha}$, $\alpha > 0$, is independent of $\boldsymbol{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. Note that the t_α-copula is elliptical and, since $G \in RV_{-\alpha}$, its tail copula satisfies (4.4.3). Choose $\alpha = 3$, then $G\boldsymbol{N}$ has finite second moment, but its fourth moment does not exist. Hence, classical factor analysis cannot be applied to $G\boldsymbol{N}$, see Proposition 4.3.5 and Theorem 4.3.7. Also, if the model with $\alpha < 8$ is considered, which has finite fourth moment but non-existing eight moment, the estimator of $\boldsymbol{\Gamma}$ will only be consistent and large sample sizes may be necessary to observe the limiting χ^2 distribution of the test statistic T. As the test statistics T_{QD}^τ and T_{QD}^λ are based on the copula of the sample, they are not affected by the existence or non-existence of moments.

We simulate 500 iid samples of length $n = 1\,000$ of the t_3-copula, calculate the Kendall's tau based estimators (4.4.4) and (4.4.6) and estimate T_{QD}^τ from these. To ensure uniqueness of the loadings, we use the restriction that $\mathbf{L}^T\mathbf{V}^{-2}\mathbf{L}$ is diagonal, hence we have $m(m-1)/2 = 1$ additional constraints, see Lawley and Maxwell (1971, Section 2.3). Using this restriction and the 2-factor setting, T_{QD}^τ should be (for a large sample) χ^2_{df} distributed with $df = d(d-1)/2 - dm + m(m-1)/2 = 26$ degrees of freedom; see Theorem 4.3.7. Therefore, we compare the 500 estimates of T_{QD}^τ with the χ^2_{26}-distribution by a QQ-plot, see Figure 4.1, left plot. From this plot we see that the distribution of T_{QD}^τ fits the χ^2_{26}-distribution quite well. Similarly, we estimate T_{QD}^λ based on the tail copula estimators (4.4.14) and (4.4.17) with weight function (4.5.1) using the same samples as for T_{QD}^τ and based on the $k = 100$ largest observations; see Figure 4.1, right plot. Also here we observe a reasonable fit to the χ^2_{26}-distribution – not as good as before since the estimators are calculated from a smaller (sub)sample. Note that under the assumption of $m = 1$ factor the corresponding T^{τ}_{QD}'s and $T^{\lambda}_{\mathrm{QD}}$'s were always larger than 600, which clearly rejects the 1-factor hypothesis.

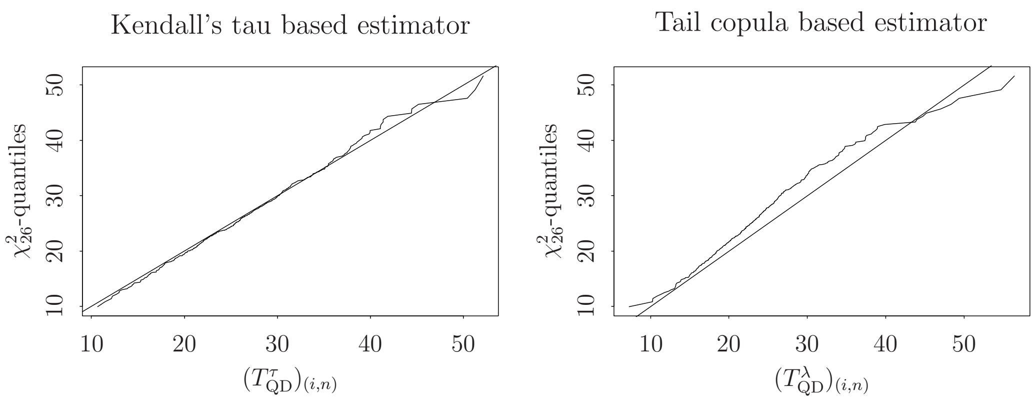

Figure 4.1: *QQ*-plot of ordered estimates $\widehat{T}$ against the χ^2_{26}-quantiles.
Left plot: T^{τ}_{QD} obtained from Kendall's tau based estimators (4.4.4) and (4.4.6).
Right plot: $T^{\lambda}_{\mathrm{QD}}$ obtained from tail copula based estimators (4.4.14) and (4.4.17).

Example 4.5.2. [Oil-currency data]
In this example we consider an 8-dimensional set of data, *(oil, s&p500, gbp, usd, chf, jpy, dkk, sek)*, i.e. we are interested in the dependence structure between the oil-price, the S&P500 index and some currency exchange rates with respect to euro. Each time series consists of 4 904 daily logreturns from May, 1985 to June, 2004. To this data set we fit a copula factor model using the $T^{\mathrm{emp}}_{\mathrm{QD}}$, T^{τ}_{QD} and $T^{\lambda}_{\mathrm{QD}}$ statistics for estimation and model selection. Estimation of $T^{\lambda}_{\mathrm{QD}}$ is based on the $k = 300$ largest observations. The values of these test statistics, based on different numbers of factors are given in Table 4.2. To estimate the number of factors, we use a 95% confidence test, i.e. we reject the

number of factors	*df*	$T^{\mathrm{emp}}_{\mathrm{QD}}$	T^{τ}_{QD}	$T^{\lambda}_{\mathrm{QD}}$	$\chi^2_{df,0.95}$
2	13	298.5	252.7	52.7	22.36
3	7	33.7	17.4	24.0	14.07
4	2	2.3	3.3	0.9	5.99

Table 4.2: Test statistics $T^{\mathrm{emp}}_{\mathrm{QD}}$, T^{τ}_{QD} and $T^{\lambda}_{\mathrm{QD}}$ of oil-currency data under different number of factors.

null hypothesis of having a m-factor model if the test statistic T is larger than the 95%-

quantile of the χ^2_{df}-distribution. This yields 4 factors under the empirical, Kendall's tau based and tail copula based test statistics.

Applying factor analysis based on the different correlation estimates (and their asymptotic covariance estimates) yield different results; see Figure 4.2. The first four plots show the loadings of the four factors, obtained from the empirical correlation estimator, Kendall's tau based and tail copula based estimator. The last plot shows the loadings of the specific factors for all three correlation estimators.

We want to emphasize that, although we have plotted the factors in the same figures, the factors obtained by the three different estimation methods are not known and may have different interpretations. We call them *empirical factors*, *Kendall's tau factors* and *tail copula factors*.

For the first factor all loadings of the different correlation estimators behave very similar with respect to factor 1, which has a weight close to one for usd. Hence, factor one can be interpreted as the *usd-factor*. It also can be seen that this factor has a positive weight for all currencies, but not for the oil-price and s&p500 (almost 0 or very small negative), and the largest dependence is observed for gbp, and jpy.

For factor 2 we observe for all correlation estimators a large weight on Swiss Francs chf, so we call it *chf-factor*. We observe that the empirical and Kendall's tau factor has almost no (or only little) correlation with oil, s&p500, gbp, usd and jpy. The weights on dkk and sek are larger and also moderate for gpd for the tail copula factor indicating that extreme dependence between all European currencies is present.

Considering factor 3, we see for the empirical and Kendall's tau factor a large loading for sek and dkk with only little impact on the other components. If scandinavian currencies were merged, then only a specific factor would remain. The tail copula factor indicates moderate dependence between oil and gbp.

From factor 4 we observe for the empirical factor a loading close to one for the oil-price and loadings close to 0 for the rest of the factors. This indicates that a 3-factor model is sufficient in this case. In combination with the model selection procedure as seen in Table 4.2 this indicates that the distribution of $T_{\mathrm{QD}}^{\mathrm{emp}}$ is far away from a χ^2 distribution. For the Kendall's tau factor there is some dependence between the European currencies and the usd. The tail copula factor behaves different: there is dependence observed between large positive jumps of s&p500 and large negative jumps of the oil price which would not be detected when only considering the other correlation estimators.

Finally, we give an interpretation of the specific factors, where we find the correlation which is not explained through the common factors. For the empirical factor oil is completely explained by factor 4, which is the specific factor for oil, and s&p500 has a loading close to one, showing there is (almost) no correlation to oil and the other currencies. For the Kendall's tau factor, oil and s&p500 are uncorrelated and uncorrelated from the rest. Contrary, for the tail copula factor, oil and s&p500 are not uncorrelated from the common factors. Oil has a rather large specific loading factor, but s&p500 is explained to a large extend by factor 4.

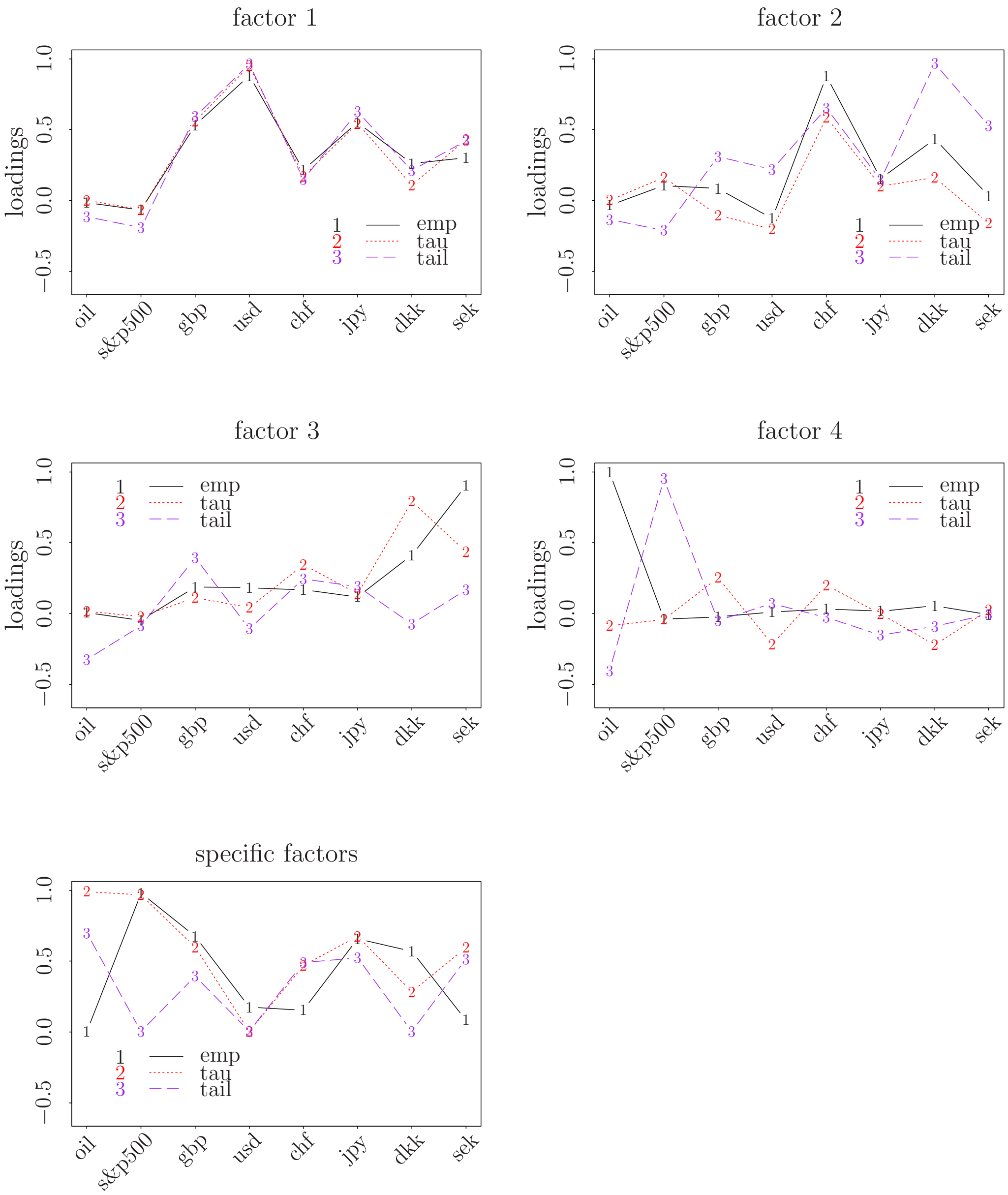

Figure 4.2: Oil-currency data: factor analysis based on 4 factors and different statistics, "emp" for the loadings $\widehat{\boldsymbol{\vartheta}}_{\text{emp}}$, "tau" for $\widehat{\boldsymbol{\vartheta}}_{\tau}$ and "tail" for $\widehat{\boldsymbol{\vartheta}}_{\lambda}$.

Upper row: loadings of factor 1 (left) and 2 (right).

Middle row: loadings of factor 3 (left) and 4 (right).

Lower row: specific factors $\text{diag}(\mathbf{V}^2)$.

4.6 Proofs

Proof of Theorem 4.4.9: Define $\widehat{\boldsymbol{t}} := \text{vec}[\widehat{\mathbf{T}}]$ and $\boldsymbol{t} := \text{vec}[\mathbf{T}]$. Since $\widehat{\boldsymbol{t}}$ is a vector of U-statistics, and, obviously,

$$E\left(\text{sign}\left((X_{1,i} - X_{2,i})(X_{1,j} - X_{2,j})\right)^2\right) \quad < \quad \infty, \quad i \neq j,$$

Lee (1990, Chapter 3, Theorem 2) applies (together with the remark at the end of p.7 therein that all results also hold for random vectors). The covariance structure is stated in Lee (1990, Section 1.4, Theorem 1), hence

$$\sqrt{n}(\widehat{\boldsymbol{t}} - \boldsymbol{t}) \quad \xrightarrow{d} \quad \mathcal{N}_{d(d-1)/2}\left(\mathbf{0}, 4\boldsymbol{\Upsilon}\right), \quad n \to \infty,$$

where $\boldsymbol{\Upsilon} = (\tau_{ij,kl} - \tau_{ij}\tau_{kl})_{1 \leq i \neq j, k \neq l \leq d}$ and $\tau_{ij,kl}$ is given in (4.4.5). Note that the Jacobian matrix $\mathbf{D} := \partial\left(\sin\left(\boldsymbol{t}\pi/2\right)\right)/\partial\boldsymbol{t}$ is a diagonal matrix with

$$\text{diag}(\mathbf{D}) \quad = \quad \frac{\pi}{2}\cos\left(\frac{\pi}{2}\boldsymbol{t}\right).$$

Hence, by the delta method (see Casella and Berger (2001, Section 5.5.4)),

$$\sqrt{n}\left(\widehat{\boldsymbol{r}} - \boldsymbol{r}\right) \quad \xrightarrow{d} \quad \mathcal{N}_{d(d-1)/2}\left(\mathbf{0}, 4\mathbf{D}^T\boldsymbol{\Upsilon}\mathbf{D}\right), \quad n \to \infty,$$

and the proof is complete.

Proof of Theorem 4.4.11: We first consider $\widehat{\tau}_{ij,kl}$ and rewrite it as a linear combination of some U-statistics. Define for $1 \leq a < b < c \leq n$

$$\begin{aligned}
\Phi_2^{ij,kl}(\boldsymbol{x}_a, \boldsymbol{x}_b) &:= \text{sign}\left[(x_{a,i} - x_{b,i})(x_{a,j} - x_{b,j})\right]\text{sign}\left[(x_{a,k} - x_{b,k})(x_{a,l} - x_{b,l})\right] \\
\Phi_{abc}^{ij,kl} &:= \Phi_{abc}^{ij,kl}(\boldsymbol{x}_a, \boldsymbol{x}_b, \boldsymbol{x}_c) \\
&:= \text{sign}\left[(x_{a,i} - x_{b,i})(x_{a,j} - x_{b,j})\right]\text{sign}\left[(x_{a,k} - x_{c,k})(x_{a,l} - x_{c,l})\right] \quad \text{and} \\
\Phi_3^{ij,kl}(\boldsymbol{x}_a, \boldsymbol{x}_b, \boldsymbol{x}_c) &:= \frac{1}{6}\left(\Phi_{abc}^{ij,kl} + \Phi_{acb}^{ij,kl} + \Phi_{bac}^{ij,kl} + \Phi_{bca}^{ij,kl} + \Phi_{cab}^{ij,kl} + \Phi_{cba}^{ij,kl}\right).
\end{aligned}$$

Hence, $\Phi_2^{ij,kl}$ and $\Phi_3^{ij,kl}$ are symmetric in their arguments. Next, define

$$\begin{aligned}
\widehat{u}_2^{ij,kl} &:= \frac{2}{n(n-1)}\sum_{1 \leq a < b \leq n}\Phi_2^{ij,kl}(\boldsymbol{X}_a, \boldsymbol{X}_b) \quad \text{and} \\
\widehat{u}_3^{ij,kl} &:= \frac{6}{n(n-1)(n-2)}\sum_{1 \leq a < b < c \leq n}\Phi_3^{ij,kl}(\boldsymbol{X}_a, \boldsymbol{X}_b, \boldsymbol{X}_c),
\end{aligned}$$

and note that both are U-statistics. Obviously,

$$E\left(\left(\Phi_2^{ij,kl}(\boldsymbol{X}_1,\boldsymbol{X}_2)\right)^2\right) < \infty \quad \text{and} \quad E\left(\left(\Phi_3^{ij,kl}(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)\right)^2\right) < \infty,$$

therefore, by Lee (1990, Chapter 3, Theorem 2), the vector of all $\widehat{u}_2^{ij,kl}$ and $\widehat{u}_2^{ij,kl}$ is consistent and asymptotically normal. Since

$$\widehat{\tau}_{ij,kl} = \frac{1}{n(n-1)^2}\left(\frac{n(n-1)}{2}\widehat{u}_2^{ij,kl} + \frac{n(n-1)(n-2)}{6}\widehat{u}_3^{ij,kl}\right),$$

$\widehat{\tau}_{ij,kl}$ is a linear combination of U-statistics and is therefore also consistent and asymptotically normal. The result then follows using the delta method.

Proof of Theorem 4.4.16: First, by homogeneity,

$$\lambda(\sqrt{2}\cos\theta_2, \sqrt{2}\sin\theta_2, \alpha, \rho) = \sqrt{2}\lambda(\cos\theta_2, \sin\theta_2, \alpha, \rho)$$

holds. Let '$\xrightarrow{w}$' denote weak convergence in the space of all functions $f : \overline{\mathbb{R}}_+^n \to \mathbb{R}$ which are locally uniformly-bounded on every compact subset of $\overline{\mathbb{R}}_+^n$. Next, extending Schmidt and Stadtmüller (2006, Theorem 6) from the bivariate to the d-dimensional setting, we have

$$\sqrt{k}\left(\widehat{\lambda}^{\text{emp}}(\boldsymbol{x};\, k) - \lambda^{\boldsymbol{X}}(\boldsymbol{x})\right) \xrightarrow{w} B(\boldsymbol{x}) - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}\lambda^{\boldsymbol{X}}(\boldsymbol{x})B_i(x_i),$$

where $B_i(x) = B(\infty,\ldots,\infty,x,\infty,\ldots,\infty)$, x is the i-th component and B is a zero mean Wiener process with covariance structure $E(B(\boldsymbol{x})B(\boldsymbol{y})) = \lambda^{\boldsymbol{X}}(\boldsymbol{x}\wedge\boldsymbol{y})$.

To show asymptotic normality we use an extended version of the classical delta-method, for details see van der Vaart and Wellner (1996, p.374). First, note that for all $i \neq j$ and for λ defined in (4.4.3)

$$\inf_{\theta\in Q_{ij}^*} |\lambda'^{\alpha}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta,\, \alpha, \rho_{ij})| > 0,$$
$$\inf_{\theta\in U_{ij}^*} |\lambda'^{\rho}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta,\, \alpha, \rho_{ij})| > 0 \quad \text{and}$$
$$\sup_{\theta\in U_{ij}^*} \left|(\lambda^{\leftarrow\rho})'^{\alpha}(\sqrt{2}\cos\theta, \sqrt{2}\sin\theta,\, \alpha, \rho_{ij})\right| < \infty.$$

Next, define $\mathcal{TC}$ as the set of all d-dimensional tail copulae. By Schmidt and Stadtmüller (2006, Theorem 1(iii)) a tail copula is Lipschitz-continuous, hence $\mathcal{TC}$ is a subset of a topological vector space. Abbreviate for λ defined in (4.4.3) and $\mu \in \mathcal{TC}$ with μ_{ij} being the ij-th marginal of μ

$$\begin{aligned}\widetilde{\alpha}_{ij}(\theta,\mu,\rho) &:= \lambda^{\leftarrow\alpha}\left(\mu_{ij}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta);\sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\rho\right), \quad \text{and}\\ \widetilde{\rho}_{ij}(\theta,\mu,\alpha) &:= \lambda^{\leftarrow\rho}\left(\mu_{ij}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta);\ \sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\alpha\right).\end{aligned}$$

Next, define for some correlation matrix $\mathbf{R} = (\rho_{ij})_{1\le i,j\le d}$

$$\begin{aligned}\alpha(\mu,\mathbf{R}) &:= \frac{1}{d(d-1)}\sum_{i\neq j}\frac{1}{W(Q^*_{ij})}\int_{\theta\in Q^*_{ij}}\widetilde{\alpha}_{ij}(\theta,\mu,\rho_{ij})\,W(d\theta),\\ \rho_{ij}(\mu,\mathbf{R}) &:= \frac{1}{W^*(U^*_{ij})}\int_{\theta\in U^*_{ij}}\widetilde{\rho}_{ij}(\theta,\mu,\alpha(\mu,\mathbf{R}))\,W^*(d\theta), \quad \text{and}\\ \boldsymbol{r}(\mu,\mathbf{R}) &:= \operatorname{vec}\left[(\rho_{ij}(\mu,\mathbf{R}))_{1\le i,j\le d}\right].\end{aligned}$$

Write $\alpha(\mu) := \alpha(\mu,\mathbf{R})$ and note that $\alpha(\mu)$ is Hadamard-differentiable, i.e. let $t_m \overset{m\to\infty}{\longrightarrow} \infty$ and $h_m \overset{m\to\infty}{\longrightarrow} h \in \mathcal{TC}$ such that $\mu + h_m/t_m \in \mathcal{TC}$ for all m. Then, using Taylor expansion,

$$\begin{aligned}&\lim_{m\to\infty} t_m\left(\alpha(\mu+h_m/t_m)-\alpha(\mu)\right)\\ &= \frac{1}{d(d-1)}\sum_{i\neq j}\frac{1}{W(Q^*_{ij})}\int_{\theta\in Q^*_{ij}}\frac{h_{ij}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta)}{\lambda'^{\alpha}\left(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\alpha(\mu),\rho_{ij}\right)}\,W(d\theta)\\ &=: \alpha'_\mu(h),\end{aligned}$$

which obviously is a linear map. Analogously, $\rho_{ij}(\mu) := \rho_{ij}(\mu,\mathbf{R})$ is Hadamard differentiable, i.e.

$$\begin{aligned}&\lim_{m\to\infty} t_m\left(\rho_{ij}(\mu+h_m/t_m)-\rho_{ij}(\mu)\right)\\ &= \frac{1}{W^*(U^*_{ij})}\int_{\theta\in U^*_{ij}}\frac{h_{ij}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta)}{\lambda'^{\rho}\left(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\alpha(\mu),\rho_{ij}\right)}+\\ &\qquad + \alpha'_\mu(h)(\lambda^{\leftarrow\rho})'^{\alpha}\left(\mu_{ij}(\sqrt{2}\cos\theta,\sqrt{2}\sin\theta);\ \sqrt{2}\cos\theta,\sqrt{2}\sin\theta,\alpha(\mu)\right)\,W^*(d\theta)\\ &=: \rho'_{ij;\mu}(h),\end{aligned}$$

and similarly for $\boldsymbol{r}(\mu,\mathbf{R})$. Since $\widehat{\mathbf{R}}_\tau - \mathbf{R} = o_p(1/\sqrt{k})$,

$$\widehat{\rho}_{ij}(k,w^*) = \rho_{ij}\left(\widehat{\lambda}^{\text{emp}}(\cdot;k),\widehat{\mathbf{R}}_\tau\right),$$

and similarly for $\widehat{\boldsymbol{r}}(k,w^*)$, the delta method yields

$$\sqrt{k}\left(\widehat{\boldsymbol{r}}(k,w^*)-\boldsymbol{r}\right) \overset{w}{\longrightarrow} \boldsymbol{r}'_{\lambda\boldsymbol{x}}(\widetilde{B}).$$

The result then follows using

$$E\left(\left(\boldsymbol{r}'_{\lambda}\boldsymbol{x}(\widetilde{B})\right)_{ij}\left(\boldsymbol{r}'_{\lambda}\boldsymbol{x}(\widetilde{B})\right)_{kl}\right) = \sigma_{\alpha}+\sigma_{ij,\alpha}+\sigma_{kl,\alpha}+\sigma_{ij,kl},$$

with σ_{α}, $\sigma_{ij,\alpha}$, $\sigma_{kl,\alpha}$, $\sigma_{ij,kl}$ defined through (4.4.18)–(4.4.20).

Chapter 5

Tails of credit default portfolios

Summary

We derive analytic expressions for the tail behavior of credit losses in a large homogeneous credit default portfolio. Our model is an extended CreditMetrics model; i.e. it is a one-factor model with a multiplicative shock-variable. We show that the first order tail behavior is robust with respect to this shock-variable. In a simulation study we compare different models for the latent variables. We fix default probability and correlation of the latent variables and the first order tail behavior of the limiting credit losses in all models and observe a completely different tail behavior leading to very different VaR estimates. For three portfolios of different credit quality we suggest a pragmatic model selection procedure and compare the fit with that of the β-model.

5.1 Introduction

We consider a homogeneous portfolio $L^{(m)} = \frac{1}{m}\sum_{j=1}^{m} L_j$ of m bonds $L_j \in \{0, 1\}$, where $L_j = 1$ indicates the default of the credit of company j. Each bond is characterized by the vector (S_j, s), where S_j is a latent variable, e.g. the equity value of company j. The number s denotes the default threshold in the sense that the bond of company j defaults, if $S_j < s$.

The credit loss of the portfolio is expressed as the fraction of defaulted bonds and the portfolio is homogeneous in the sense that all bonds have the same characteristics; i.e. the vector $(S_1, S_2, \ldots, S_m)$ follows a factor model

$$S_j := W s^*(X, Y_j), \tag{5.1.1}$$

where $W > 0$, $X \in \mathbb{R}$ and $(Y_j)_{j\in\mathbb{N}}$ is an iid sequence of real random variables. The Y_j are interpreted as a company-specific risk factors, X is a common risk factor (which can be extended to a vector of common factors) and W is a global risk factor and allows for a tuning of the model.

A well-known example for $s^*(\cdot,\cdot)$ is the CreditMetrics model as described in Gupton, Finger, and Bhatia (1997). We consider an *extended CreditMetrics model* given by

$$S_j = W(aX + bY_j), \quad a, b > 0 \text{ and } W > 0,\ X, Y_j \in \mathbb{R} \text{ random.} \tag{5.1.2}$$

The *CreditMetrics model* corresponds to $W = 1$, $X, Y_j \overset{iid}{\sim} \mathcal{N}(0,1)$ and $a = \sqrt{\rho}$, $b = \sqrt{1-\rho}$ for some $\rho \in (0,1)$, modelling the correlation between S_i and S_j for $i \neq j$. One popular extension of this model takes $W = \sqrt{\nu/\chi^2_\nu}$, which yields for $(S_1, \ldots, S_m)$ a multivariate t_ν distribution, called the *multivariate t-model*.

A treatment of different credit portfolio models with a finite number of loans can be found in Frey and McNeil (2000, 2002, 2003) and in Frey, McNeil, and Nyfeler (2001).

For the limiting portfolio $L := \lim_{m\to\infty} L^{(m)}$ it can be shown (see Theorem 5.2.3) that L is a random variable and the limit is in the almost sure sense. For model (5.1.2) with $W \equiv 1$ Lucas, Klaasen, Spreij, and Straetmans (2003) show under weak regularity conditions that the tail behavior of L is *Weibull*-like, i.e. $P(L > q) = (1-q)^\alpha \mathcal{L}(1/(1-q))$, $q \in (0,1)$, for some $\alpha > 0$ and a *slowly varying* function $\mathcal{L}$ (see Definition 5.2.7 for the term Weibull-like and Definition 5.2.6 for the concepts of regular and slow variation).

For a random variable $W > 0$ the result remains true with the same α but a different slowly varying function $\mathcal{L}$ appears. We indicate the influence of W in Section 5.3 by simulation, showing that it has an important influence on the right-tail behavior of L. In Section 5.4 we fit four (extended) CreditMetrics models to three portfolios of different credit quality. We also investigate the fit of a simple β-model. This model, however, proves as being too simplistic in most real world credit portfolios. The extended CreditMetrics model proves to be superior provided the shock-variable W is chosen correctly.

All proofs are gathered in the Appendix.

5.2 Results

First, we give some notations used throughout the chapter.

Notation 5.2.1. *(i) Random variables are always denoted by capital letters.*

(ii) $F_\cdot$ *denotes the distribution function of the random variable '·' and* $f_\cdot$ *denotes its density, e.g.* F_X *and* f_X *are the distribution function and density of* X*, respectively. Further, let* $\overline{F}_\cdot := 1 - F_\cdot$ *denote the tail-distribution of '·'.*

(iii) Let $h = h(x_1, x_2)$ *be a function of two variables. Then* $D_2 h := \partial h / \partial x_2$*.*

(iv) $\mathbf{1}_{\boldsymbol{A}}$ *denotes the indicator function of the set* $\boldsymbol{A}$*.*

(v) We write $a(x) \sim b(x)$ *as* $x \to x_0$*, if* $\lim_{x \to x_0} a(x)/b(x) = 1$*.*

(vi) We write $a(x) = o(1)$ *as* $x \to \infty$*, if* $\lim_{x \to \infty} a(x) = 0$*.* □

We shall investigate the tail-distribution of the limiting portfolio credit loss as defined in the following definition in combination with Theorem 5.2.3

Definition 5.2.2. *Let* $L_j := \mathbf{1}_{\{S_j < s\}} = \mathbf{1}_{\{W s^*(X, Y_j) < s\}}$ *denote the default indicator of the bond of company* j *and define the* portfolio credit loss *by*

$$L^{(m)} \quad := \quad \frac{1}{m} \sum_{j=1}^{m} L_j = \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}_{\{W s^*(X, Y_j) < s\}}.$$

The (almost sure) limit of $L^{(m)}$ *as* $m \to \infty$ *is called* limiting portfolio credit loss *and denoted by* L*.* □

Theorem 5.2.3. *Consider the setting of Definition 5.2.2. Then*

$$\begin{aligned} \lim_{m \to \infty} L^{(m)} &= \lim_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} L_j \\ &\overset{\text{a.s.}}{=} \quad E(L_1 | W, X) = P(S_1 < s | W, X) =: L. \end{aligned}$$

□

Considering the variance of L, we observe the following lemma.

Lemma 5.2.4. *(i) Choose the setting of Definition 5.2.2 with* $p_{\text{loss}} := P(L_j = 1)$*, then* $0 \leq \sqrt{\text{Var}(L)} \leq \sqrt{p_{\text{loss}}(1 - p_{\text{loss}})}$*. The upper bound is obtained for* $L_i \overset{\text{a.s.}}{=} L_j \ \forall i, j$ *and the* $\text{Var}(L) = 0$ *is obtained for* L_i *independent of* $L_j \ \forall i \neq j$*.*

(ii) In the extended CreditMetrics model (5.1.2) *the upper bound is obtained for* $a = 1$ *and* $b = 0$*, and the lower bound is obtained for* $a = 0$*,* $b = 1$ *and* $W \equiv \text{const}$*.*

□

Next, we introduce our key assumptions on the factor model (5.1.1) and the risk factors.

Assumption 5.2.5. *(i) $0 < W \sim F_W$, $X \sim F_X$, $(Y_j)_{j\in\mathbb{N}}$ are iid with $Y_1 \sim F_Y$ and all random variables are independent.*

(ii) Denote by $\mathcal{S}$, $\mathcal{W}$, $\mathcal{X}$ and $\mathcal{Y}$ the supports of S_j, W, X and Y_j, respectively, and let $\mathcal{W} \subseteq (0,\infty)$, $\inf \mathcal{X} = -\infty$ and $\sup \mathcal{Y} = +\infty$. We further assume that F_X and F_Y have densities f_X and f_Y, respectively, and that f_X is monotone on some interval $(-\infty, z_X)$ and f_Y is monotone on some interval (z_Y, ∞).

(iii) The factor model $s^*(x,y)$ is strictly increasing, differentiable in both components and the inverse functions exist on its support; i.e. for all $s \in \mathcal{S}$, $w \in \mathcal{W}$, and $x \in \mathcal{X}$ there exists an inverse function $y^*(s/w,x) \in \mathcal{Y}$ and for all $s \in \mathcal{S}$, $w \in \mathcal{W}$, and $y \in \mathcal{Y}$ there exists an inverse function $x^*(s/w,y) \in \mathcal{X}$, so that

$$s = ws^*(x^*(s/w,y),y) = ws^*(x,y^*(s/w,x)).$$

(iv) We assume $\lim_{y\to\infty} F_X\left(x^*(0,y)\right)/\overline{F}_Y(y) < \infty$.

(v) The default threshold s is negative. □

Assumption 5.2.5 is nothing but Assumption 1 and the comment before Lucas et al. (2003, Assumption 2A), amended by some further regularities.

Assumption 5.2.5(iii) says that we only consider factor models, where, given three components of (S_j, W, X, Y_j), the fourth is uniquely determined. Also note that with Theorem 5.2.3 it follows that $P(S_1 < s|W,X) =: L \stackrel{\mathrm{d}}{=} F_Y\left(y^*(s/W,X)\right)$.

Assumption 5.2.5(iv) is needed since we extend the standard latent variable model $s^*(X,Y_j)$ by the multiplicative factor W. Note that the default probability $P(L_j = 1)$ is in general small and therefore, if $E(S_j) = 0$, we always have $s < 0$, hence Assumption 5.2.5(v) is not restrictive.

Assumptions 5.2.5 hold for a large number of factor models. For instance, they are satisfied by the CreditMetrics model as well as for the multivariate t-model. In the following we focus on the extended CreditMetrics model (5.1.2) and turn our attention to the right tail behavior of the limiting portfolio credit losses L. From the right tail behavior we can deduce the *riskyness* of the portfolio.

Before we specify the different types of distributions of X and Y_j further, we introduce the concept of *regular variation*.

Definition 5.2.6. *(i) A positive, Lebesgue measurable function r is called* regularly varying at infinity with index $\alpha \in \mathbb{R}$ *and we write $r \in \mathcal{R}_\alpha$, if*

$$r(tx)/r(x) \stackrel{x\to\infty}{\longrightarrow} t^\alpha,\ t > 0.$$

If $\mathcal{L} \in \mathcal{R}_0$, then $\mathcal{L}$ is called slowly varying at infinity *and we write $\mathcal{L} \in \mathcal{R}_0$.*

(ii) $r \in \mathcal{R}_\alpha$ if and only if $r(x) = x^\alpha \mathcal{L}(x)$ for $\mathcal{L} \in \mathcal{R}_0$.

(iii) If $X \sim F$ with $\overline{F} \in \mathcal{R}_{-\alpha}$ for some $\alpha \geq 0$ holds, then the random variable X is called regularly varying at infinity with index $-\alpha$ *and we write $X \in \mathcal{R}_{-\alpha}$.* □

For more details on the concept of regular variation we refer to Bingham, Goldie, and Teugels (1989).

If we want to determine large losses of the limiting portfolio L we are interested in its right tail behavior near 1 and we use extreme value theory as the natural tool to describe this tail.

Definition 5.2.7. *We say that the random variable X or the distribution function F of X belongs to the* maximum domain of attraction *of the* Weibull *distribution*

$$\Psi_\kappa(x) = \exp\left(-\left(\max\{-x, 0\}\right)^\kappa\right), \quad \kappa > 0,$$

if for the iid sequence $X_1, X_2, \ldots \stackrel{iid}{\sim} F$ there exist norming constants $c_n > 0$, $d_n \in \mathbb{R}$ such that (as $n \to \infty$)

$$\left(\max\{X_1, \ldots, X_n\} - d_n\right)/c_n \stackrel{\mathrm{d}}{\longrightarrow} \Psi_\kappa.$$

We write $X \in DA(\Psi_\kappa)$ or $F \in DA(\Psi_\kappa)$, and it can be shown that in this case F has a finite right endpoint $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} < \infty$. It also can be shown, that $F \in DA(\Psi_\kappa)$ if and only if $\overline{F}(x) = (x_F - x)^\kappa \cdot \mathcal{L}(1/(x_F - x))$ with $\mathcal{L} \in \mathcal{R}_0$, $x_F < \infty$ and $\kappa > 0$. □

For more details on extreme value theory we refer to Embrechts, Klüppelberg, and Mikosch (1997) or to Resnick (1987).

The following two assumptions classify the different regimes of tail behavior of the risk factors $X \sim F_X$ and $Y_j \sim F_Y$. The first regime assumes polynomially decreasing tails of the risk factors.

Assumption 5.2.8. *(i) $F_X(-\cdot) \in \mathcal{R}_{-\mu_X}$, $\mu_X > 0$, i.e. $F_X(-x) = x^{-\mu_X}\mathcal{L}_X(x)$, $x > 0$, and $\mathcal{L}_X \in \mathcal{R}_0$.*

(ii) $Y_j \sim F_Y$ *and* $\overline{F}_Y \in \mathcal{R}_{-\nu_Y}$, $\nu_Y > 0$, *i.e.* $\overline{F}_Y(y) = y^{-\nu_Y}\mathcal{L}_Y(y)$, $y > 0$, *and* $\mathcal{L}_Y \in \mathcal{R}_0$.

(iii) Let $s < 0$ *be the threshold from Definition 5.2.2 and consider the function* $x^*(\cdot,\cdot)$ *defined in Assumption 5.2.5(iii). Define*

$$\zeta(w,y) \quad := \quad \frac{yD_2x^*(s/w,y)}{x^*(s/w,y)}.$$

Assume $\lim_{y\to\infty}\zeta(w,y) = \zeta \in (0,\infty)$ *for any* $w \in (0,\infty)$ *pointwise. Further, assume that there exists an integrable (w.r.t.* F_W*) function* u *such that* $\zeta(w,y) \le u(w)$ *for all* $w \in (0,\infty)$*, for all* $y \in (y_0,\infty)$ *and some* y_0. □

The second regime assumes exponentially decreasing tails of the risk factors.

Assumption 5.2.9. *(i)* $F_X(-x) = r_X(x)\exp\left(-\mu_X x^{\mu_2}(1+\varepsilon_X(x))\right)$, $x > 0$, *where* $\varepsilon_X(x) = o(1)$, $r_X \in \mathcal{R}_{\mu_1}$, $\mu_X,\mu_2 > 0$ *and* $\mu_1 \in \mathbb{R}$. *Further, let also the derivatives* ε'_X *and* r'_X *be ultimately monotone for* $x \to \infty$.

(ii) $\overline{F}_Y(y) = r_Y(y)\exp\left(-\nu_Y y^{\nu_2}(1+\varepsilon_Y(y))\right)$, $y > 0$, *where* $\varepsilon_Y(y) = o(1)$, $r_Y \in \mathcal{R}_{\nu_1}$, $\nu_Y,\nu_2 > 0$ *and* $\nu_1 \in \mathbb{R}$. *Further, let also the derivatives* ε'_Y *and* r'_Y *be ultimately monotone for* $y \to \infty$.

(iii) Let $s < 0$ *be the fixed threshold from Definition 5.2.2 and consider the function* $x^*(\cdot,\cdot)$ *defined in Assumption 5.2.5(iii). Define*

$$\zeta(w,y) \quad := \quad \frac{\mu_2\left(-x^*(s/w,y)\right)^{\mu_2-1}\left(-D_2x^*(s/w,y)\right)}{\nu_2 y^{\nu_2-1}}$$

Assume $\lim_{y\to\infty}\zeta(w,y) = \zeta \in (0,\infty)$ *for any* $w \in (0,\infty)$ *pointwise. Further, assume that there exists an integrable (w.r.t.* F_W*) function* u *such that* $\zeta(w,y) \le u(w)$ *for all* $w \in (0,\infty)$*, for all* $y \in (y_0,\infty)$ *and some* y_0. □

Note that Assumptions 5.2.8 and 5.2.9 are slightly stronger than Lucas et al. (2003, Assumptions 2A and 2B), since we use the existence of a density of L in the proof of the following Theorem 5.2.10.

We now determine the right tail behavior of the limiting portfolio credit loss distribution.

Theorem 5.2.10. *Consider the setting of Assumption 5.2.5. If Assumptions 5.2.8 or 5.2.9 are satisfied, then* $L \in DA(\psi_\kappa)$ *with* $\kappa = \zeta\mu_X/\nu_Y > 0$, *i.e. there exists* $\mathcal{L} \in \mathcal{R}_0$ *such that*

$$P(L > q) = (1-q)^{\zeta\mu_X/\nu_Y}\mathcal{L}(1/(1-q)), \quad q \in (0,1). \tag{5.2.1}$$

□

For $W \equiv 1$ this result has been proved in Lucas et al. (2003, Theorems 2 and 3). Hence, our result shows that the tail of the portfolio loss is in first order robust with respect to a shock variable W. Consequently, any difference between $W \equiv 1$ and a random $W > 0$ can only be found in the second order tail expansion, the slowly varying function $\mathcal{L}(1/(1-q))$.

As an example, we derive an analytic expression of $\mathcal{L}(1/(1-q))$ in the extended CreditMetrics framework, both, in the setting of Assumptions 5.2.8 and 5.2.9:

Theorem 5.2.11. *Given the extended CreditMetrics model* (5.1.2) *with* $X \sim t_{\mu_X}$, $Y_j \sim t_{\nu_Y}$ *and* $W > 0$ *such that* $\mu_X \geq \nu_Y$. *Let Assumptions 5.2.5 hold. Then the distribution of* L *is of the form (5.2.1) with* $\kappa = \mu_X/\nu_Y$ *and* $\mathcal{L} \in \mathcal{R}_0$ *satisfies for* $q \to 1$ *the relation*

$$\mathcal{L}\left(\frac{1}{1-q}\right) \sim C_{\mu_X}\mu_X^{(\mu_X-1)/2}\int_0^\infty\left(-\frac{s}{aw}(1-q)^{1/\nu_Y} + \frac{b}{a}C_{\nu_Y}^{1/\nu_Y}\nu^{(\nu_Y-1)/(2\nu_Y)}\right)^{-\mu_X}\mathrm{d}F_W(w).$$

□

Theorem 5.2.12. *Given the extended CreditMetrics model* (5.1.2) *with* $X, Y_j \sim \mathcal{N}(0,1)$ *and* $W > 0$ *such that* $E(1/W) < \infty$ *and* $b \geq a$. *Let Assumptions 5.2.5 hold. Then the distribution of* L *is of the form (5.2.1) with* $\kappa = b^2/a^2$ *and* $\mathcal{L} \in \mathcal{R}_0$ *satisfies for* $q \to 1$ *the relation*

$$\mathcal{L}\left(\frac{1}{1-q}\right) \sim \int_0^\infty \exp\left(-\frac{s^2}{2aw} + \frac{sb}{a^2w}\sqrt{-2\ln(1-q)}\right)\frac{(-2\ln(1-q))^{b^2/(2a^2)}}{\left|\frac{s}{aw} - \frac{b}{a}\sqrt{-2\ln(1-q)}\right|}\,\mathrm{d}F_W(w).$$

□

Remark 5.2.13. (i) In the setting of Theorem 5.2.12 we require $b \geq a > 0$. The natural choice in this model is $a = \sqrt{\rho}$ and $b = \sqrt{1-\rho}$ for $\rho \in (0,1)$ modelling the correlation between S_i and S_j for $i \neq j$. Then, $b \geq a$ is equivalent to $\rho \leq 1/2$ and this is always given in practice.

(ii) The first order tail behavior is a function of the correlation ρ only.

(iii) As can be seen in the proof, for the CreditMetrics model Assumptions 5.2.5(iv) and 5.2.8(iii) or 5.2.9(iii) are superfluous. However, in the extended model, one can easily construct examples, where these restrictions are essential. □

Setting $W \equiv 1$ in Theorem 5.2.12 we immediately obtain Lucas et al. (2003, Theorem 6).

Corollary 5.2.14 (Lucas et al. (2003), Theorem 6). *For the CreditMetrics model with $b \geq a$ the tail-distribution of L is of the form (5.2.1) with $\kappa = b^2/a^2$ and $\mathcal{L} \in \mathcal{R}_0$ satisfies for $q \to 1$ the relation*

$$\mathcal{L}\left(\frac{1}{1-q}\right) \sim \frac{a}{b}\exp\left(-\frac{s^2}{2a} + \frac{sb}{a^2}\sqrt{-2\ln(1-q)}\right)(-2\ln(1-q))^{(b^2-a^2)/(2a^2)}.$$

□

5.3 A simulation study

We focus on the extended CreditMetrics model (5.1.2). Denote the default probability by $p_{\text{loss}} = P(S_j \leq s)$ and we assume that $p_{\text{loss}} < 1/2$. We consider different distributions of W, X and Y_j and show their influence on the tail-distribution of the limiting portfolio credit loss L. We consider the following examples.

Model 5.3.1. *(1) $W \equiv 1$ and $X, Y_j \overset{iid}{\sim} \mathcal{N}(0,1)$ and $b \geq a$.*

(2) $W \overset{\text{d}}{=} \sqrt{4/\chi_4^2}$ and $X, Y_j \overset{iid}{\sim} \mathcal{N}(0,1)$ and $b \geq a$.

(3) $W \equiv 1$ and $X \sim t_{\mu_X}$, $Y_j \sim t_{\nu_Y}$ and $\mu_X \geq \nu_Y > 2$.

(4) $W \overset{\text{d}}{=} \sqrt{4/\chi_4^2}$ and $X \sim t_{\mu_X}$, $Y_j \sim t_{\nu_Y}$ and $\mu_X \geq \nu_Y > 2$. □

As shown in Theorems 5.2.11 and 5.2.12, all these models fall into the framework of our assumptions, i.e. for $q \in (0,1)$ there are functions $\mathcal{L}_1, \dots, \mathcal{L}_4 \in \mathcal{R}_0$ such that

$$\begin{aligned} P(L > q) &= (1-q)^{b^2/a^2}\mathcal{L}_{1,2}(1/(1-q)), \text{ in case of model 1 and 2,} \\ P(L > q) &= (1-q)^{\mu_X/\nu_Y}\mathcal{L}_{3,4}(1/(1-q)), \text{ in case of model 3 and 4.} \end{aligned}$$

As indicated in Remark 5.2.13 the restriction $b \geq a$ for model 1 and 2 is quite natural corresponding to $\rho < 1/2$; see Table 5.1 for some scenarios. The restriction $\mu_X \geq \nu_Y$ for model 3 and 4 can be seen in the same spirit as we choose $\mu_X/\nu_Y = b^2/a^2 \geq 1$. The bound $\nu_Y, \mu_X > 2$ is needed to ensure finite variance of S_j.

To make the four models comparable, we fix the following parameters

- the default probability $p_{\text{loss}} := P(S_j \leq s)$,
- the correlation-structure $\rho := \text{Corr}(S_i, S_j)\ \forall i \neq j$ and

- the first order tail behavior $\kappa = \zeta\mu_X/\nu_Y$ of the limiting portfolio credit loss L, given by Theorem 5.2.10.

For all models we have $\text{Corr}(S_i, S_j) = a^2 EX^2/(a^2 EX^2 + b^2 EY_i^2)\ \forall i \neq j$. Let $a = \sqrt{\rho}$, $b = \sqrt{1-\rho}$ in models 1, 2 and $a = \sqrt{\rho(\mu_X - 2)/\mu_X}$, $b = \sqrt{(1-\rho)(\nu_Y - 2)/\nu_Y}$, $\mu_X, \nu_Y > 2$ in models 3, 4. Then we have always the same correlation $\rho \in (0,1)$ in all models.

By Theorem 5.2.12 we have $\kappa = b^2/a^2 = (1-\rho)/\rho$ in model 1 and 2 as the parameter of the first order tail behavior. In model 3 and 4 we get $\kappa = \mu_X/\nu_Y$ (by Theorem 5.2.11), therefore we choose $\mu_X = 2/\rho$ and $\nu_Y = 2/(1-\rho)$ and this leads to $a = b = \sqrt{\rho(1-\rho)}$. Hence we have the same κ in all models.

The threshold s is the p_{loss}-quantile of S_j. Since $S_j \sim \mathcal{N}(0,1)$ in model 1 and $S_j \sim t_4$ in model 2, we can read off this quantile from standard tables. In model 3 and 4, we choose s as the empirical p_{loss}-quantile of S_j. The simulation run length is 10^7, which should suffice to obtain a reliable estimate.

In choosing the specific default probabilities and correlations we follow Frey et al. (2001), i.e. we consider three rating groups of decreasing credit quality, which we label A, B and C; see Table 5.1. This leads to the (rounded) parameters given in Table 5.2.

group	A	B	C
p_{loss}	0.01%	0.50%	7.50%
ρ	2.58%	3.80%	9.21%

Table 5.1: Values for default probability and correlation of the three credit quality groups.

As stated in Theorem 5.2.3 we have $L \stackrel{d}{=} F_Y(s/(bW) - Xa/b)$ and we simulate L by this distributional equality, see Figures 5.1 to 5.3 corresponding to the three groups. Each of Figures 5.1 to 5.3 shows four graphs, each with four curves, corresponding to the different models (1)-(4) with parameters as given in Tables 5.1 and 5.2.

The upper left graph corresponds to the tail-distribution $\overline{L}(q)$ of the limiting portfolio, where the arguments q are chosen such that $0 \leq \overline{L}(q) \leq 0.1$ for all four models; the lower left graph is similar but zoomed in, i.e. q is such that $0 \leq \overline{L}(q) \leq 0.01$. The right graphs show the quantile functions or the *Value-at-Risk* $L^{\leftarrow}(p) = \text{VaR}_p$ of the portfolios with $0.9 \leq p \leq 1$ and $0.99 \leq p \leq 1$, respectively.

In Table 5.3 the VaR_p of all models in the three groups is given for p running through the different values $95\%, 99\%, 99.5\%, 99.9\%, 99.95\%$.

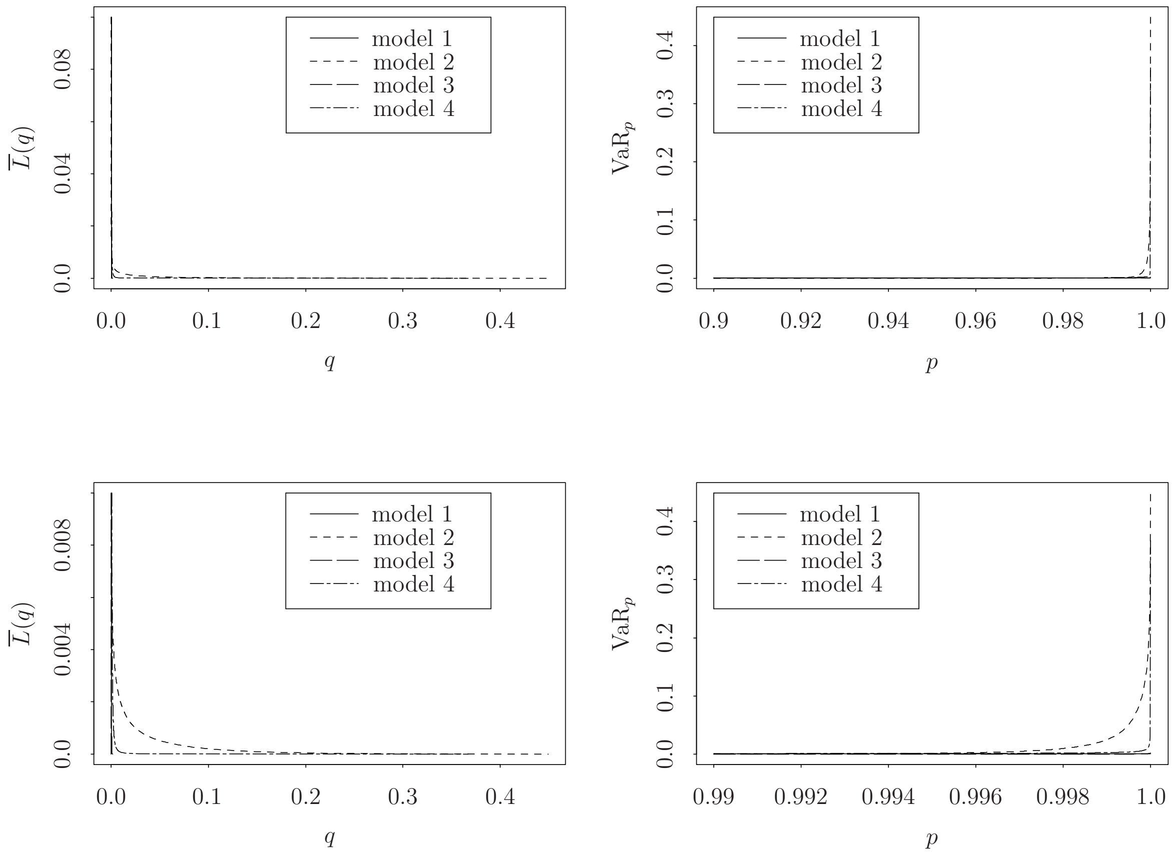

Figure 5.1: Tail-distribution and Value-at-Risk of the four models with group A-parameter setting.

We observe in all groups that model 2 leads to a portfolio with larger quantiles than model 1 and, similarly, model 4 gives larger quantiles than model 3; this is obviously due to W. Although three parameters are the same in all models, we observe a completely different behavior of the four models in their right tails. As can be seen in Table 5.3, the 99.95%-quantile of model 2 in group A is 90 times larger than in case of model 1 and even 440 times larger than in case of model 3. In group B we observe in model 2 an up to 25 times larger 99.95%-quantile than in model 3 and in group C the riskyness of the models turn where model 4 shows up to 50% larger quantiles than model 2.

To quantify the different portfolio behavior further we also estimate empirically the standard deviation of L, see Table 5.4. The (rounded) 95% confidence intervals are, as usual, based on the asymptotic χ^2_{n-1} distribution of the empirical variance. We observe that model 2 has a larger empirical deviation than model 1 and, similarly, model 4 shows

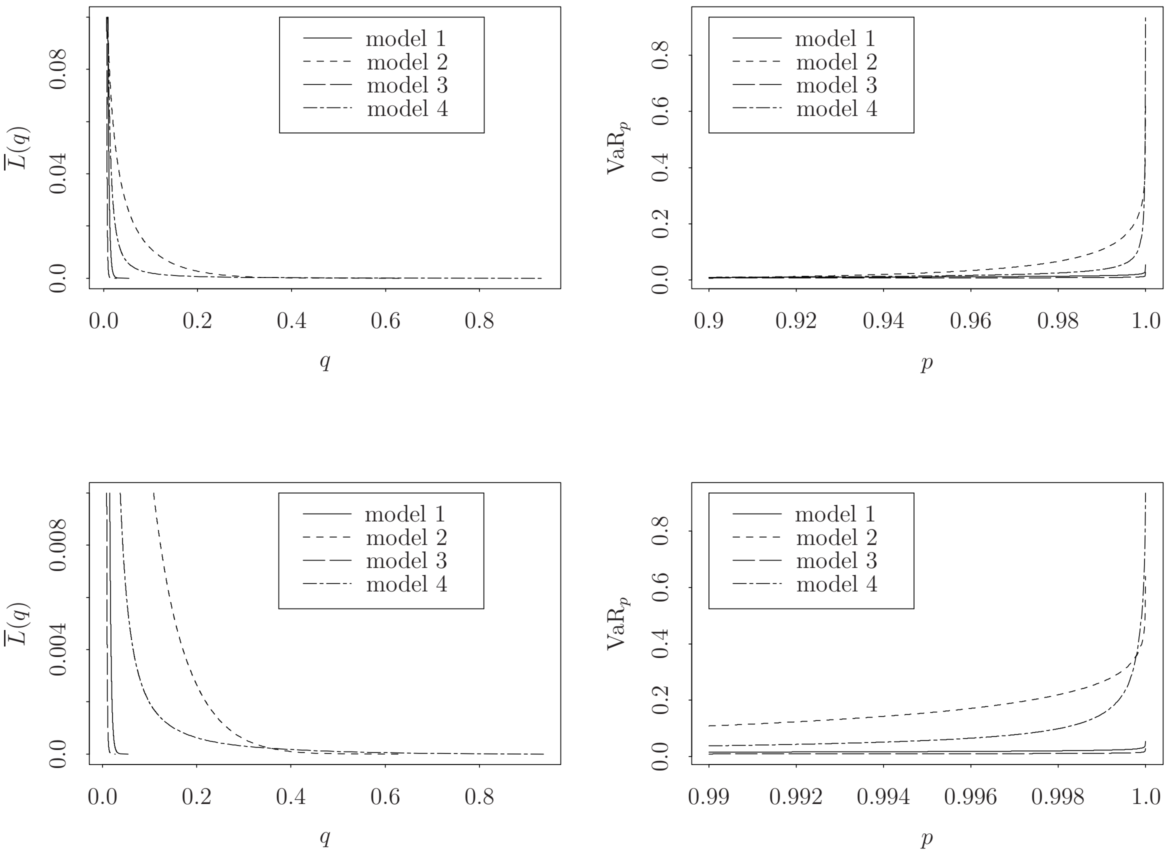

Figure 5.2: Tail-distribution and Value-at-Risk of the four models with group B-parameter setting.

larger deviation than model 3. As in case of the VaR the differences of the standard deviations are not negligible: in group A model 2 shows 850 times more deviation than model 3; see Table 5.4. From Lemma 5.2.4 we get an upper bound for the standard deviation and observe in all our models a quite small standard deviation compared to the upper bound, see also Table 5.4. The meaning of the last line in Table 5.4 will be explained in the following section.

5.4 Cutting Gordon's knot

Recall that for all models of section 5.3 the parameters where chosen such that default probability, correlation and first order tail behavior are the same for all models in each group A-C. Nonetheless, we observe completely different upper tails for the different

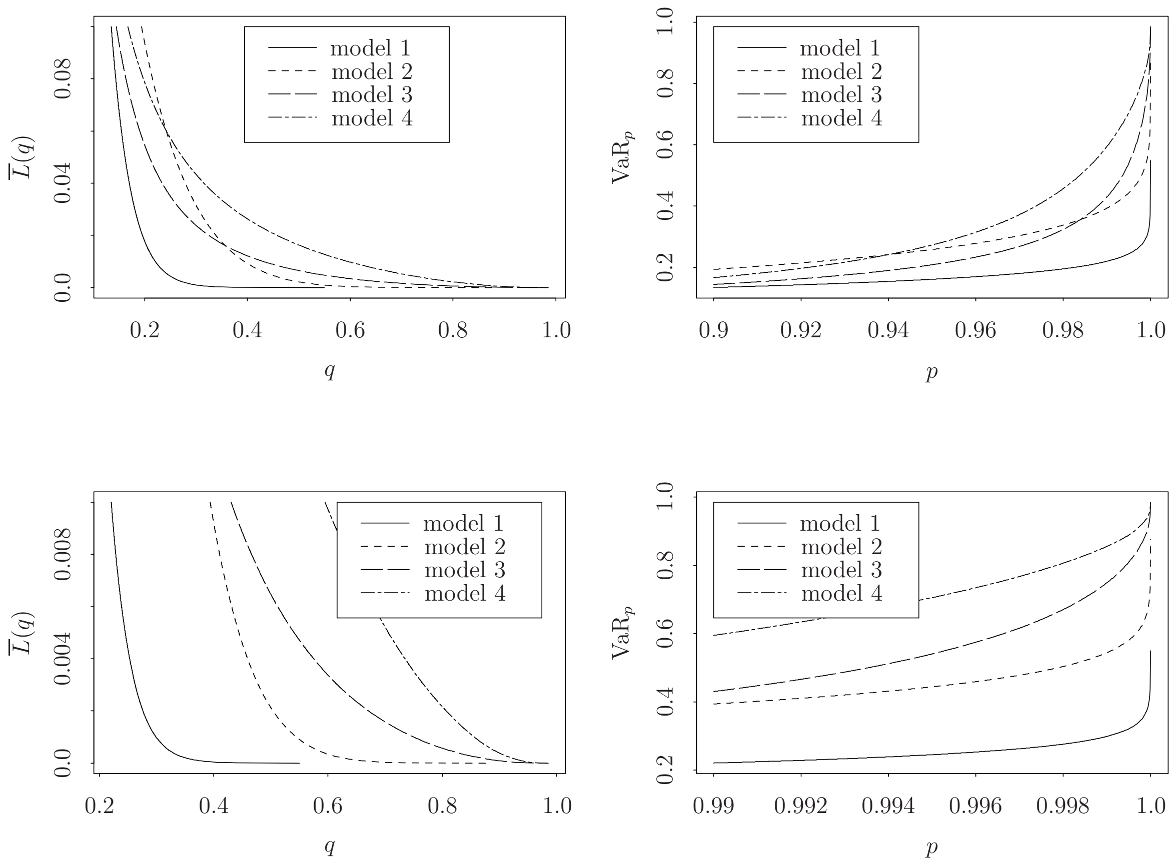

Figure 5.3: Tail-distribution and Value-at-Risk of the four models with group C-parameter setting.

models. This indicates that a naive quantile estimator based on extreme value theory may be grossly misleading. Such a method would concentrate on the parameter κ in Theorem 5.2.10 and replace the slowly varying function $\mathcal{L}$ by a constant, see Embrechts et al. (1997, Chapter 6) for details. However, as can be seen in Theorem 5.2.12, $\mathcal{L}$ is far away from being constant and has a strong influence near the right endpoint $q = 1$.

To overcome the problem, which model to choose, we suggest in the following a pragmatic approach, which originates in the β-model. The β model is a simple model often used in practice, where the parameters are estimated by matching the first two moments; see e.g. Bluhm, Overbeck, and Wagner (2003, page 39). The $\beta(c, d)$-distribution has density

$$f_{\beta(c,d)}(q) = \frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} q^{c-1}(1-q)^{d-1}, \quad 0 < q < 1, \quad c, d > 0.$$

From Embrechts et al. (1997, Example 3.3.17) we know that the $\beta(c, d)$-distribution satisfies the weak requirement of being in $DA(\Psi_d)$. As our main focus is on VaR-estimation, we fit besides the location parameter the first order tail behavior. Since $\kappa = (1-\rho)/\rho$ and $E\beta(c,d) = c/(c+d)$ we obtain

$$c = \frac{1-\rho}{\rho}\frac{p_{\text{loss}}}{1-p_{\text{loss}}} \quad \text{and} \quad d = \frac{1-\rho}{\rho}. \tag{5.4.1}$$

This means we match the default probability p_{loss} and the correlation ρ.

We observe that VaR estimated from the β-model compared to our models 1-4 is slightly more moderate but roughly of the same order as for model 2 in all groups; see Table 5.3.

The question arises, if there is any further advantage of the latent variable models 1-4 in comparison to the simple and easy to fit β-model for VaR estimation, which after all, has the correct first order tail behaviour. One drawback of the β-model is that it has no economic interpretation in the credit risk context. From a statistical point of view, models 2 and 4 constitute a much richer class of models in the sense that more parameters can be specified.

One parameter, which we have not considered up to now is the standard deviation (see Table 5.4) and here we can observe substantial differences between the models. As the first order tail behavior is determined by ρ solely, it is independent of W. As W acts as a random standard deviation of the factor models, it is natural to match the empirical standard deviation by choosing a proper W. In our simulations we observe for models 2 and 4 that the standard deviation of L_ν is decreasing in ν_W. For the normal factor model 2 we can proof this by asymptotic expansion. Consequently, we can estimate ν_W by matching the standard deviation.

Theorem 5.4.1. *In the setting of model 2, let* $W = W_\nu \stackrel{\text{d}}{=} \sqrt{\nu_W/\chi^2_{\nu_W}}$ *and denote* $L_j = L_{j,\nu}$ *and* $L = L_\nu$. *Then, the standard deviation of* L_ν *is decreasing in* ν *for sufficiently small default probability* $p_{\text{loss}} = P(L_{j,\nu} = 1) = P(S_j < s)$. □

We conclude this section by a comparison of the extended CreditMetrics models 2 and 4 with the β-model estimated from the parameters p_{loss} and ρ given in Table 1. The estimated parameters for group A-C are given in Table 5.2. We choose σ from the parameters (5.4.1) of the β-model; the estimates are given in the last line in Table 5.4. We see that in case of group C the standard deviation of model 3 is already slightly larger than in the β-case, therefore we set $\nu_W = \infty$ (corresponding to $W \equiv 1$) for model 4 in group C. All results are summarized in Table 5.5.

In Figure 5.4 we plot the tail-distribution (left column) and the VaR_p (right column), where the upper, middle and lower row correspond to group A, B and C, respectively. In

Table 5.6 we also give the VaR_p estimates of model 2, 4 and β-model in the three groups for certain values of p. We observe now that model 2 and the β-model are very similar in all groups, indicating that the β-model gives a reasonable approximation for model 2, provided the standard deviations of both models coincide. In other words, the similarity of model 2 and the β-model suggests model 2 as a substantial improvement of the β-model.

As to model 4, we see that in group A the quantiles of model 2 and β are roughly three times larger than in model 4. In group B, all three models are comparable and in group C model 4 behaves roughly 50% riskier than the other models. We shall further comment on model 4 in the next section.

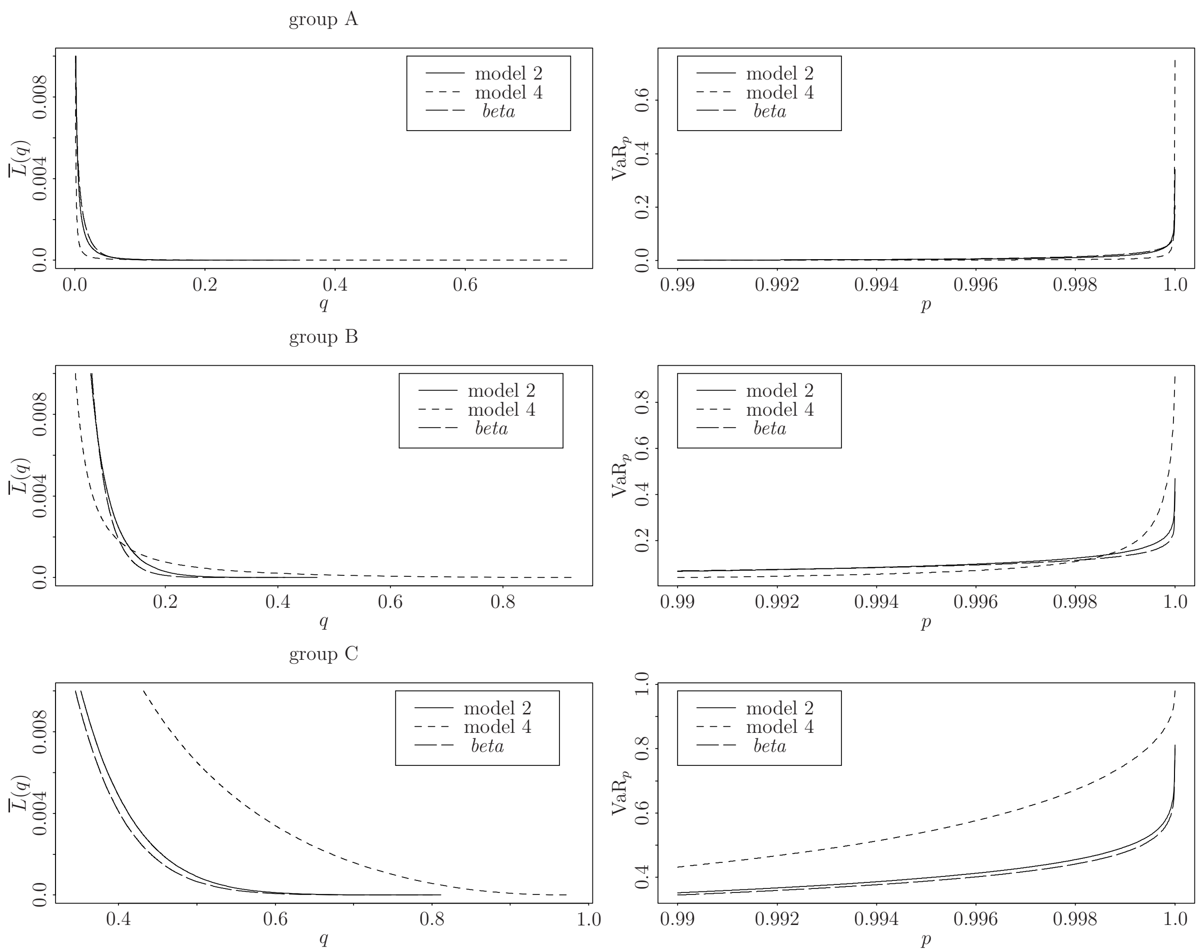

Figure 5.4: Tail-distribution and Value-at-Risk of model 2, 4 and β-model in the three groups.

5.5 A word of warning

In the heavy-tailed models 3 and 4 we restrict the parameters to $\mu_X \geq \nu_Y$, i.e. we consider only X being not heavier-tailed than Y_j. As can be seen in Table 5.2 we always have $\mu_X > \nu_Y$ with a rather large ratio $\mu_X/\nu_Y > 9.8$. We did this for good reasons. Because, if $\nu = \mu_X = \nu_Y$, then this models a very extreme economic situation, the more extreme, the smaller ν is. In this case X (and Y_j) have extremely heavy tails and, thus, have with very high probability extremely large realizations. Consequently, it can happen that a large negative observation of X dominates all Y_j such that almost the whole portfolio defaults. This would model an economy which fluctuates wildly. In that case the limiting portfolio credit loss behaves like the model built on $S_j = \min\{aX, bY_j\}$.

Corollary 5.5.1. *Define* $L_\wedge^{(m)} := \frac{1}{m}\sum_{j=1}^m L_j^\wedge$ *with default indicators* $L_j^\wedge := \mathbf{1}_{\{\min\{aX,bY_j\}\leq s\}}$. *Let* $X, Y_1, \ldots, Y_m \overset{iid}{\sim} t_\nu$. *Then*

$$L^\wedge := \lim_{m\to\infty} L_\wedge^{(m)} \overset{\text{a.s.}}{=} \begin{cases} 1, & \textit{with probability } F_{t_\nu}(s/a), \\ F_{t_\nu}(s/b), & \textit{with probability } \overline{F}_{t_\nu}(s/a). \end{cases}$$

□

Theorem 5.5.2. *Choose the model* $L_\wedge^{(m)}$ *and* $L^\wedge$ *as in the setting above. Let* $L^{(m)}$ *and* L *correspond to model 3 with* $\mu_X = \nu_Y =: \nu$. *Further, choose the same default threshold* s *for both models.*

(i) Let m *be fixed. Then, for any* $q \in \{0, 1/m, 2/m \ldots, 1\}$, *it holds that*

$$\lim_{s\to-\infty} P\left(L^{(m)} = q \middle|\, L_\wedge^{(m)} = q\right) = 1.$$

(ii) Let $\varepsilon > 0$. *Then,* $\lim_{s\to-\infty} P\left(|L - L^\wedge| < \varepsilon \mid L^\wedge\right) \overset{\text{a.s.}}{=} 1$.

□

From Theorem 5.5.2(ii) conclude that in the setting of model 3, where s, $\rho = a^2/(a^2 + b^2)$ and $\mu_X = \nu_Y = \nu$ are small, the limiting portfolio credit loss L degenerates in the sense that most of the mass is near the point $F_{t_\nu}(s/b)$ (when the Y_j's dominate the portfolio) and some very rare events can be observed close to 1 (when X dominates the portfolio). From Theorem 5.5.2(i) conclude that this behavior also can be observed for portfolios with a finite number of loans. Of course, model 4 has the same structure; the difference to model 3 being that large fluctuations are multiplied by a random W.

5.6 Conclusion

In this chapter we derived the tail behavior of aggregate credit losses extending results of Lucas et al. (2003). We enriched the one factor latent variable model by a positive multiplicative shock variable W. In the models, where the latent variables follow a multivariate normal or t-distribution, we observed that first order tail behavior is a function of the correlation between the latent variables. In particular, W has no influence on the first order tail behavior of the limiting credit loss portfolio.

In a simulation study we observed an impact of the second order tail behavior on the quantiles by comparing four different models. We fitted the models by matching default probability, correlation between latent variables and first order tail behavior. In some credit scenario we observed quantiles that were up to 440 times larger than in another scenario.

To offer some decision support to the risk manager on which model to choose, we compared the VaR estimated from the β-model with the VaR estimated from the four extended CreditMetrics models. Fixing default probability and first order tail behavior we observed a similar (slightly more moderate) performance of the β-model and the multivariate t-model. From Section 5.5 we learned to be aware of the influence of heavy-tailed latent variables as the limiting credit loss portfolio may degenerate. This suggests the β-model as a simple model based on the fit of two quantities of interest, either matching the first two moments, or, perhaps more advisable in the context of risk management and VaR estimation, loss probability and correlation.

The multivariate t-model offers an improved fit by the shock variable W. We showed that W can influence the standard deviation without having influence on the other parameters. As for small loss probabilities the standard deviation of the limiting credit loss distribution decreases in ν, we estimate ν by matching the standard deviation. Consequently, the multivariate t-model improves the fit of the β-model.

Model 1: $X, Y_i \overset{iid}{\sim} \mathcal{N}(0,1)$ and $W \equiv 1$:							
		a	b	s	κ	μ_X	ν_Y
group	A	.161	.987	-3.73	37.8	.500	.500
	B	.195	.981	-2.58	25.3	.500	.500
	C	.303	.953	-1.44	9.86	.500	.500

Model 2: $X, Y_i \overset{iid}{\sim} \mathcal{N}(0,1)$ and $W \overset{\mathrm{d}}{=} \sqrt{4/\chi^2_4}$:							
		a	b	s	κ	μ_X	ν_Y
group	A	.161	.987	-13.0	37.8	.500	.500
	B	.195	.981	-4.60	25.3	.500	.500
	C	.303	.953	-1.78	9.86	.500	.500

Model 3: $X \sim t_{\mu_X}$, $Y_i \sim t_{\nu_Y}$ and $W \equiv 1$:							
		a	b	s	κ	μ_X	ν_Y
group	A	.159	.159	-10.1	37.8	77.5	2.05
	B	.191	.191	-1.81	25.3	52.6	2.08
	C	.289	.289	-.782	9.86	21.7	2.20

Model 4: $X \sim t_{\mu_X}$, $Y_i \sim t_{\nu_Y}$ and $W \overset{\mathrm{d}}{=} \sqrt{\chi^2_1}$:							
		a	b	s	κ	μ_X	ν_Y
group	A	.159	.159	-10.2	37.8	77.5	2.05
	B	.191	.191	-1.88	25.3	52.6	2.08
	C	.289	.289	-.679	9.86	21.7	2.20

Table 5.2: Parameter setting of the four models in the three groups (given by Table 5.1).

VaR$_p$ for group A, $p_{\text{loss}} = 0.0001$ and $\rho = 0.0258$.						
	p	95%	99%	99.5%	99.9%	99.95%
model	1	$2.18 \cdot 10^{-4}$	$3.29 \cdot 10^{-4}$	$3.82 \cdot 10^{-4}$	$5.14 \cdot 10^{-4}$	$5.76 \cdot 10^{-4}$
	2	$1.54 \cdot 10^{-8}$	$1.76 \cdot 10^{-4}$	$1.46 \cdot 10^{-3}$	$2.48 \cdot 10^{-2}$	$5.12 \cdot 10^{-2}$
	3	$1.09 \cdot 10^{-4}$	$1.12 \cdot 10^{-4}$	$1.13 \cdot 10^{-4}$	$1.15 \cdot 10^{-4}$	$1.16 \cdot 10^{-4}$
	4	$2.98 \cdot 10^{-4}$	$7.30 \cdot 10^{-4}$	$1.06 \cdot 10^{-3}$	$2.46 \cdot 10^{-3}$	$3.52 \cdot 10^{-3}$
	β	$1.91 \cdot 10^{-8}$	$1.10 \cdot 10^{-3}$	$4.73 \cdot 10^{-3}$	$2.37 \cdot 10^{-2}$	$3.47 \cdot 10^{-2}$

VaR$_p$ for group B, $p_{\text{loss}} = 0.005$ and $\rho = 0.038$.						
	p	95%	99%	99.5%	99.9%	99.95%
model	1	0.0107	0.0152	0.0173	0.0221	0.0242
	2	0.0254	0.108	0.155	0.265	0.308
	3	0.00715	0.00871	0.00942	0.0113	0.0122
	4	0.0143	0.0376	0.0568	0.151	0.226
	β	0.0285	0.069	0.0886	0.135	0.155

VaR$_p$ for group C, $p_{\text{loss}} = 0.075$ and $\rho = 0.0921$.						
	p	95%	99%	99.5%	99.9%	99.95%
model	1	0.162	0.221	0.245	0.299	0.321
	2	0.259	0.394	0.444	0.544	0.581
	3	0.209	0.431	0.541	0.750	0.810
	4	0.274	0.595	0.706	0.856	0.889
	β	0.233	0.345	0.388	0.478	0.513

Table 5.3: VaR$_p$, p = 95%, 99%, 99.5%, 99.9%, 99.95%, for the four models and the fitted β-distribution in the three groups.

		group		
		A	B	C
	$\sigma_{\max}$	0.01	0.0705	0.263
		$\widehat{\sigma}_L$		
model	1	$6.54 \cdot 10^{-5} \pm 2 \cdot 10^{-6}$	$3.00 \cdot 10^{-3} \pm 9 \cdot 10^{-5}$	$4.50 \cdot 10^{-2} \pm 2 \cdot 10^{-3}$
	2	$2.93 \cdot 10^{-3} \pm 9 \cdot 10^{-5}$	$2.16 \cdot 10^{-2} \pm 7 \cdot 10^{-4}$	$8.72 \cdot 10^{-2} \pm 3 \cdot 10^{-3}$
	3	$3.38 \cdot 10^{-6} \pm 1 \cdot 10^{-7}$	$1.18 \cdot 10^{-3} \pm 4 \cdot 10^{-5}$	$7.84 \cdot 10^{-2} \pm 3 \cdot 10^{-3}$
	4	$3.28 \cdot 10^{-4} \pm 1 \cdot 10^{-5}$	$1.24 \cdot 10^{-2} \pm 4 \cdot 10^{-4}$	$1.07 \cdot 10^{-1} \pm 4 \cdot 10^{-3}$
	β	$1.61 \cdot 10^{-3}$	$1.37 \cdot 10^{-2}$	$7.71 \cdot 10^{-2}$

Table 5.4: Estimated standard deviations $\widehat{\sigma}_L$ with 95%CI of the estimator for the four models in the three groups. The last line shows the standard deviation of the fitted β-model.

		group		
	$\widehat{\nu_W}$	A	B	C
model	2	7.77	8.22	5.74
	4	2.65	3.72	0

Table 5.5: Estimated ν_W for model 2 and 4 with $W \sim \sqrt{\nu_W / \chi^2_{\nu_W}}$.

VaR$_p$ for group A, $p_{\text{loss}} = 0.0001$, $\rho = 0.0258$ and $\sigma = 1.61 \cdot 10^{-3}$.						
	p	95%	99%	99.5%	99.9%	99.95%
model	2	$6.91 \cdot 10^{-5}$	$1.69 \cdot 10^{-3}$	$4.19 \cdot 10^{-3}$	$1.88 \cdot 10^{-2}$	$2.99 \cdot 10^{-2}$
	4	$2.66 \cdot 10^{-4}$	$9.62 \cdot 10^{-4}$	$1.65 \cdot 10^{-3}$	$5.81 \cdot 10^{-3}$	$1.00 \cdot 10^{-2}$
	beta	$1.91 \cdot 10^{-8}$	$1.10 \cdot 10^{-3}$	$4.73 \cdot 10^{-3}$	$2.37 \cdot 10^{-2}$	$3.47 \cdot 10^{-2}$

VaR$_p$ for group B, $p_{\text{loss}} = 0.005$, $\rho = 0.038$ and $\sigma = 1.37 \cdot 10^{-2}$.						
	p	95%	99%	99.5%	99.9%	99.95%
model	2	0.0249	0.0673	0.0906	0.151	0.180
	4	0.0145	0.0400	0.0617	0.170	0.257
	beta	0.0285	0.0693	0.0886	0.135	0.155

VaR$_p$ for group C, $p_{\text{loss}} = 0.075$, $\rho = 0.0921$ and $\sigma = 7.71 \cdot 10^{-2}$.						
	p	95%	99%	99.5%	99.9%	99.95%
model	2	0.234	0.352	0.398	0.495	0.531
	4	0.209	0.432	0.542	0.750	0.810
	beta	0.233	0.345	0.388	0.478	0.513

Table 5.6: VaR$_p$, $p = 95\%$, 99%, 99.5%, 99.9%, 99.95% for model 2, 4 and β-model in the three groups.

5.7 Proofs

Proof of Lemma 5.2.4: As in Definition 5.2.2 set $L_i := \mathbf{1}_{\{W s^*(X,Y_i) \le s\}}$ with $EL_i = p_{\text{loss}}$, $L^{(m)} = \frac{1}{m}\sum_{i=1}^m L_j$. We observe

$$\operatorname{Var}\left(\sum_{i=1}^m L_j\right) = E\left(\left(\sum_{i=1}^m L_j\right)^2\right) - \left(E\left(\sum_{i=1}^m L_j\right)\right)^2 \tag{5.7.1}$$

$$= \sum_{i,j=1}^m E(L_i L_j) - m^2 p_{\text{loss}}^2 \le m^2 p_{\text{loss}}(1 - p_{\text{loss}}), \tag{5.7.2}$$

since $E(L_i L_j) \le E(L_i) = p_{\text{loss}}$. Hence, $\operatorname{Var} L^{(m)} \le p_{\text{loss}}(1 - p_{\text{loss}})$ for all m, and, obviously, $\operatorname{Var} L^{(m)} = p_{\text{loss}}(1 - p_{\text{loss}})$ holds for $L_i \overset{\text{a.s.}}{=} L_j$.

By Theorem 5.2.3 (independent of Lemma 5.2.4), we have $\lim_{m\to\infty} L^{(m)} \overset{\text{a.s.}}{=} L$ and L has bounded support $(0,1)$, hence $\operatorname{Var} L = \operatorname{Var}\left(\lim_{m\to\infty} L^{(m)}\right) = \lim_{m\to\infty} \operatorname{Var} L^{(m)} \le p_{\text{loss}}(1 - p_{\text{loss}})$.

In the extended CreditMetrics setting $S_j = W(aX + bY_j)$ obviously we have for $a = 1$, $b = 0$ that $L_i \overset{\text{a.s.}}{=} L_j$, for all i, j, hence $L^{(m)} \overset{\text{a.s.}}{=} L_1$, therefore $\operatorname{Var} L = \operatorname{Var} L_1 = p_{\text{loss}}(1 - p_{\text{loss}})$ and for $W = \text{const} \in (0, \infty)$, $a = 0$, $b = 1$ we have $L_1, L_2, \ldots \overset{iid}{\sim} \operatorname{Ber}(p_{\text{loss}})$, therefore $\operatorname{Var} L = 0$. □

Proof of Theorem 5.2.3: Given W and X, the indicator variables $L_j = \mathbf{1}_{\{S_j \le s\}}$ are iid, hence a conditional law of large numbers holds as $m \to \infty$ with

$$L^{(m)} = \frac{1}{m}\sum_{j=1}^m \mathbf{1}_{\{S_j \le s\}} = \frac{1}{m}\sum_{j=1}^m \mathbf{1}_{\{W s^*(X,Y_j) \le s\}}$$

$$\xrightarrow{\text{a.s.}} \; E\left(\mathbf{1}_{\{S_1 \le s\}} \,|W, X\right) = P\left(S_1 \le s \,|W, X\right) =: L.$$

Furthermore, by Assumption 5.2.5(iv), $s^*(\cdot,\cdot)$ is increasing and invertible with respect to the second component, therefore

$$L = P(W s^*(X, Y_1) \le s | W, X) = P(Y_1 \le y^*(s/W, X) | W, X) \overset{\text{d}}{=} F_Y\left(y^*(s/W, X)\right).$$

□

For the proof of Theorem 5.2.10 we need the following Lemmas 5.7.1, 5.7.2 and 5.7.3.

Lemma 5.7.1 (Smith (1983, Theorem 10.3, Chapter 13))**.** *Let μ be a finite measure on $\boldsymbol{A} \subset \mathbb{R}^m$, $\boldsymbol{B}$ an open interval in $\mathbb{R}$ and $h : \boldsymbol{A} \times \boldsymbol{B} \to \mathbb{R}^n$ defined by $(w, t) \mapsto h(w, t)$. Assume that the following holds.*

(i) For almost every $t \in \boldsymbol{B}$, the function $h(\cdot, t)$ is measurable on $\boldsymbol{A}$ and for some t it is integrable.

(ii) For almost every $w \in \boldsymbol{A}$, the function $h(w, \cdot)$ is $\mathcal{C}^1$ on $\boldsymbol{B}$.

(iii) There is an integrable $u : \boldsymbol{A} \to \mathbb{R}$ such that

$$|D_2 h(w,t)| \leq u(w) \quad \text{for all } t \in \boldsymbol{B} \text{ and almost all } w \in \boldsymbol{A}.$$

Then the function $\int h(w, \cdot)\, \mathrm{d}\mu(w)$ is $\mathcal{C}^1$ and satisfies

$$\frac{\partial}{\partial t} \int h(w,t)\, \mathrm{d}\mu(w) \quad = \quad \int D_2 h(w,t)\, \mathrm{d}\mu(w).$$

□

Lemma 5.7.2. *(i) Choose the setting of Assumption 5.2.8, then*

$$f_X(-x) \sim \frac{\mu_X}{x} F_X(-x) \text{ as } x \to \infty \quad \text{and} \quad f_Y(y) \sim \frac{\nu_Y}{y} \overline{F}_Y(y) \text{ as } y \to \infty.$$

(ii) Choose the setting of Assumption 5.2.9, then

$$\begin{aligned} f_X(-x) &\sim \mu_X \mu_2 x^{\mu_2 - 1} F_X(-x), \quad \text{as } x \to \infty, \text{ and} \\ f_Y(y) &\sim \nu_Y \nu_2 y^{\nu_2 - 1} \overline{F}_Y(y), \quad \text{as } y \to \infty. \end{aligned}$$

Proof: In the setting of Assumption 5.2.8 just apply the *Monotone Density Theorem*, e.g. Bingham et al. (1989, Theorem 1.7.2), since the densities f_X and f_Y are ultimately monotone. In the setting of Assumption 5.2.9 we have (the asymptotic behavior of f_Y is shown similarly)

$$F_X(-x) \quad = \quad r_X(x) \exp\left(-\mu_X x^{\mu_2}(1 + \varepsilon_X(x))\right).$$

We obtain

$$f_X(-x) = F_X(-x) \mu_X \mu_2 x^{\mu_2 - 1} \left(1 + \varepsilon_X(x) - \frac{x \varepsilon_X'(x)}{\mu_2} + \frac{x^{1-\mu_2}}{\mu_X \mu_2} \frac{r_X'(x)}{r_X(x)}\right).$$

As r_X' is ultimately monotone, the monotone density theorem yields $r_X'(x)/r_X(x) \sim c/x$ as $x \to \infty$. Since $\mu_2 > 0$, it follows that

$$\frac{x^{1-\mu_2}}{\mu_X \mu_2} \frac{r_X'(x)}{r_X(x)} \sim \frac{c}{\mu_X \mu_2} x^{-\mu_2} \overset{x \to \infty}{\longrightarrow} 0.$$

Considering $x\varepsilon'_X(x)$ choose x such that $\varepsilon'_X(\xi)$ is monotone for all $\xi \geq x$. Note that $\varepsilon_X(x) = o(1)$ and monotonicity of ε'_X implies $\varepsilon'_X(x) = o(1)$. Without loss of generality let $\varepsilon'_X > 0$ be decreasing. Hence there exists $\delta > 0$ such that

$$-\varepsilon_X(x) = \int_x^\infty \varepsilon'_X(\xi)\,\mathrm{d}\xi \geq \sum_{i=\lfloor x\rfloor+1}^{\infty} \int_i^{i+1} \varepsilon'_X(\xi)\,\mathrm{d}\xi \geq \sum_{i=\lfloor x\rfloor+2}^{\infty} \varepsilon'_X(i) \in [0,\delta). \tag{5.7.3}$$

Therefore, $i\varepsilon'_X(i) \overset{i\to\infty}{\longrightarrow} 0$, hence (by monotonicity of ε'_X), $x\varepsilon'_X(x) = o(1)$ holds. □

Lemma 5.7.3. *Consider the setting of Assumption 5.2.5 and let q_0 be close to 1. If Assumptions 5.2.8 or 5.2.9 are satisfied, then, for $q \in (q_0, 1)$, L has density*

$$f_L(q) \;=\; -\frac{1}{f_Y\left(F_Y^{\leftarrow}(q)\right)} \int_0^\infty f_X\left(x^*\left(s/w, F_Y^{\leftarrow}(q)\right)\right) D_2 x\left(s/w, F_Y^{\leftarrow}(q)\right) \mathrm{d}F_W(w)\,.$$

Proof: From Theorem 5.2.3 we have

$$L \overset{\mathrm{d}}{=} F_Y\left(y^*(s/W, X)\right).$$

Let q be close to 1; then $L(W, X)$ is larger than q, if $y^*(s/W, X)$ is close to ∞, since Y has support unbounded to the right. By Assumption 5.2.5(iii) F_Y is strictly increasing near its right endpoint, hence the (continuous) inverse $F_Y^{\leftarrow}$ exists. By independence of W and X we have

$$\begin{aligned} P(L > q) &= P\left(F_Y\left(y^*(s/W, X)\right) > q\right) \\ &= P\left(y^*(s/W, X) > F_Y^{\leftarrow}(q)\right) = P\left(X < x^*\left(s/W, F_Y^{\leftarrow}(q)\right)\right) \\ &= \int_0^\infty F_X\left(x^*\left(s/w, F_Y^{\leftarrow}(q)\right)\right) \mathrm{d}F_W(w), \end{aligned} \tag{5.7.4}$$

where the inequality sign is reversed since $y^*(s/W, X)$ is decreasing in X. By Assumption 5.2.5(iii) X and Y_j have ultimately monotone densities f_X and f_Y, respectively.

To show existence and to derive an analytic representation of $f_L(q)$, we set

$$h(w, q) := F_X\left(x^*\left(s/w, F_Y^{\leftarrow}(q)\right)\right) \tag{5.7.5}$$

and show that h satisfies the conditions of Lemma 5.7.1. Since $x^*(\cdot,\cdot)$ is continuous we have that $h(\cdot, t)$ is measurable on $(0,\infty)$ and, since $|h| \leq 1$, it is also integrable with respect to F_W for some $q \in (0, 1)$. Therefore Lemma 5.7.1(i) is applies. Next we have to show that $h(w, \cdot) : (q_0, 1) \to (0, 1)$ is $\mathcal{C}^1$ on $(q_0, 1)$ (as we consider $q \to 1$ we do not need continuity of h for all $q \in (0,1)$). We choose q_0 large enough such that $F_Y^{\leftarrow}$ is $\mathcal{C}^1$ and denote $y := F_Y^{\leftarrow}(q)$. To show that $F_X(x^*(s/w,y))$ is $\mathcal{C}^1$ first note that $x^*(s/w,\cdot)$ is $\mathcal{C}^1$ and decreasing (by Assumption 5.2.5(iv)). Therefore, $\lim_{y\to\infty} x^*(s/w,y) = c \geq -\infty$ and, by Assumption 5.2.8(iii), $c < 0$. Assuming $c > -\infty$ implies $y^*(s/w, c-1) = \infty \notin \mathcal{Y}$. This contradicts Assumption 5.2.5(iii) as $c-1 \in \mathcal{X}$. Therefore $\lim_{y\to\infty} x^*(s/w,y) = -\infty$, hence $h(w,\cdot)$ is $\mathcal{C}^1$ and Lemma 5.7.1 applies.

To show that Lemma 5.7.1(iii) holds observe that

$$D_2 h(w,q) = D_2 x^*(s/w, F_Y^{\leftarrow}(q)) f_X(x^*(s/w, F_Y^{\leftarrow}(q)))/f_Y(F_Y^{\leftarrow}(q)),$$

as $x^*(s/\cdot, y)$ is increasing. Define $y_0 := F_Y^{\leftarrow}(q_0)$, $y := F_Y^{\leftarrow}(q)$, and $x_{w,y} := x^*(s/w,y)$ and choose the setting of Assumption 5.2.8. Then

$$\begin{aligned} |D_2 h(w,q)| &= \frac{f_X\left(x^*\left(s/w, F_Y^{\leftarrow}(q)\right)\right)}{f_Y\left(F_Y^{\leftarrow}(q)\right)} \left|D_2 x^*\left(s/w, F_Y^{\leftarrow}(q)\right)\right| = \frac{f_X\left(x_{w,y}\right)}{f_Y\left(y\right)} |D_2 x_{w,y}| \\ &= \left|\frac{y D_2 x_{w,y}}{x_{w,y}}\right| \frac{|x_{w,y}| f_X(x_{w,y})}{F_X(x_{w,y})} \frac{\overline{F}_Y(y)}{y f_Y(y)} \frac{F_X(x_{w,y})}{\overline{F}_Y(y)}. \end{aligned} \tag{5.7.6}$$

By Lemma 5.7.2(i) we have $f_Y(y) \sim \nu_Y / y\overline{F}_Y(y)$ and therefore $\overline{F}_Y(y)/(y f_Y(y)) \leq 1/\nu_Y + \varepsilon_Y$ for all $y \geq y_0$ and an $\varepsilon_Y > 0$. Similarly, $|x_{w,y}| f_X(x_{w,y})/F_X(x_{w,y}) \to \mu_X$ as $y \to \infty$. As $x_{w,y}$ is increasing in w, $|x_{\infty,y}| f_X(x_{\infty,y})/F_X(x_{\infty,y}) \leq \mu_X + \varepsilon_X$ for all $y \geq y_0$ and an $\varepsilon_X > 0$ implies $|x_{w,y}| f_X(x_{w,y})/F_X(x_{w,y}) \leq \mu_X + \varepsilon_X$ for all $y \geq y_0$, an $\varepsilon_X > 0$ and for all $w \in (0,\infty)$. By Assumption 5.2.5(iv),

$$\lim_{y\to\infty} \frac{F_X(x_{w,y})}{\overline{F}_Y(y)} \leq \lim_{y\to\infty} \frac{F_X(x^*(0,y))}{\overline{F}_Y(y)} < \infty,$$

i.e. $\sup_{w\in(0,\infty),\, y\in(y_0,\infty)} F_X(x_{w,y})/\overline{F}_Y(y) = C < \infty$. Note that there exists a function $u(w)$, integrable with respect to F_W, such that $y D_2 x_{w,y}/x_{w,y} \leq u(w)$ for all y (Assumption 5.2.8(iii)). Hence $|D_2 h(w,q)|$ is dominated by an integrable function $u(w)$ and Lemma 5.7.1(iii) is satisfied. Showing that there is an integrable upper bound $u(w)$ such that $|D_2 h(w,q)| \leq u(w)\ \forall q \in (q_0, 1)$ in the setting of Assumption 5.2.9 is proved similarly using the asymptotic behavior of f_X and f_Y, given in Lemma 5.7.2(ii).

Therefore, by Lemma 5.7.1, we can interchange integration and differentiation and get the result. □

Proof of Theorem 5.2.10: By Lemma 5.7.3, L has density f_L. If we observe $\lim_{q\to 1}(1-q) f_L(q)/\overline{F}_L(q) = \kappa$ then we know from Embrechts, Klüppelberg, and Mikosch (1997, Corollary 3.3.13) that $L \in DA(\Psi_\kappa)$. Substituting $y = F_Y^{\leftarrow}(q)$ (hence $1-q = \overline{F}_Y(y)$), we obtain

$$\begin{aligned}
&\lim_{q\to 1} \frac{(1-q) f_L(q)}{\overline{F}_L(q)} \\
&= \lim_{q\to 1} \frac{(1-q)\int_0^\infty -f_X\left(x\left(s/w, F_Y^{\leftarrow}(q)\right)\right) D_2 x\left(s/w, F_Y^{\leftarrow}(q)\right) \mathrm{d}F_W(w)}{f_Y\left(F_Y^{\leftarrow}(q)\right)\int_0^\infty F_X\left(x\left(s/w, y\right)\right) \mathrm{d}F_W(w)} \\
&= \lim_{y\to\infty} \frac{\int_0^\infty -\overline{F}_Y(y)/f_Y(y) f_X\left(x\left(s/w, y\right)\right) D_2 x\left(s/w, y\right) \mathrm{d}F_W(w)}{\int_0^\infty F_X\left(x\left(s/w, y\right)\right) \mathrm{d}F_W(w)}.
\end{aligned} \tag{5.7.7}$$

Now we consider the setting of Assumption 5.2.8 and denote the integrands of (5.7.7) by

$$\begin{aligned}
h^*(w,y) &:= -\overline{F}_Y(y)/f_Y(y) f_X\left(x^*\left(s/w, y\right)\right) D_2 x^*\left(s/w, y\right) \quad \text{and} \\
h(w,y) &:= F_X\left(x^*\left(s/w, y\right)\right).
\end{aligned}$$

Choose $w \in (0,\infty)$ fixed, then Lemma 5.7.2(i) yields

$$\lim_{y\to\infty} \frac{h^*(w,y)}{h(w,y)} = \frac{\mu_X}{\nu_Y} \lim_{y\to\infty} \frac{F_X\left(x^*(s/w,y)\right) y D_2 x^*(s/w,y)/x^*(s/w,y)}{F_X\left(x^*(s/w,y)\right)}.$$

By Assumption 5.2.8(iii) we have $\lim_{y\to\infty} y D_2 x^*(s/w,y)/x^*(s/w,y) = \zeta$, hence

$$\lim_{y\to\infty} \frac{h^*(w,y)}{h(w,y)} = \zeta \frac{\mu_X}{\nu_Y},$$

for almost every $w \in (0,\infty)$. Similarly to (5.7.6), $h^*(w,y)$ is dominated by an integrable function $u(w)$ for all $q \in (q_0, 1)$. Therefore we can apply the *Dominated Convergence Theorem* and with (5.7.7) we get

$$\lim_{q\to 1} \frac{(1-q) f_L(q)}{\overline{F}_L(q)} = \lim_{y\to\infty} \frac{\int_0^\infty h^*(w,y)\,\mathrm{d}F_W(w)}{\int_0^\infty h(w,y)\,\mathrm{d}F_W(w)} = \zeta \frac{\mu_X}{\nu_Y}.$$

Hence, $L \in DA(\Psi_{\zeta\mu_X/\nu_Y})$.

In the setting of Assumption 5.2.9 the same result is obtained similarly using the asymptotic behavior of f_X and f_Y given in Lemma 5.7.2(ii). □

Proof of Theorem 5.2.11: We have $X \sim t_{\mu_X}$ and $Y_j \sim t_{\nu_Y}$. The t_ν density f_ν is given by

$$f_\nu(x) = C_\nu \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} = C_\nu |x|^{-\nu-1} \left(\frac{1}{\nu} + \frac{1}{x^2}\right)^{-(\nu+1)/2}, \quad C_\nu = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}.$$

We immediately obtain that for $x > 0$ the t_ν distribution function F_ν is bounded by

$$\begin{aligned}\underline{f}_\nu(x) := \frac{C_\nu}{\nu} x^{-\nu}\left(\frac{1}{\nu}+\frac{1}{x^2}\right)^{-(\nu+1)/2} &= \frac{C_\nu}{\nu}\left(x^{-2/(\nu+1)}+\frac{1}{\nu}x^{2-2/(\nu+1)}\right)^{-(\nu+1)/2} \\ \leq \ F_\nu(-x) = \overline{F}_\nu(x) \quad \leq \quad \frac{C_\nu}{\nu}x^{-\nu}\left(\frac{1}{\nu}\right)^{-(\nu+1)/2} &=: \overline{f}_\nu(x). \end{aligned} \tag{5.7.8}$$

Note that $\underline{f}_\nu(x) \sim \overline{f}_\nu(x)$ as $x \to \infty$. To get the asymptotic behavior of $\overline{F}_\nu$, we show that $\underline{f}_\nu^{\leftarrow}(q) \sim \overline{f}_\nu^{\leftarrow}(q)$ as $q \to 0$. First, we obtain

$$\overline{f}_\nu^{\leftarrow}(q) = C_\nu^{1/\nu}\nu^{(\nu-1)/(2\nu)}q^{-1/\nu}. \tag{5.7.9}$$

Straightforward calculation yields

$$\underline{f}_\nu\left(\overline{f}_\nu^{\leftarrow}(q)\right) \quad = \quad q\left(1+\left(\frac{\nu}{C_\nu}\right)^{2/\nu}q^{2/\nu}\right) \sim q, \quad \text{as } q \to 0,$$

hence

$$\underline{f}_\nu^{\leftarrow}(q) \sim \overline{f}_\nu^{\leftarrow}(q) \sim \overline{F}_\nu^{\leftarrow}(q) \quad \text{as } q \to 0. \tag{5.7.10}$$

Note that $x^*(s/w, y) = s/(aw) - yb/a$ and $F_Y^{\leftarrow}(q) = \overline{F}_Y^{\leftarrow}(1-q)$, therefore

$$\begin{aligned} F_X\left(x^*\left(\frac{s}{w}, F_Y^{\leftarrow}(q)\right)\right) &\sim C_{\mu_X}\mu_X^{(\mu_X-1)/2}\left(-\frac{s}{aw}+\frac{b}{a}C_{\nu_Y}^{1/\nu_Y}\nu^{(\nu_Y-1)/(2\nu_Y)}(1-q)^{-1/\nu_Y}\right)^{-\mu_X} \\ = \ (1-q)^{\mu_X/\nu_Y}C_{\mu_X}\mu_X^{(\mu_X-1)/2}&\left(-\frac{s}{aw}(1-q)^{1/\nu_Y}+\frac{b}{a}C_{\nu_Y}^{1/\nu_Y}\nu^{(\nu_Y-1)/(2\nu_Y)}\right)^{-\mu_X}. \end{aligned}$$

Note that this asymptotic behavior holds uniformly for all $w \in (0, \infty)$, since $x^*(s/w, y)$ is decreasing in w and the asymptotic behavior also holds for $w = \infty$. Applying (5.7.4) yields

$$\begin{aligned} P(L > q) &= \int_0^\infty F_X\left(x^*\left(\frac{s}{w}, F_Y^{\leftarrow}(q)\right)\right) \mathrm{d}F_W(w) \\ &\sim (1-q)^{\mu_X/\nu_Y}C_{\mu_X}\mu_X^{(\mu_X-1)/2}\int_0^\infty\left(-\frac{s}{aw}(1-q)^{1/\nu_Y}+\frac{b}{a}C_{\nu_Y}^{1/\nu_Y}\nu^{(\nu_Y-1)/(2\nu_Y)}\right)^{-\mu_X}\mathrm{d}F_W(w). \end{aligned} \tag{5.7.11}$$

We observe that

$$\lim_{y\to\infty}\frac{F_X(x^*(0,y))}{\overline{F}_Y(y)} = C\lim_{y\to\infty}(y)^{\nu_Y-\mu_X} = \begin{cases} \infty, & \mu_X < \nu_Y, \\ C, & \mu_X = \nu_Y, \\ 0, & \mu_X > \nu_Y, \end{cases}$$

for some constant $C < \infty$, i.e. the limit is finite, if X is not heavier-tailed than Y_j. Since $\mu_X \geq \nu_Y$, the upper limit is finite, hence Assumption 5.2.5(iv) is satisfied. We observe that the t-distribution falls into the setting of Assumption 5.2.8(i) and (ii). Considering Assumption 5.2.8(iii) we obtain $yD_2x^*(s/w,y)/x^*(s/w,y) = 1/\left(1 - s/(bwy)\right) \nearrow 1$ for all w pointwise and 1 is an integrable upper bound. Hence we can apply Theorem 5.2.10 and obtain $\kappa = \mu_X/\nu_Y$. Comparing this to (5.7.11) gives the desired result. □

For the proof of Theorem 5.2.12 we need the following lemma.

Lemma 5.7.4. *Let Φ and ϕ denote the distribution function and density of the standard normal distribution, respectively. Taking $0 < C < 1$, then for $x \geq \sqrt{1/(1-C)}$,*

$$C\phi(x)/x \leq \Phi(-x) \leq \phi(x)/x \quad \text{and} \quad C\phi(x)/x \leq \overline{\Phi}(x) \leq \phi(x)/x.$$

Moreover, Mill's Ratio holds: $\Phi(-x) = \overline{\Phi}(x) \sim \phi(x)/x$ as $x \to \infty$.

Proof: Gänssler and Stute (1977, Lemma 1.19.2) shows $(1/x - 1/x^3)\phi(x) \leq \Phi(-x) \leq \phi(x)/x$ for all $x > 0$. Then, $1/x - 1/x^3 \geq C/x$ holds for $x > 0$ if and only if $x \geq \sqrt{1/(1-C)}$. The limit relation follows from this as well. □

Proof of Theorem 5.2.12: We have $X, Y_j \overset{iid}{\sim} \mathcal{N}(0,1)$, $S_j = W(aX + bY_j)$. Let Φ and ϕ be the distribution function and density of the standard normal distribution, respectively. Applying (5.7.4) yields

$$P(L > q) = \int_0^\infty \Phi\left(s/(aw) - \Phi^{-1}(q)b/a\right) \mathrm{d}F_W(w). \tag{5.7.12}$$

Let $\varepsilon \in (0,1)$ and $x \leq -\sqrt{1/\varepsilon}$, then, by Lemma 5.7.4,

$$(1-\varepsilon)\phi(x)/|x| \leq \Phi(x) \leq \phi(x)/|x|. \tag{5.7.13}$$

The integrand of (5.7.12) increases in w, hence, for $\Phi^{-1}(q)b/a \geq \sqrt{1/\varepsilon}$,

$$\frac{\int_0^\infty \Phi\left(s/(aw) - \Phi^{-1}(q)b/a\right) \mathrm{d}F_W(w)}{\int_0^\infty \phi\left(s/(aw) - \Phi^{-1}(q)b/a\right) / \left|s/(aw) - \Phi^{-1}(q)b/a\right| \mathrm{d}F_W(w)} \in [1-\varepsilon, 1].$$

Therefore, as $q \to 1$,

$$\begin{aligned}
P(L > q) &= \int_0^\infty \Phi\left(s/(aw) - \Phi^{-1}(q)b/a\right) \mathrm{d}F_W(w) \\
&\sim \int_0^\infty \phi\left(s/(aw) - \Phi^{-1}(q)b/a\right) / \left|s/(aw) - \Phi^{-1}(q)b/a\right| \mathrm{d}F_W(w) \\
&= \left(\frac{\phi\left(\Phi^{-1}(q)\right)}{\Phi^{-1}(q)}\right)^{b^2/a^2} \int_0^\infty \exp\left(-\frac{s^2}{2aw} + \frac{sb\Phi^{-1}(q)}{a^2 w}\right) \frac{\left(\Phi^{-1}(q)\right)^{b^2/a^2}}{\left|\frac{s}{aw} - \frac{b}{a}\Phi^{-1}(q)\right|} \mathrm{d}F_W(w).
\end{aligned}$$

Note that $\Phi^{-1}(q) = \overline{\Phi}^{-1}(1-q)$. Then, again by Lemma 5.7.4, $\phi\left(\Phi^{-1}(q)\right)/\Phi^{-1}(q) \sim 1-q$, as $q \to 1$.

By Mill's Ratio, Lemma 5.7.2(ii) holds, hence $X, Y_j \overset{iid}{\sim} \mathcal{N}(0,1)$ fall into the setting of Assumption 5.2.9(i) and (ii) with $\mu_X = \nu_Y = 1/2$. We have $\zeta(w,y) = (b/a)(s/(aw) - yb/a)/y$ and $\lim_{y\to\infty} \zeta(w,y) = b^2/a^2$. Note that $|\zeta(w,y)| = c_1/(wy) + c_2 \le c_1/(wy_0) + c_2$ for some $c_1, c_2 < \infty$ and all $y \ge y_0$. As $E(1/W) < \infty$, $(c_1/(wy_0) + c_2)$ is an integrable upper bound, hence Assumption 5.2.9(iii) is satisfied. From Lemma 5.7.4 we obtain

$$\begin{aligned} \lim_{y\to\infty} \frac{F_X(x^*(0,y))}{\overline{F}_Y(y)} &= \lim_{y\to\infty} \frac{\overline{\Phi}(yb/a)}{\overline{\Phi}(y)} \\ &= \frac{a}{b} \lim_{y\to\infty} \exp\left(\frac{y^2}{2}\left(1 - \frac{b^2}{a^2}\right)\right) = \begin{cases} \infty, & b < a, \\ a/b, & b = a, \\ 0, & b > a. \end{cases} \end{aligned}$$

Hence, Assumption 5.2.5(iv) is satisfied and, obviously, also Assumptions 5.2.5(i)-(iii) and (v) are. Therefore, by Theorem 5.2.3, $P(L > q) = (1-q)^{b^2/a^2} \mathcal{L}(1/(1-q))$, where $\mathcal{L}$ satisfies for $q \to 1$ the relation

$$\mathcal{L}\left(\frac{1}{1-q}\right) \sim \int_0^\infty \exp\left(-\frac{s^2}{2aw} + \frac{sb\Phi^{-1}(q)}{a^2 w}\right) \frac{\left(\Phi^{-1}(q)\right)^{b^2/a^2}}{\left|\frac{s}{aw} - \frac{b}{a}\Phi^{-1}(q)\right|} \mathrm{d}F_W(w). \tag{5.7.14}$$

Choose $c > 1$ fixed and $x \ge \left(\sqrt{c}/(\sqrt{c}-1)\right)^{1/2}$, then

$$f_c(x) := c^{-1/2}\phi(x)/x \le \overline{\Phi}(x) \le \phi(x)/x =: f_1(x).$$

Therefore, since $\overline{\Phi}(x)$ is decreasing, $f_c^{-1}(q) \le \overline{\Phi}^{-1}(q) \le f_1^{-1}(q)$, $q_0 \le q < 1$, and some q_0. Taking logarithm, we want a solution of $\ln f_c(x) = \ln(1-q)$, i.e.

$$\frac{1}{2}x^2 + \ln x + \frac{1}{2}\ln(2\pi c) = -\ln(1-q).$$

By asymptotic expansion, similarly to Resnick (1987, Section 1.5, Example 2) we obtain

$$f_c^{-1}(1-q) = \sqrt{-2\ln(1-q)} - \frac{\ln(-\ln(1-q)) + \ln(4\pi c)}{2\sqrt{-2\ln(1-q)}} + o\left(1/\sqrt{-\ln(1-q)}\right). \tag{5.7.15}$$

Since $f_c^{-1}(1-q) \sim f_1^{-1}(1-q)$, as $q \to 1$, $f_1^{-1}(1-q) \sim \overline{\Phi}^{-1}(1-q)$ holds. Note that $f_1^{-1}(1-q) - f_c^{-1}(1-q) = O\left(1/\sqrt{-\ln(1-q)}\right) \overset{q\to 1}{\longrightarrow} 0$, hence also $\exp\left(-f_c^{-1}(1-q)\right) \sim \exp\left(-\overline{\Phi}^{-1}(1-q)\right)$ holds.

Substituting $\Phi^{-1}(q)$ in (5.7.14) by (5.7.15), we obtain the desired result. $\square$

Proof of Theorem 5.4.1: From (5.7.1) we have

$$\mathrm{Var}\sum_{j=1}^{m} L_j \;=\; 2\sum_{i\neq j} E(L_iL_j) + mp_{\mathrm{loss}}(1-mp_{\mathrm{loss}}), \tag{5.7.16}$$

where, since $F_\nu(s) := P(W_\nu s^*(X,Y_i)\le s) = p_{\mathrm{loss}}$ and $s<0$,

$$E(L_iL_j) = P\left(W_\nu s^*(X,Y_i)\le F_\nu^{\leftarrow}(p_{\mathrm{loss}}), W_\nu s^*(X,Y_j)\le F_\nu^{\leftarrow}(p_{\mathrm{loss}})\right).$$

As mentioned in the introduction, $(S_i^\nu, S_j^\nu) := (W_\nu s^*(X,Y_i), W_\nu s^*(X,Y_j))$ has a bivariate t_ν-distribution with correlation ρ. We apply now a dependence measure, called *lower tail-dependence coefficient*, defined by

$$\lambda \;:=\; \lim_{p\to 0} P\left(S_j^\nu \le F_\nu^{\leftarrow}(p)\,|S_i^\nu \le F_\nu^{\leftarrow}(p)\right).$$

Hult and Lindskog (2002) observed, that in case of a multivariate t_ν-distribution and with $\rho := \mathrm{Corr}(S_i^\nu, S_j^\nu)$

$$\lambda = \lambda(\nu) \;=\; \left(\int_{\arccos((1+\rho)/2)}^{\pi/2} \cos^\nu(v)\,\mathrm{d}v\right) \Big/ \left(\int_0^{\pi/2} \cos^\nu(v)\,\mathrm{d}v\right).$$

Kostadinov (2005, Lemma 2.2) shows that $\lambda(\nu)$ is strictly decreasing in ν. Let $\nu_1<\nu_2$, hence $\lambda(\nu_1)-\lambda(\nu_2) =: \varepsilon > 0$. Since $P\left(S_j^{\nu_i}\le F_{\nu_i}^{\leftarrow}(p)\,|S_i^{\nu_i}\le F_{\nu_i}^{\leftarrow}(p)\right)\to\lambda(\nu_i)$, $i=1,2$, there exists $p_\varepsilon>0$ such that for all $p\le p_\varepsilon$

$$\left|\lambda(\nu_i) - P\left(S_j^{\nu_i}\le F_{\nu_i}^{\leftarrow}(p)\,|S_i^{\nu_i}\le F_{\nu_i}^{\leftarrow}(p)\right)\right| < \frac{\varepsilon}{2}, \quad i=1,2.$$

Hence, for all $p\le p_\varepsilon$,

$$P\left(S_j^{\nu_1}\le F_{\nu_1}^{\leftarrow}(p)\,|S_i^{\nu_1}\le F_{\nu_1}^{\leftarrow}(p)\right) \;>\; P\left(S_j^{\nu_2}\le F_{\nu_2}^{\leftarrow}(p)\,|S_i^{\nu_2}\le F_{\nu_2}^{\leftarrow}(p)\right).$$

Since

$$\begin{aligned} &P\left(S_j^\nu \le F_\nu^{\leftarrow}(p_{\mathrm{loss}})\,|S_i^\nu\le F_\nu^{\leftarrow}(p_{\mathrm{loss}})\right)\\ &= \frac{1}{p_{\mathrm{loss}}} P\left(S_j^\nu \le F_\nu^{\leftarrow}(p_{\mathrm{loss}}),\, S_i^\nu\le F_\nu^{\leftarrow}(p_{\mathrm{loss}})\right) = \frac{1}{p_{\mathrm{loss}}} E(L_iL_j), \end{aligned}$$

also $E(L_iL_j)$ is decreasing in ν, if p_{loss} is sufficiently small. Applying (5.7.16), $\text{Var}\left(L^{(m)}\right)$ is decreasing in ν, hence $\text{Var}L$ is. $\square$

Proof of Corollary 5.5.1: We can rewrite

$$\begin{aligned} L_j^{\wedge} = \mathbf{1}_{\min\{aX,bY_j\}\leq s} &= \mathbf{1}_{\{aX\leq s\}\cup\{bY_j\leq s\}} \\ &= \mathbf{1}_{\{bY_j\leq s\}} + \left(1-\mathbf{1}_{\{bY_j\leq s\}}\right)\mathbf{1}_{\{aX\leq s\}}\,. \end{aligned}$$

Hence

$$\begin{aligned} L_{\wedge}^{(m)} &= \frac{1}{m}\sum_{j=1}^{m} L_j^{\wedge} = \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}_{\{bY_j\leq s\}} + \mathbf{1}_{\{aX\leq s\}}\frac{1}{m}\sum_{j=1}^{m}\left(1-\mathbf{1}_{\{bY_j\leq s\}}\right) \\ &=: \frac{1}{m}\sum_{j=1}^{m} B_j + \mathbf{1}_{\{aX\leq s\}}\frac{1}{m}\sum_{j=1}^{m}(1-B_j)\,, \end{aligned}$$

where

$$\begin{aligned} B_1, B_2, \ldots &\overset{iid}{\sim} \text{Ber}\left(F_X\left(s/b\right)\right) \text{ and} \\ (1-B_1),(1-B_2),\ldots &\overset{iid}{\sim} \text{Ber}\left(1-F_X\left(s/b\right)\right) \end{aligned}$$

are iid Bernoulli sequences. Therefore, for $m\to\infty$, $L^{(m)}$ converges almost surely to $F_X\left(s/b\right)+\left(1-F_X\left(s/b\right)\right)\mathbf{1}_{\{aX\leq s\}}$. $\square$

Proof of Theorem 5.5.2:

<u>*(i):*</u> $X, Y_1 \overset{iid}{\sim} t_\nu$ are regularly varying on $\mathbb{R}^-$, i.e. $F_{t_\nu}(-\cdot)\in\mathcal{R}_{-\nu}$. Hence

$$P\left(aX+bY_1\leq s\right)\sim P\left(\min\left\{aX,bY_1\right\}\leq s\right)\,,\quad s\to-\infty\,,$$

see for instance Example 3.2 in combination with Goldie and Klüppelberg (1998, Definition 1.1). For convenience we define $A := aX$ and $B := bY_1$. Then, A and B are independent in $\mathcal{R}_{-\nu}$ satisfying $P(A>s)/P(B>s)\overset{s\to\infty}{\longrightarrow} c\in(0,\infty)$ and $P(A\leq s)/P(B\leq s)\overset{s\to-\infty}{\longrightarrow} c\in(0,\infty)$. Let F_A and F_B denote the dfs of A and B, respectively. Writing $x\wedge y := \max\{x,y\}$ and $x\vee y := \max\{x,y\}$ we have

$$\begin{aligned} P\left(L_1^{\wedge}=1\,|L_1=1\right) &= P\left(A\wedge B<-s\,|A+B<-s\right) \\ &= P\left(A\vee B>s\,|A+B>s\right) = \frac{P\left(A\vee B>s,\ A+B>s\right)}{P\left(A+B>s\right)} \\ &= 1-\frac{P\left(A\vee B\leq s,\ A+B>s\right)}{P\left(A+B>s\right)}\,. \end{aligned} \tag{5.7.17}$$

For illustration purposes see Figure 5.5. The set $\{(x,y) \in \mathbb{R}^2 : \ x \vee y > s, \ x + y > s\}$ is the hatched area in Figure 5.5 above the lines $\{x + y = s\}$ and $\{x \vee y = s\}$; the set $\triangle := \{(x,y) : \ x \vee y \leq s, \ x + y > s\}$ is the triangle with edges $(s,0)$, (s,s) and $(0,s)$. Let $\|(x,y)\|_1 := |x| + |y|$ denote the 1-norm, then for any $\varepsilon > 0$ it holds that

$$\begin{aligned} \triangle &= \{(x,y) : \ x \vee y \leq s, \ x + y > s\} \\ &\subset \ \left\{(x,y) : \ \varepsilon < \frac{y}{x} < \frac{1}{\varepsilon}, \ \|(x,y)\|_1 > s\right\} \\ &\quad \cup \{(x,y) : \ (1-\varepsilon)s < x < s, \ 0 < y < \varepsilon s\} \\ &\quad \cup \{(x,y) : \ 0 < x < \varepsilon s, \ (1-\varepsilon)s < y < s\} \\ &=: \ \boldsymbol{S}_1 \cup \boldsymbol{S}_2 \cup \boldsymbol{S}_3, \end{aligned} \tag{5.7.18}$$

where $\boldsymbol{S}_1$ can be identified in Figure 5.5 as the set between the two lines through the points $(0,0)$, $(s, \varepsilon s)$ and $(0,0)$, $(\varepsilon s, s)$ and above the line $\{x + y = s\}$; the sets $\boldsymbol{S}_2$ and $\boldsymbol{S}_3$ represent in the figure the two small rectangles.

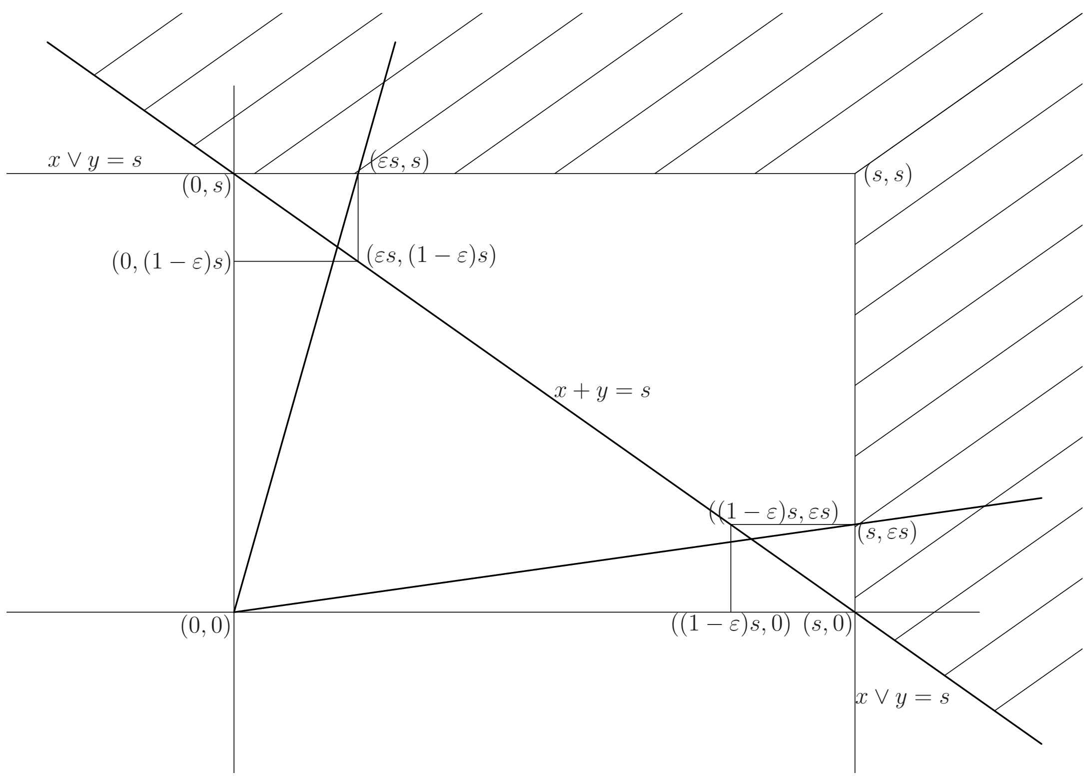

Figure 5.5: Illustration of (5.7.17) and (5.7.18).

By Resnick (2004, section 4.1 and 4.3), the vector (A, B) is bivariate regularly varying.

More precisely, let $\|\cdot\|$ be any norm on $\mathbb{R}^2$, then

$$\frac{P\left(\|(A,B)\| \geq t,\, (A,B)/\|(A,B)\| \in \cdot\right)}{P\left(\|(A,B)\| \geq t\right)} \overset{t\to\infty}{\longrightarrow} \Theta(\cdot).$$

Θ is a measure on the unit simplex $\mathfrak{S} = \{\boldsymbol{s} \in \mathbb{R}^2 : \|\boldsymbol{s}\| = 1\}$ called *spectral measure*. Since A and B are independent, Θ is concentrated on $(1,0)$, $(-1,0)$, $(0,1)$ and $(0,-1)$. Note that, using the 1-norm, symmetry of A and B yields

$$\frac{P\left(\|(A,B)\|_1 > s\right)}{P\left(A+B > s\right)} < \frac{P\left(\|(A,B)\|_1 > s\right)}{P\left(A+B > s,\, A,B > 0\right)} = 4.$$

Hence, for any $\varepsilon > 0$, we have

$$P\left((A,B) \in \boldsymbol{S}_1 \,|A+B > s\right) = \frac{P\left(A+B > s,\, \varepsilon < B/A < 1/\varepsilon\right)}{P\left(A+B > s\right)} \tag{5.7.19}$$

$$= \frac{P\left(A,B>0,\, \varepsilon < B/A < 1/\varepsilon,\, \|(A,B)\|_1 > s\right)}{P\left(\|(A,B)\|_1 > s\right)} \frac{P\left(\|(A,B)\|_1 > s\right)}{P\left(A+B > s\right)}$$

$$< 4\frac{P\left(A,B>0,\, \varepsilon < B/A < 1/\varepsilon,\, \|(A,B)\|_1 > s\right)}{P\left(\|(A,B)\|_1 > s\right)} \overset{s\to\infty}{\longrightarrow} 4\Theta\left(\mathcal{S}_1^n\right) = 0,$$

since $\mathcal{S}_1^n := \{\boldsymbol{s}/\|\boldsymbol{s}\|_1 : \boldsymbol{s} \in \mathcal{S}_1\}$ has no points on the axes. Considering rectangle $\boldsymbol{S}_2$ we obtain as $s \to \infty$ using again $P(A+B > s) \sim P(A \vee B > s)$

$$P\left((A,B) \in \boldsymbol{S}_2 \,|A+B > s\right) \leq \frac{P\left((1-\varepsilon)s < A < s\right) P(0 < B < \varepsilon s)}{P\left(A+B > s\right)}$$

$$\sim \frac{P\left((1-\varepsilon)s < A < s\right) P(0 < B < \varepsilon s)}{P\left(A \vee B > s\right)}$$

$$= \frac{\left(F_A(s) - F_A((1-\varepsilon)s)\right)\left(F_B(\varepsilon s) - F_B(0)\right)}{\overline{F}_A(s) + \overline{F}_B(s) - \overline{F}_A(s)\overline{F}_B(s)}$$

$$= \frac{s^{-\nu}\left((1-\varepsilon)^{-\nu}\mathcal{L}_A((1-\varepsilon)s) - \mathcal{L}_A(s)\right)\left(\frac{1}{2} - s^{-\nu}\varepsilon^{-\nu}\mathcal{L}_B(\varepsilon s)\right)}{s^{-\nu}\left(\mathcal{L}_A(s) + \mathcal{L}_B(s)\right) - s^{-2\nu}\mathcal{L}_A(s)\mathcal{L}_B(s)}$$

$$\longrightarrow \frac{1}{2(1+c)}\left((1-\varepsilon)^{-\nu} - 1\right) < \frac{\nu\varepsilon}{(1+c)}\,. \tag{5.7.20}$$

The last convergence holds since $F_A(s) = 1 - (s)^{-\nu}\mathcal{L}_A(s)$, $F_B(s) = 1 - (s)^{-\nu}\mathcal{L}_B(s)$, $\mathcal{L}_B(s)/\mathcal{L}_A(s) \overset{s\to\infty}{\longrightarrow} c$ and $\mathcal{L}_A, \mathcal{L}_B \in \mathcal{R}_0$; $(1-\varepsilon)^{-\nu} - 1 < 2\nu\varepsilon$ holds for ε small enough since $(1-\varepsilon)^{-\nu} - 1 \sim \nu\varepsilon$ as $\varepsilon \to 0$. Combining (5.7.19) and (5.7.20), (5.7.18) yields

$$\lim_{s\to\infty} \frac{P\left(A \vee B \leq s,\, A+B > s\right)}{P\left(A+B > s\right)} < K\varepsilon\,, \quad \forall\varepsilon > 0\,,$$

for some constant $0 < K < \infty$. Hence the latter limit equals 0 and applying this to (5.7.17) yields

$$P\left(L_1 = 1\,|L_1^{\wedge} = 1\right) \overset{s\to-\infty}{\longrightarrow} 1.$$

Therefore, for all m and any $q \in \{0, 1/m, 2/m, \ldots, 1\}$, we conclude

$$P\left(L^{(m)} = q \,|\, L_{\wedge}^{(m)} = q\right) \overset{s \to -\infty}{\longrightarrow} 1.$$

<u>*(ii):*</u> Now we consider the limiting case $m \to \infty$. Recall that

$$L \overset{\mathrm{d}}{=} F_{t_\nu}(y^*(s, X)) = F_{t_\nu}\left((s - aX)/b\right)$$

and from Corollary 5.5.1 we know that $L^{\wedge} \in \{F_{t_\nu}(s/b), 1\}$ holds. Further, $L^{\wedge} = F_{t_\nu}(s/b)$ if $s - aX < 0$. Define $\mathcal{B}_\varepsilon(x) := (x - \varepsilon, x + \varepsilon)$, then

$$\begin{aligned} & P\left(L \in \mathcal{B}_\varepsilon\left(F_{t_\nu}(s/b)\right) \,|\, L^{\wedge} = F_{t_\nu}(s/b)\right) \\ \geq \; & P\left(s - aX \leq bF_{t_\nu}^{\leftarrow}\left(F_{t_\nu}(s/b)(1+\varepsilon)\right) \,\big|\, s - aX < 0\right) \\ & -P\left(s - aX < bF_{t_\nu}^{\leftarrow}\left(F_{t_\nu}(s/b)(1-\varepsilon)\right) \,\big|\, s - aX < 0\right). \end{aligned} \tag{5.7.21}$$

By (5.7.9) and (5.7.10) we have $F_{t_\nu}^{\leftarrow}(qc) \sim F_{t_\nu}^{\leftarrow}(q)c^{-1/\nu}$ as $q \to 0$, hence

$$bF_{t_\nu}^{\leftarrow}\left(F_{t_\nu}(s/b)(1 \pm \varepsilon)\right) \sim (1 \pm \varepsilon)^{-1/\nu} s \quad \text{as } s \to -\infty.$$

From $s - aX \leq (1 \pm \varepsilon)^{-1/\nu} s$ follows $aX \geq \left(1 - (1 \pm \varepsilon)^{-1/\nu}\right) s$. Since $1 - (1 - \varepsilon)^{-1/\nu} < 0$, we conclude (as $s \to -\infty$)

$$\begin{aligned} & P\left(s - aX < bF_{t_\nu}^{\leftarrow}\left(F_{t_\nu}(s/b)(1-\varepsilon)\right) \,\big|\, s - aX < 0\right) \\ \sim \; & P\left(aX > \left(1 - (1-\varepsilon)^{-1/\nu}\right) s \big|\, aX > s\right) \longrightarrow 0. \end{aligned}$$

Since $1 - (1 + \varepsilon)^{-1/\nu} > 0$, we conclude (as $s \to -\infty$)

$$\begin{aligned} & P\left(s - aX < bF_{t_\nu}^{\leftarrow}\left(F_{t_\nu}(s/b)(1+\varepsilon)\right) \,\big|\, s - aX < 0\right) \\ \sim \; & P\left(aX \geq \left(1 - (1+\varepsilon)^{-1/\nu}\right) s \big|\, aX > s\right) \longrightarrow 1. \end{aligned}$$

Therefore, (5.7.21) converges to 1 as $s \to -\infty$ for all $\varepsilon > 0$.

We have $L^{\wedge} = 1$ if $aX \leq s$, hence, similarly to (5.7.21),

$$\begin{aligned} & P(L > 1 - \varepsilon |\, L^{\wedge} = 1) = P\left(aX \leq s - bF_{t_\nu}^{\leftarrow}(1-\varepsilon) \big|\, aX \leq s\right) \\ = \; & \frac{F_{t_\nu}\left(\left(s - bF_{t_\nu}^{\leftarrow}(1-\varepsilon)\right)/a\right)}{F_{t_\nu}(s/a)}. \end{aligned}$$

From (5.7.8) we know $F_{t_\nu}(-x) \sim Cx^{-\nu}$ as $x \to \infty$ for some constant C. Hence, $P(L > 1 - \varepsilon |\, L^{\wedge} = 1) \to 1$ as $s \to -\infty$ for all $\varepsilon > 0$. □

Bibliography

Anderson, T. (2003). *An introduction to multivariate statistical analysis* (Third ed.). Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley.

Basrak, B., R. Davis, and T. Mikosch (2002). Regular variation of GARCH processes. *Stochastic Process. Appl. 99*(1), 95–115.

Bentler, P. and P. Dudgeon (1996). Covariance structure analysis: Statistical practice, theory and directions. *Annu. Rev. Psychol. 47*, 563–593.

Bingham, N., C. Goldie, and J. Teugels (1989). *Regular variation*, Volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press.

Bluhm, C., L. Overbeck, and C. Wagner (2003). *An introduction to credit risk modeling*. London: Chapman & Hall.

Browne, M. (1982). Covariance structures. In D. Hawkins (Ed.), *Topics in applied multivariate analysis*. Cambridge University Press.

Browne, M. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British J. Math. Statist. Psych. 37*(1), 62–83.

Browne, M. and A. Shapiro (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. *Linear Algebra Appl. 82*, 169–176.

Browne, M. and A. Shapiro (1987). Adjustments for kurtosis in factor analysis with elliptically distributed errors. *J. Roy. Statist. Soc. Ser. B 49*(3), 346–352.

Cambanis, S., S. Huang, and G. Simons (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal. 11*(3), 368–385.

Casella, G. and R. Berger (2001). *Statistical inference* (Second ed.). Belmont, CA: Duxbury Press.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer.

Coles, S. and J. Tawn (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B 53*(2), 377–392.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1*(2), 223–236.

Dacorogna, M., H. Hauksson, T. Domenig, U. Müller, and G. Samorodnitsky (2001). Multivariate extremes, aggregation and risk estimation. CeNDEF Workshop Papers, January 2001 P2, Universiteit van Amsterdam, Center for Nonlinear Dynamics in Economics and Finance. Available at `http://ideas.repec.org/p/ams/cdws01/p2.html`.

Danielsson, J., L. de Haan, L. Peng, and C. de Vries (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal. 76*(2), 226–248.

de Haan, L. and J. de Ronde (1998). Sea and wind: multivariate extremes at work. *Extremes 1*(1), 7–45.

de Haan, L. and L. Peng (1998). Comparison of tail index estimators. *Statist. Neerlandica 52*(1), 60–70.

de Haan, L. and S. Resnick (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 40*(4), 317–337.

de Haan, L. and S. Resnick (1993). Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models 9*(2), 275–309.

de Haan, L. and U. Stadtmüller (1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A 61*(3), 381–395.

Demarta, S. and A. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*(1), 111–129.

Draisma, G., L. de Haan, L. Peng, and T. Pereira (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes 2*(4), 367–404.

Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal. 64*(1), 25–47.

Drees, H. and E. Kaufmann (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl. 75*(2), 149–172.

Einmahl, J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Process. Appl. 70*(1), 31–58.

Einmahl, J., L. de Haan, and X. Huang (1993). Estimating a multidimensional extreme-value distribution. *J. Multivariate Anal. 47*(1), 35–47.

Einmahl, J., L. de Haan, and D. Li (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist. 34*(4).

Einmahl, J., L. de Haan, and V. Piterbarg (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist. 29*(5), 1401–1423.

Einmahl, J., L. de Haan, and A. Sinha (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl. 70*(2), 143–171.

Embrechts, P., L. de Haan, and X. Huang (2000). Modelling multivariate extremes. In P. Embrechts (Ed.), *Extremes and integrated risk management*, pp. 59–67. UBS Warburg and Risk Books.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling extremal events for insurance and finance*. Berlin: Springer.

Embrechts, P., F. Lindskog, and A. McNeil (2003). Modelling dependence with copulas and applications to risk management. In S. Rachev (Ed.), *Handbook of heavy tailed distributions in finance*, Chapter 8, pp. 329–384. Amsterdam: Elsevier.

Falk, M., J. Hüsler, and R.-D. Reiss (2004). *Laws of small numbers: extreme and rare events* (Second ed.). Lecture Notes in Statistics. Basel: Birkhäuser.

Fang, K., S. Kotz, and K. Ng (1990). *Symmetric multivariate and related distributions*, Volume 36 of *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

Frahm, G., M. Junker, and R. Schmidt (2005). Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance Math. Econom. 37*(1), 80–100.

Frahm, G., M. Junker, and A. Szimayer (2003). Elliptical copulas: applicability and limitations. *Statist. Probab. Lett. 63*(3), 275–286.

Frey, R. and A. McNeil (2000). Modelling dependent defaults. Available at `http://www.defaultrisk.com/pp_corr_17.htm`.

Frey, R. and A. McNeil (2002). Var and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. *Journal of Banking & Finance 26*(7), 1317–1334. Available at `http://ideas.repec.org/a/eee/jbfina/v26y2002i7p1317-1334.html`.

Frey, R. and A. McNeil (2003). Dependent defaults in models of portfolio credit risk. *Journal of Risk 6*(1), 59–92.

Frey, R., A. McNeil, and M. Nyfeler (2001, October). Copulas and credit models. *RISK*, 111–114.

Galambos, J. (1987). *The asymptotic theory of extreme order statistics* (Second ed.). Melbourne: Krieger Publishing.

Gänssler, P. and W. Stute (1977). *Wahrscheinlichkeitstheorie*. Berlin: Springer.

Geluk, J. and L. de Haan (1987). *Regular variation, extensions and Tauberian theorems*, Volume 40 of *CWI Tract*. Amsterdam: Stichting Mathematisch Centrum Centrum voor Wiskunde en Informatica.

Goldie, C. M. and C. Klüppelberg (1998). Subexponential distributions. In *A practical guide to heavy tails*, pp. 435–459. Boston: Birkhäuser.

Gumbel, E. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc. 55*, 698–707.

Gupton, G., C. Finger, and M. Bhatia (1997). Creditmetrics – technical document. Technical report, Morgan Guaranty Trust Company.

Heffernan, J. and J. Tawn (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B 66*(3), 497–546.

Higham, N. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal. 22*(3), 329–343.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist. 3*(5), 1163–1174.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics 19*, 293–325.

Hsing, T., C. Klüppelberg, and G. Kuhn (2004). Dependence estimation and visualization in multivariate extremes with applications to financial data. *Extremes 7*(2), 99–121.

Hu, L., P. Bentler, and Y. Kano (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin 112*, 351–362.

Huang, X. (1992). *Statistics of bivariate extremes*. Ph. D. thesis, Erasmus University Rotterdam, Tinbergen Research Institute, no. **22**.

Hult, H. and F. Lindskog (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. in Appl. Probab. 34*(3), 587–608.

Joe, H. (1997). *Multivariate models and dependence concepts*, Volume 73 of *Monographs on Statistics and Applied Probability*. London: Chapman & Hall.

Kendall, M. and J. Gibbons (1990). *Rank correlation methods* (Fifth ed.). London: Edward Arnold.

Klüppelberg, C. and G. Kuhn (2006). Copula structure analysis based on robust and extreme dependence measures. Submitted.

Klüppelberg, C., G. Kuhn, and L. Peng (2005a). Estimating the tail dependence of an elliptical distribution. Submitted.

Klüppelberg, C., G. Kuhn, and L. Peng (2005b). Multivariate tail copula: modeling and estimation. Submitted.

Kostadinov, K. (2005). Non-parametric estimation of elliptical copula with application to credit risk. Preprint, Munich University of Technology.

Kuhn, G. (2002). Multivariate value at risk schätzung von zins swap sätzen. Master's thesis, Munich University of Technology.

Kuhn, G. (2005). Tails of credit default portfolios. Submitted.

Lawley, D. and A. Maxwell (1971). *Factor analysis as a statistical method* (Second ed.). New York: Elsevier.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extremes and related properties of random sequences and processes*. New York: Springer.

Ledford, A. and J. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*(1), 169–187.

Ledford, A. and J. Tawn (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B 59*(2), 475–499.

Lee, A. (1990). *U-statistics*, Volume 110 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker Inc.

Lee, S. and P. Bentler (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure modules. *South African Statist. J. 14*(2), 121–136.

Lindskog, F., A. McNeil, and U. Schmock. Kendall's tau for elliptical distributions. In *Credit risk – measurement, evaluation and management*. Heidelberg: Physica-Verlag.

Ling, S. and L. Peng (2004). Hill's estimator for the tail index of an ARMA model. *J. Statist. Plann. Inference 123*(2), 279–293.

Lucas, A., P. Klaasen, P. Spreij, and S. Straetmans (2003). Tail behaviour of credit loss distribution for general latent factor models. *Applied Mathematical Finance 10*(4), 337–357.

McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative risk management*. Princeton University Press.

Muirhead, R. and C. Waternaux (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika 67*(1), 31–43.

Nelsen, R. (1999). *An introduction to copulas*, Volume 139 of *Lecture Notes in Statistics*. New York: Springer.

Pickands, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, Volume 49, pp. 859–878 and 894–902.

Qi, Y. (1997). Almost sure convergence of the stable tail empirical dependence function in multivariate extreme statistics. *Acta Math. Appl. Sinica (English Ser.) 13*(2), 167–175.

Resnick, S. (1987). *Extreme values, regular variation, and point processes*, Volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. New York: Springer-Verlag.

Resnick, S. (2004). On the foundations of multivariate heavy-tail analysis. *J. Appl. Probab. 41A*, 191–212.

Satorra, A. and P. Bentler (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika 66*(4), 507–514.

Schlather, M. and Tawn (2000). Properties of extremal coefficients. Technical Report ST 00-01, Department of Mathematics and Statistics, Lancaster University.

Schlather, M. and J. A. Tawn (2003). A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika 90*(1), 139–156.

Schmidt, R. (2003). Credit risk modelling and estimation via elliptical copulae. In Bohl, Nakhaeizadeh, Rachev, Ridder, and Vollmer (Eds.), *Credit risk: measurement, evaluation and management*, pp. 267–289. Heidelberg: Physica Verlag.

Schmidt, R. and U. Stadtmüller (2006). Non-parametric estimation of tail dependence. *Scand. J. Statist. 33*, 307–335.

Sklar, A. (1996). Random variables, distribution functions, and copulas—a personal look backward and forward. In *Distributions with fixed marginals and related topics (Seattle, WA, 1993)*, Volume 28 of *IMS Lecture Notes Monogr. Ser.*, pp. 1–14. Hayward, CA: Inst. Math. Statist.

Smith, K. (1983). *Primer of modern analysis* (Second ed.). Undergraduate texts in mathematics. New York: Springer.

Smith, R. (2003). Statistics of extremes, with applications in environment, insurance and finance. In B. Finkenstädt and H. Rootzén (Eds.), *Extreme Values in Finance, Telecommunications and the Environment*, Chapter 1, pp. 1–78. Chapman & Hall/CRC Press.

Steiger, J. (1994). Factor analysis in the 1980's and the 1990's: some old debates and some new developments. In I. Borg and P. Mohler (Eds.), *Trends and perspectives in empirical social research*, pp. 201–224. Walter de Gruyter, Berlin.

Steiger, J., A. Shapiro, and M. Browne (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika 50*(3), 253–264.

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

van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes.* New York: Springer.

van Praag, B., T. Dijkstra, and J. van Velzen (1985). Least-squares theory based on general distributional assumptions with an application to the incomplete observations problem. *Psychometrika 50*(1), 25–36.

Besser ist aufhören,
denn überfüllen.

Die Klinge immerfort geschärft,
bleibt nicht lange Klinge.

Der Saal mit Gold und Jade vollgestopft,
ist nicht vor Räubern zu bewahren.

Glanz und Ehren mit Hochmut gepaart,
ziehn sich selbst ins Verderben.

Zurückziehn nach getanem Werk.
so ist das Dau des Himmels.
Laotse