Portfolio Credit Risk Modelling With Heavy-Tailed Risk Factors

Krassimir Kolev Kostadinov

Munich University of Technology

Center of Mathematical Sciences

GKAAM

Portfolio Credit Risk Modelling With Heavy-Tailed Risk Factors

Krassimir Kostadinov

2006

Acknowledgments

It is a pleasure for me to thank my advisor Prof. Dr. Claudia Klüppelberg for her patience and for her support in all aspects of work and life.

The financial support of the Deutsche Forschungsgemeinschaft through the Graduate Program "Applied Algorithmic Mathematics" is gratefully acknowledged.

Special thanks to all my colleagues at the graduate program and at the department for Statistics at the Munich University of Technology for their encouragement throughout my three years there. Thanks to Gabriel Kuhn and Dr. Martin Hillenbrand for investing their time and efforts in the area of credit risk modelling and for the numerous fruitful discussions. Thanks to Christian Schwarz and the other students from the seminar on credit risk modelling in 2004 for their work and for their challenging questions.

Thanks to Dr. Christian Bluhm for introducing me to the newest practical problems in the area of credit risk management. Thanks to Prof. Dr. Philip Jorion and to the anonymous referee from *The Journal of Risk* for their helpful comments and the suggested improvements to my paper. Thanks to Dr. Liang Peng for the careful proofreading and the excellent advises and references he gave me.

I got (unexpectedly for me) really a lot of feedback on my work, from practitioners, students and researchers at all levels. Thank you all.

Nothing could have been done without the love and support of my family.

Abstract

During the last decade, the dependencies between financial assets have increased due to globalization effects and relaxed market regulation. The standard industrial methodologies like RiskMetrics [117] and CreditMetrics [74] model the dependence structure in the derivatives or in the credit portfolio by assuming multivariate normality of the underlying risk factors. It has been well recognized that many financial assets exhibit a number of features which contradict the normality assumption – namely asymmetry, skewness and heavy tails. Moreover, asset return data suggests also a dependence structure which is quite different from the Gaussian. Recent empirical studies indicate that especially during highly volatile and bear markets the probability for joint extreme events leading to simultaneous losses in a portfolio could be seriously underestimated under the normality assumption. Theoretically, Embrechts et al. [48] show that the traditional dependence measure (the linear correlation coefficient) is not always suited for a proper understanding of the dependency in financial markets. When it comes to measuring the dependence between extreme losses, other measures (e.g. the tail dependence coefficient) are more appropriate. This is particularly important in the credit risk framework, where the risk factors actually enter the model only to introduce a dependence structure in the portfolio. Clearly, appropriate multivariate models suited for extreme events are needed.

In this thesis, we consider a portfolio credit risk model in the spirit of CreditMetrics [74]. With respect to the marginal losses, we retain and enhance all features of that model and we incorporate not only the default risk, but also the rating migrations, the credit spread volatility and the recovery risk. The dependence structure in the portfolio is given by a set of underlying risk factors which we model by a general multivariate elliptical distribution. On the one hand, this model retains the standard Gaussian model as a special case. On the other hand, by introducing a heavy-tailed "global shock" affecting the credits simultaneously across regions and business sectors, we obtain a more flexible model for joint extreme losses.

The goals of the thesis are twofold.

First, we consider the calibration of the model. The main result is a new method for statistical estimation of the dependence structure (the copula) of a random vector with arbitrary marginals and elliptical copula. Within our method, we calibrate the linear correlation coefficients using the whole available sample of observations and the non-linear (tail) dependence coefficients using only the extreme observations. Special attention is put to the estimation of the tail dependence coefficients, where additional results aiming at a

lower variance of the estimates are provided. The particular application of the method to the calibration of the credit risk model is given in detail, and several simulation studies and real data examples are presented.

Second, we investigate the portfolio loss distribution. In particular, we derive an upper bound of its tail, which is especially accurate at high loss levels. Given the complexity of our model, we obtain this result using a mixture of analytic techniques and Monte Carlo simulation. An approximation of the Value-at-Risk and a new method to determine the contributions of the individual credits to the overall portfolio risk is provided. The impact of the heavy-tailed model on the overall portfolio risk and on the risk structure as given by the risk contributions is investigated. We conclude that the heavy-tailed assumption has important consequences in all aspects of risk management.

iv

Contents

Chapter 1

Introduction

1.1 Basics of portfolio credit risk modelling

Credit risk is a broad category of risks faced by the financial institutions. It includes (see Duffie and Singleton [40], Chapter 1):

(1) **default risk** – the risk that an obligor will not be able to meet in time its financial obligations;

(2) **recovery risk** – the risk that in case of default the market value of the collateral of an obligor will not be sufficient to cover at least the expected part of the losses;

(3) **credit quality risk** – the risk that the credit quality of an obligor, measured by some internal or external for the financial institution rating system, will decrease;

(4) **credit spread risk** – the risk of a reduction in the market value of a credit security (e.g. corporate bond or default swap).

Historically banks have been managing credit risk by imposing rigorous underwriting standards, credit limit enforcement and one-by-one obligor monitoring. This has changed dramatically since the late 1990s, when portfolio credit risk has become a key risk management challenge (see Allen and Saunders [1], Chapter 1). The amount of credit risk taken by the banks has increased, and thus the need for more sophisticated techniques. At the same time, credit derivatives like swaps and forwards have become very popular, leading to globalization of the credit exposures. In particular, collateralised debt obligations (CDOs) and other credit risk securities transferring the risk of a whole credit portfolio have appeared on the market. Last but not least, portfolio credit risk started to play a crucial role in the determination of capital requirements (see BIS [9]).

The primary reason to have a quantitative portfolio approach to credit risk is that one can more systematically address the concentration risk. Concentration risk refers to the additional portfolio risk resulting from an increased exposure to one obligor or groups of dependent obligors, e.g. by industry, by location, etc. During the last decade, the de-
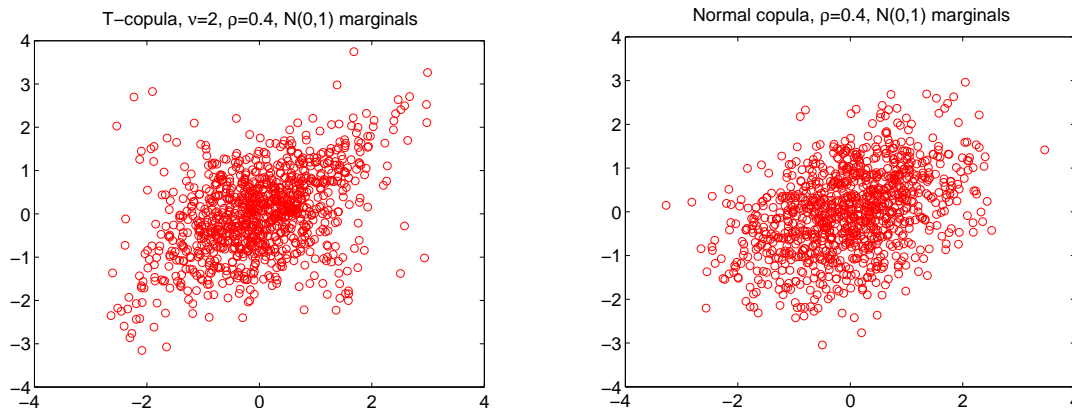
Figure 1.1: The linear correlation is not enough.
Left plot: 1000 i.i.d. copies of a bivaraite random vector with standard normal $N(0,1)$ marginals, correlation $\rho = 0.4$ and t-copula with $\nu = 4$ degrees of freedom (see Example 3.2.1).
Right plot: 1000 i.i.d. copies of a bivariate Gaussian random vector with $N(0,1)$ marginals and correlation $\rho = 0.4$. Obviously there are much more joint extremes in the left plot, indicating a non-linear dependence.
The joint extremes are quite important in market VaR analysis and in credit risk.

pendencies between the financial assets in general have increased due to globalization effects and relaxed market regulation. The standard industrial methodologies like Risk-Metrics [117] and CreditMetrics [74] model the dependence structure in the derivatives or in the credit portfolio by assuming multivariate normality of the underlying risk factors. It has been well recognized (see Mandelbrot [107] for a classsical, or Cont [24] for a recent study), that many financial assets exhibit a number of features which contradict the normality assumption – namely asymmetry, skewness and heavy tails. Moreover, asset returns data suggests also a dependence structure which is quite different from the Gaussian (see Fortin and Kuzmics [56]). In particular, empirical studies like Junker and May [85] and Malevergne and Sornette [106] indicate that, especially during highly volatile and bear markets, the probability for joint extreme events leading to simultaneous losses in a portfolio could be seriously underestimated under the normality assumption. Theoretically, Embrechts et al. [48] show that the traditional dependence measure (the linear correlation coefficient) is not always suited for a proper understanding of the dependency in financial markets. When it comes to measuring the dependence between extreme losses, other measures (e.g. the tail dependence coefficient) are more appropriate. In the credit risk framework, Frey et al. [62] provide examples and insight on the impact of a violated Gaussian assumption on the tail of the credit portfolio loss distribution. Holding the marginal losses of the individual credits fixed and introducing tail dependence through heavy-tailed risk factors, Frey et al. [62] conclude that the overall portfolio risk increases

drastically as compared to the Gaussian (tail-independent) case. Clearly, multivariate models suitable for extreme events are needed.

Even when the dependence in the credit portfolio is modelled through Gaussian risk factors, the special structure of models like CreditMetrics [74] makes the computation of the portfolio loss distribution an analytically intractable problem. Monte Carlo simulation is extensively used in such cases. However, this method is computationally intensive, in particular when one has to determine the probability for an extremely high loss in a high-dimensional portfolio (see Glasserman and Li [70]). Furthermore, for risk management purposes it is important to measure also the contributions of the marginal credits to the overall portfolio risk. CreditMetrics [74] suggest the so called 'marginal standard deviation' as a risk measure, as this quantity can be computed by simple numerical methods. This approach has numerous drawbacks, see for instance Overbeck [122], for a particular discussion in the credit risk framework, or Embrechts et al. [48] for a more general discussion. On the other hand, estimating the marginal risk contributions with respect to a down-sided risk measure like Value-at-Risk (VaR) or Expected Shortfall (ES) by simulation requires even greater computational effort (see Overbeck [122] or Merino and Nyfeler [114]). Current solutions to these problems are improved Monte Carlo methods (Merino and Nyfeler [114], Glasserman and Li [70]) or analytic approximations (Glasserman [67], Kuhn [94]).

In Section 1.2 we present the main concepts of the CreditMetrics model, which is the base of most of the portfolio credit risk models used in practice. Then, in Section 1.3 we describe briefly the generalizations which we suggest. We conclude the first chapter with a summary of the main results to follow.

1.2 The CreditMetrics model

CreditMetrics is a methodology based on the estimation of the distribution of the changes in the value (profits or losses) of a credit risk portfolio over a given time horizon, usually one year. The changes in the value are related to the eventual migrations in the credit quality of the obligors, measured by a credit rating system, including up- and downgrades, as well as default. A detailed description of the model is its technical document [74].

The approach of CreditMetrics applies primarily to bonds and loans which are both treated in the same manner. It can be easily extended to any type of financial claims for which we can derive the loss (or, more precisely, the profit / loss) distribution of the credit at the risk horizon, for all possible credit ratings including default. Receivables, loan commitments and financial letters of credit are treated in the original model [74], a simple extension of it as in Finger [51] allows also the treatment of credit derivatives like

swaps and forwards.

The structure of the model is given in Figure 1.2. There are four 'building blocks'. In the first building block, the loss distribution at the time horizon of each single credit in the portfolio is specified separately. In the second one, the multivariate dependence structure in the portfolio is modelled by means of a set of risk factors. In the third one, the standard deviation of the portfolio loss and the contribution of every individual credit to it are computed analytically. In the forth one, Monte Carlo simulation is used to determine the portfolio loss distribution.

Block 1: First one has to specify a rating system, with rating categories, and assign each of the obligors in the portfolio to an initial rating category. The exact rating system and the number of rating categories is not significant. It can be Moody's, or Standard&Poor's, or the internal rating system of a bank. However, a strong assumption made by CreditMetrics is that all obligors are credit-homogeneous within the rating category, i.e. with the same rating migration probabilities and the same default probability over the time horizon. These probabilities are collected in a stochastic matrix (migration matrix).

The risk horizon is usually one year, but in principle it can be arbitrary. However, it should be consistent with the availability of historical default and rating migration data which is used to calibrate the migration matrix. More details are given in Section 4.1.1, see also Duffie and Singleton [40], Chapter 4.

Then, one has to specify and calibrate a model for the loss at the risk horizon for each possible future rating category, including default, for each credit in the portfolio. The nature and the complexity of this model depends crucially on the type of the credit (e.g. loan, bond or derivative). Market data like credit spreads, as well as rating agencies' data like recovery rates for different obligor classes, is used for the calibration. More details are given in Section 4.1.2, see also Schönbucher [129], Section 8.5.

Specifying the loss distribution for each possible future rating category and the probability that the obligor migrates to this category is sufficient to determine the loss distribution at the risk horizon of the single credit (**the marginal loss distribution**). This completes the first building block of CreditMetrics.

Block 2: The second building block of CreditMetrics is to specify the dependence structure in the portfolio. It has been observed (see Nickell et al. [120]), that rating migrations and defaults vary significantly with the business cycle. Furthermore, obligors operating in the same region (country) and/or business sector tend to default simultaneously (over short time intervals). Thus defaults and rating migrations are dependent events, and a multivariate model for them is needed.

The suggested model is in the spirit of Merton [115]. Default (or rating migration) is triggered when the asset (log-)returns of the obligor fall below a certain barrier. The

Figure 1.2: The four building blocks of the CreditMetrics model.
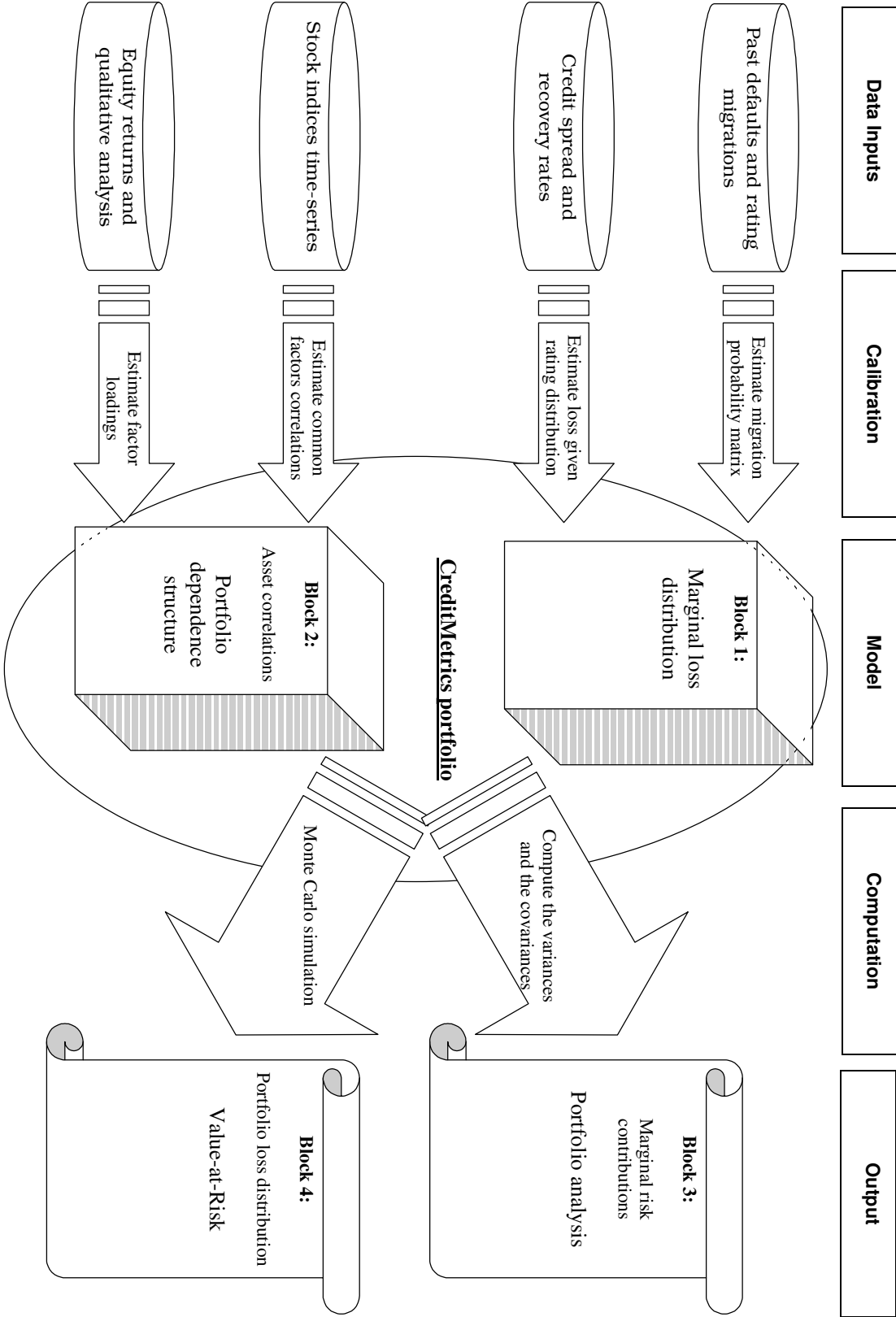
asset returns are modelled by a multivariate normal random vector. Each asset return is decomposed into an obligor-specific part and a common part with corresponding loadings by means of a linear factor model. The correlations between the common factors are estimated from observable macroeconomic stock indices. The factor loadings are estimated by a combination of qualitative and quantitative analysis of equity returns data. The resulting **asset correlations** are the final output of building block 2. More details are given in Section 3.1 and in Section 4.4, see also Bluhm et al. [16], Section 2.4.1.

Block 3: The aim of the third building block is to perform a portfolio analysis. The key issue is to compute the marginal contributions of the individual credits to the overall portfolio risk. Following the classical theory on portfolio selection (i.e. Markowitz [108]) the risk is measured by **the standard deviation**. In the CreditMetrics model, this quantity can be computed analytically. More details are given in Section 3.3 and in Section 7.2, see also Tasche [134].

Block 4: It has been well recognized in CreditMetrics [74], that the credit loss distributions are typically highly skewed and asymmetric. Therefore, knowing the standard deviation of the portfolio loss distribution (building block 3) is not sufficient to characterize the risk of extreme losses, as the standard deviation is a "two-sided" risk measure (see also Embrechts et al. [48] for a more general discussion). In order to measure the risk of extreme losses, one needs the tail of the distribution function. Since in the CreditMetrics model this cannot be computed analytically, a Monte Carlo simulation is applied. **The VaR** and other down-sided risk measures are obtained as an output. We present more details in Section 5.1, see also Glasserman and Li [70].

1.3   Generalizations of CreditMetrics

Various generalizations of the CreditMetrics model are possible.

With respect to the *rating modelling*, a strong assumption made in CreditMetrics [74] is that the obligors are credit-homogeneous within a rating category, i.e. the default and rating migration probabilities are determined by the initial rating category of the obligor. However, there is a significant amount of evidence that different obligors of the same rating have different credit qualities, i.e. different default and rating migration probabilities. In particular, the previous rating of an obligor, the time spent in the current rating, and the presence of the obligor in watch-lists published by the rating agencies are important determinants of these probabilities, see Section 4.1.1 and the references therein. For this reason, in the thesis we consider the marginal default and migration probabilities as **attributes of the individual credit** and not of the rating category of the obligor. A similar extension to the standard CreditMetrics framework is the popular in the industry

KMV model [89, 90].

A considerable amount of research has been done in the development of *rating-based evaluation models for single-obligor credit risks* like bonds and single-name credit derivatives, see e.g. Schönbucher [129], Chapter 8, for a detailed review. Some of these models are listed in Section 4.1.2. We do not go into detail here, since our focus is more on the portfolio view of credit risk. Instead we adopt a general formulation. In the thesis the profit/loss of the individual credit, given its credit rating, is modelled by an **arbitrary random variable**.

Our main focus is on *modelling the dependence between the default and rating migration* events. Following the approach in CreditMetrics [74], we model this dependence by introducing a random vector with continuous marginal distributions, which represents the asset returns of the obligors in the portfolio. We use the general class of **elliptical distributions**. On the one hand, our model retains the standard Gaussian model as a special case, and has a similar factor structure. On the other hand, by introducing a heavy-tailed "global shock", affecting the obligors' assets simultaneously across regions and business sectors, we obtain a more flexible model for joint extreme losses. Similar models are considered in Frey and McNeil [59], Daul et al. [30] and others.

1.4   Summary of main results

The structure of the thesis is as follows.

In Chapter 2 we introduce the necessary mathematical background. In Section 2.1 we define and explain the notion of copulas. We give some of the properties of important dependency measures like the linear correlation, Kendall's tau and the tail dependence coefficient. In Section 2.2 we introduce the concept of regularly varying random variables and random vectors. In Section 2.3 we focus on the class of elliptical distributions, where we recall and extend whenever necessary some classical and more recent results. Special attention is paid to the properties of the dependence measures from Section 2.1 in the framework of elliptical copulas and in particular to the results concerning the extremal dependence structure of the regularly varying elliptical random vectors.

In Chapter 3 we introduce our heavy-tailed model and compare some of its properties to the Gaussian model used for instance in CreditMetrics [74]. We show a preliminary result which allows us to split the calibration of the model into a calibration of the marginal losses and a calibration of an elliptical copula. Expressions for the moment generating function of the portfolio loss and its first two moments are derived in Section 3.3.

In Chapter 4 we provide a calibration procedure for the model. We start with the marginal parameters. Some classical or more recent methods for the calibration of the

marginal losses given rating and the marginal rating migration and default probabilities are given in Section 4.1. Then, in Section 4.2, we consider the estimation of several important dependence measures for a bivariate random vector: Kendall's tau, the linear correlation and the tail dependence coefficient. Special attention is put to the estimation of the tail dependence coefficient, where new results aiming at a lower variance of the estimates are provided. A key result in this chapter is stated in Section 4.3, where a non-parametric method for estimation of an elliptical copula with arbitrary marginals is suggested. Our approach leads to consistent estimates of the correlation coefficients based on the whole sample of observations and, at the same time, consistent estimates of the tail dependence coefficients, which are based on the joint extreme observations. In Section 4.4, the method is applied to the portfolio credit risk model; additional results regarding the estimation of the common and specific risk factors' loadings are provided. Numerical examples, investigating the accuracy and the robustness of the method, as well as real-data studies are presented in Section 4.5.

In Chapter 5 we obtain the portfolio loss distribution by means of Monte Carlo simulation. We compare the impact of the different model assumptions (Gaussian vs heavy-tailed) on the tail of the loss distribution. An importance sampling algorithm reducing the variance of the Monte Carlo estimate of the tail is provided. We pay particular attention to its application in the framework of heavy-tailed risk factors, thus extending the work of Glasserman and Li [70].

In Chapter 6 we derive an upper bound of the tail of the portfolio loss distribution which is particularly accurate at high loss levels. As it is not possible to compute this upper bound explicitly, we use a stochastic approximation algorithm and Monte Carlo simulation. Under weak regularity conditions, we prove a.s. convergence of the proposed algorithm. Numerical support for the accuracy and the computational efficiency of the method is given in Section 6.3.

In Chapter 7 we use the derived approximation to provide a semi-Monte Carlo, semi-analytic method to determine the marginal contributions of the individual credits to the overall portfolio risk. We compare the risk contributions obtained by our method to the Expected Shortfall (ES) contributions suggested for instance in Overbeck [122]. We find out that the two methods give similar results, however, our method has lower computational costs. Finally, we investigate the impact of the different model assumptions (Gaussian vs heavy-tailed) on the risk structure of the portfolio, as given by the marginal risk contributions.

We conclude in Chapter 8.

Chapter 2

Preliminaries

2.1 Copulas and dependence measures

In this section we describe the main issues in modelling dependence with copulas. For more details see Embrechts et al. [47], Joe [83], Nelsen [118]. We summarize the necessary results without proofs.

Definition 2.1.1. *For $d \in \mathbb{N} \setminus \{0\}$ a d-dimensional distribution function (d.f.) with uniformly distributed on $[0,1]$ marginals is called* copula. $\qquad\square$

The following theorem is known as Sklar's Theorem. It is perhaps the most important result regarding copulas, and is used in essentially all applications of copulas.

Theorem 2.1.2. *[Sklar [133]] Let H be a d-dimensional d.f. with marginals $F_1, \ldots, F_d$. Then there exists a copula C such that for all $y \in \mathbb{R}^d$*

$$H(y_1, \ldots, y_d) = C(F_1(y_1), \ldots, F_d(y_d)).$$

If $F_1, \ldots, F_d$ are all continuous, then C is unique; otherwise C is uniquely determined on $\mathrm{Ran}F_1 \times \ldots \times \mathrm{Ran}F_d$. Conversely, if C is a copula and $F_1, \ldots, F_d$ are d.f.s, then the function H defined as above is a d-dimensional d.f. with marginals $F_1, \ldots, F_d$. $\qquad\square$

From Sklar's theorem we see that for continuous multivariate d.f.s., the univariate marginals and the multivariate dependence structure can be separated, and that the dependence structure can be represented by a copula.

Corollary 2.1.3. *Let H be a d-dimensional d.f. with continuous marginals $F_1, \ldots, F_d$ and copula C. Then for every $u \in [0,1]^d$*

$$C(u_1, \ldots, u_d) = H(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)),$$

where $F_j^{-1}(u_j)$, $j = 1, \ldots, d$, denotes the generalized inverse function. $\qquad\square$

Let Y be a random vector with continuous marginals $F_1, \ldots, F_d$ and a joint distribution H. By means of Theorem 2.1.2 we have

$$P(Y_1 < y_1, \ldots, Y_d < y_d) = C(F_1(y_1), \ldots, F_d(y_d)) \,.$$

Since $P(Y_1 < y_1, \ldots, Y_d < y_d) = \prod_{j=1}^{d} P(Y_j < y_j)$ if and only if the components of Y are independent, we get the following proposition.

Proposition 2.1.4. *If Y is a random vector with continuous marginals, then the components of Y are independent if and only if the copula has the form*

$$C(u_1, \ldots, u_d) = \prod_{j=1}^{d} u_j \,.$$

$\square$

One nice property of copulas is that for strictly monotone transformations of the random variables (r.v.s), copulas are either invariant, or change in certain simple ways. Note that if the d.f. of the r.v. X is continuous, and if α is a strictly monotone function whose domain contains the range of X (Ran$X$), then the d.f. of $\alpha(X)$ is also continuous. For a proof of the next proposition see Embrechts et al. [47], Theorems 2.6 and 2.7.

Proposition 2.1.5. *Let Y be a random vector with continuous marginals and copula C. If $\alpha_1, \ldots, \alpha_d$ are strictly increasing functions on $\mathrm{Ran}Y_1, \ldots, \mathrm{Ran}Y_d$, respectively, then also the random vector $(\alpha_1(Y_1), \ldots, \alpha_d(Y_d))$ has copula C. If $\alpha_1, \ldots, \alpha_d$ are strictly monotone on $\mathrm{Ran}Y_1, \ldots, \mathrm{Ran}Y_d$, respectively, and, without loss of generality, α_1 is strictly decreasing, then*

$$\begin{aligned}
C_{\alpha_1(Y_1), \ldots, \alpha_d(Y_d)}(u_1, \ldots, u_d) &= C_{\alpha_2(Y_2), \ldots, \alpha_d(Y_d)}(u_2, \ldots, u_d) - \\
&\quad C_{Y_1, \alpha_2(Y_2), \ldots, \alpha_d(Y_d)}(1 - u_1, u_2, \ldots, u_d) \,.
\end{aligned}$$

$\square$

Copulas provide a natural way to study the dependence between r.v.s. For practical purposes, however, it is often needed to summarize the dependence between two r.v.s in a single number, i.e. to use a dependence measure. As a direct consequence of Proposition 2.1.5, the dependence measures which are copula properties, i.e. which are determined only by the copula regardless of the marginals, are invariant under strictly increasing transformations of the underlying r.v.s.

The linear correlation is most frequently used in practice as a measure of dependence. However, since the linear correlation is not copula-based, it can often be quite misleading

and should not be taken as the canonical dependence measure within the copula concept. Below we recall the basic properties of the linear correlation and then continue with some copula-based measures of dependence.

Definition 2.1.6. *Let X and Y be r.v.s with $0 < \text{var}(X) < \infty$ and $0 < \text{var}(Y) < \infty$. The* linear correlation *between X and Y is defined as*

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}},$$

where $\text{cov}(X,Y) = E[XY] - E[X]E[Y]$. $\qquad\square$

The linear correlation is a measure of linear dependence. In the case of a perfect linear dependence, i.e. $Y = aX + b$, $a \in \mathbb{R} \setminus \{0\}$, $b \in \mathbb{R}$, we have $|\rho_{X,Y}| = 1$. More important is that the converse also holds. Otherwise $|\rho_{X,Y}| < 1$. Furthermore, the linear correlation has the property that $\rho_{aX+b,cY+d} = \text{sign}(ac)\rho_{X,Y}$. Hence the linear correlation is invariant under strictly increasing linear transformations.

The linear correlation is easily manipulated under linear operations. Let A and B be $m \times d$ matrixes, $a, b \in \mathbb{R}^m$ and X, Y be d-dimensional random vectors. Then $\text{cov}(AX + a, BY + b) = A\text{cov}(X,Y)B'$. Therefore $\text{var}(aX) = a'\text{cov}(X,X)a$, i.e. the variance of a linear combination is fully determined by the pairwise covariances between the components, a property which is crucial in portfolio theory.

Next we will study the concept of perfect dependence. For a proof of the next proposition see Embrechts et al. [48].

Proposition 2.1.7. *Let (X,Y) be a bivariate random vector. Then its copula $C_{X,Y}(u,v) = \min(u,v)$ if and only if there exist increasing functions $\alpha, \beta : \mathbb{R} \to \mathbb{R}$ and a r.v. Z, such that $(X,Y) \stackrel{d}{=} (\alpha(Z), \beta(Z))$. In this case X and Y are called* comonotonic. $\qquad\square$

Corollary 2.1.8. *Let (X,Y) be a random vector with continuous marginals F and G resp. and copula $C_{X,Y}(u,v) = \min(u,v)$. Then $Y = G^{-1}(F(X))$ a.s.* $\qquad\square$

Next we provide an alternative to the linear correlation dependence measure.

Definition 2.1.9. *Let (X,Y) and $(\widetilde{X}, \widetilde{Y})$ be bivariate random vectors with continuous and common marginals F (of X and $\widetilde{X}$) and G (of Y and $\widetilde{Y}$). Then (X,Y) and $(\widetilde{X}, \widetilde{Y})$ are called* concordant (disordant) *if $(X - \widetilde{X})(Y - \widetilde{Y}) > 0 (< 0)$.* $\qquad\square$

The following theorem can be found in Nelsen [118], p.127.

Theorem 2.1.10. *Let (X,Y) and $(\widetilde{X}, \widetilde{Y})$ be independent random vectors with continuous and common marginals F (of X and $\widetilde{X}$) and G (of Y and $\widetilde{Y}$). Let C and $\widetilde{C}$ denote the*

copulas of (X, Y) and $(\widetilde{X}, \widetilde{Y})$ resp. Denote with τ the probability of concordance minus the probability of disordance, i.e.

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

Then

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

$\square$

Definition 2.1.11. Kendall's tau *for the bivariate random vector (X, Y) with continuous marginals is defined as*

$$\tau = P((\widehat{X} - \widetilde{X})(\widehat{Y} - \widetilde{Y}) > 0) - P((\widehat{X} - \widetilde{X})(\widehat{Y} - \widetilde{Y}) < 0)\,, \tag{2.1}$$

where $(\widehat{X}, \widehat{Y})$ and $(\widetilde{X}, \widetilde{Y})$ are independent copies of (X, Y). $\square$

By means of Theorem 2.1.10, the Kendall's tau is a copula property. As a consequence, the Kendalls' tau is invariant under any increasing componentwise transformations (unlike the linear correlation). For some examples and additional advantages of this dependence measure see Embrechts et al. [48].

The concept of tail dependence relates to the amount of dependence in the upper-right-quadrant or the lower-left-quadrant tail of a bivariate distribution. It is a concept that is relevant for the study of the dependence between extreme events.

Definition 2.1.12. *Let (X, Y) be a random vector with continuous marginal d.f.s F and G resp. and copula C. The coefficient of* lower tail dependence *of (X, Y) is*

$$\lambda_L = \lim_{u \to 0} P(X < F^{-1}(u) \mid Y < G^{-1}(u)) = \lim_{u \to 0} \frac{C(u, u)}{u}\,, \tag{2.2}$$

if the limit exists. If $\lambda_L > 0$, (X, Y) are called lower tail $-$ dependent. *Otherwise if $\lambda_L = 0$, (X, Y) are called* lower tail $-$ independent. *In a similar way one may define the coefficient of* upper tail dependence *as*

$$\lambda_U = \lim_{u \to 1} P(X > F^{-1}(u) \mid Y > G^{-1}(u)) = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u}\,. \tag{2.3}$$

$\square$

From (2.2) and (2.3) we find that the tail dependence coefficient is a copula property, hence the amount of tail dependence is invariant under strictly increasing transformations on the marginal r.v.s. For copulas without closed analytical form another expression for the tail dependence coefficient is more useful. Consider a pair of uniformly distributed on

[0, 1] r.v.s (U, V) with copula C. Note that $P(V < v \,|\, U = u) = \frac{\partial C(u,v)}{\partial u}$ and $P(V > v \,|\, U = u) = 1 - \frac{\partial C(u,v)}{\partial u}$, and similarly when conditioning on V. Then

$$
\begin{aligned}
\lambda_L &= \lim_{u \to 0} \frac{C(u, u)}{u} \\
&= \lim_{u \to 0} \left(\frac{\partial}{\partial s} C(s, t)_{|\, s=t=u} + \frac{\partial}{\partial t} C(s, t)_{|\, s=t=u} \right) \\
&= \lim_{u \to 0} \left(P(V < u \,|\, U = u) + P(U < u \,|\, V = u) \right) .
\end{aligned}
\tag{2.4}
$$

Next we define a more general measure for extremal dependence.

Definition 2.1.13. *Let (X, Y) be a bivariate random vector with continuous marginals F and G resp. and copula C. For $x, y > 0$ we call*

$$
\lambda(x, y) = \lim_{u \to 0} \frac{1}{u} P(F(X) < ux, G(Y) < uy) = \lim_{u \to 0} \frac{C(ux, uy)}{u} ,
$$

(provided the limit extists) lower tail copula. *Similarly one may define an* upper tail copula. □

Note that $\lambda(1, 1)$ in Definition 2.1.13 is exactly equal to the lower tail dependence coefficient. Furthermore, the tail copula is uniquely determined by the copula of the random vector, and hence is invariant under marginal transformations.

In Section 2.3 we come back to the dependence measures correlation, Kendall's tau and tail dependence, as well as to the tail copula, in the framework of elliptical distributions. We conclude this section with a definition of a dependence measure which is particularly suitable for working with discrete binary r.v.s.

Definition 2.1.14. *Let X_1, X_2 be binary $\{0, 1\}$ r.v.s with $P(X_i = 1) = p_i > 0$, $i = 1, 2$. The* odds ratio *is defined as*

$$
\mathrm{OR}(X_1, X_2) = \frac{P(X_1 = 1, X_2 = 1) P(X_1 = 0, X_2 = 0)}{P(X_1 = 1, X_2 = 0) P(X_1 = 0, X_2 = 1)} .
\tag{2.5}
$$

□

For an interpretation of this measure, consider a model where $X_i = 1$, $i = 1, 2$, is the undesirable outcome, e.g. in a simple credit risk model it means the default of obligor i. Note first that that $\mathrm{OR}(X_1, X_2) = 1$ if and only X_1 and X_2 are independent Bernoulli r.v.s. Further, (2.5) can be rewritten as

$$
\mathrm{OR}(X_1, X_2) = \frac{\mathrm{O}(X_1 \,|\, X_2 = 1)}{\mathrm{O}(X_1 \,|\, X_2 = 0)} ,
\tag{2.6}
$$

where the odds

$$
\mathrm{O}(X_1 \,|\, X_2 = k) = \frac{P(X_1 = 1 \,|\, X_2 = k)}{P(X_1 = 0 \,|\, X_2 = k)} ,
$$

for $k = 0, 1$, is the chance of default to non-default of credit 1, given that credit 2 has a state k. If for example $\mathrm{OR}(X_1, X_2) = 10$, this means that the chance of default to non-default of credit 1, given that credit 2 has defaulted, is 10 times larger than the chance of default to non-default of credit 1, given that credit 2 has not defaulted.

If the conditional probability of default for credit 1 is small, for given $X_2 = k$, $k = 0, 1$, then the odds $\mathrm{O}(X_1 \mid X_2 = k)$ approximates $P(X_1 = 1 \mid X_2 = k)$, $k = 0, 1$. In this case (2.6) can be considered as an approximation of the relative default risk of credit 1 when obligor 2 has defaulted compared to when obligor 2 has not defaulted. Note that (2.5) is symmetric with respect to the credits, so we also have the representation

$$\mathrm{OR}(X_1, X_2) = \frac{\mathrm{O}(X_2 \mid X_1 = 1)}{\mathrm{O}(X_2 \mid X_1 = 0)},$$

for

$$\mathrm{O}(X_2 \mid X_1 = k) = \frac{P(X_2 = 1 \mid X_1 = k)}{P(X_2 = 0 \mid X_1 = k)}, \quad k = 0, 1.$$

The odds ratio is a quite popular concept in the modelling and the estimation of multivariate binary events. For a detailed description of the various modelling techniques see e.g. Collett [23], Czado [26]. We come back to this issue in the framework of our credit risk model in Section 3.2.

2.2   Regular variation

The concept of regular variation plays a crucial role in the study of extreme events, both in the one-dimensional case and in the multivariate extremes of a random vector. We start with some classical results from the one-dimensional extreme value theory, see Embrechts et al. [46] for a detailed exposition. In Section 2.2.2 we continue with the multivariate regime.

2.2.1   One-dimensional regular variation

We first recall the definition of regular variation for one-dimensional r.v.s.

Definition 2.2.1. *A Lebesque-measurable function F is called* regularly varying *with index $\alpha > 0$ if for all $x > 0$*

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

When $\alpha = 0$, the function F is said to be a slowly varying *function. The (non-degenerate) r.v. X is said to be* regularly varying *with* tail index $\alpha > 0$ *if its tail is a regularly varying function with index $\alpha > 0$.*   □

In the one-dimensional case, the extremal behaviour of a sequence of r.v.s can be illustrated by the behaviour of their maxima. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of r.v.s and denote $M_n = \max(X_1, \ldots, X_n)$, for $n \in \mathbb{N} \setminus \{0\}$. The following result is the basis of classical extreme value theory.

Theorem 2.2.2. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. r.v.s with non-degenerate d.f. If there exist norming constants $c_n > 0$, $d_n \in \mathbb{R}$, $n \in \mathbb{N} \setminus \{0\}$, and some non-degenerate r.v. M such that*

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} M \,, n \to \infty \,, \tag{2.7}$$

then the d.f. of M belongs to the type of one of the following three d.f.s.

Frechet: $\Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \,, \\ \exp(-x^{-\alpha}) & x > 0 \,, \quad \alpha > 0 \,; \end{cases}$

Weibull: $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & x \leq 0 \,, \quad \alpha > 0 \,, \\ 1 & x > 0 \,; \end{cases}$

Gumbel: $\Lambda(x) = \exp(-e^{-x}) \ x \in \mathbb{R}.$ □

Details of the proof are for instance to be found in Resnik [124], Proposition 0.3. The three types of d.f.s in Theorem 2.2.2 are called *extreme value distributions.*

Definition 2.2.3. *The d.f. of the r.v. X is said to belong to the* maximum domain of attractition *of the extreme value distribution H, if there exist norming constants $c_n > 0$, $d_n \in \mathbb{R}$, $n \in \mathbb{N}$, such that (2.7) holds and M has d.f. H.* □

The concept of regular variation is crucial when one has to determine the domain of attraction of F.

Of particular interest in financial applications are the distributions in the domain of attraction of the Frechet distribution, see for instance Embrechts et al. [46], Chapter 6. The next proposition charaterizes the distributions in this domain.

Proposition 2.2.4. *The d.f. F belongs to the maximum domain of attraction of the Frechet distribution Φ_α, $\alpha > 0$, if and only if the tail $\overline{F}(x) = x^{-\alpha}L(x)$, $x > 0$, where L is a slowly varying function.* □

For a proof see Embrechts et al. [46], Theorem 3.3.7. If the r.v. X has d.f. F in the domain of attraction of Φ_α, then X is regularly varying and its tail decreases quite slowly (at a, roughly said, polynomial rate). Note that this implies, for instance, that $E[X^\beta] = \infty$ for every $\beta \geq \alpha$. Thus, X is a 'heavy-tailed' r.v.

The following d.f. is a convinient (from a statistical point of view) representation of the three types of extreme value d.f.s.

Definition 2.2.5. *The* standardized generalized extreme value distribution (GEV) *is defined as*

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right) & \xi \neq 0 \\ \exp\left(-\exp(-x)\right) & \xi = 0 \,, \end{cases}$$

where the support is

$$\begin{array}{ll} x > -\frac{1}{\xi} & \xi > 0 \\ x < -\frac{1}{\xi} & \xi < 0 \\ x \in \mathbb{R} & \xi = 0 \,. \end{array}$$

One can introduce the related location-scale family $H_{\xi;\mu,\sigma}$ by replacing the argument x above with $\frac{x-\mu}{\sigma}$ for some $\mu \in \mathbb{R}$ and some $\sigma > 0$. The support has to be adjusted accordingly. We will refer to $H_{\xi;\mu,\sigma}$ as GEV.   □

In the above definition, $\xi = 1/\alpha > 0$ corresponds to the Frechet case in Theorem 2.2.2, $\xi = -1/\alpha < 0$ to the Weibull case and $\xi = 0$ to the Gumbel case. Thus, the GEV distribution describes the limit distribution of normalized maxima of an i.i.d sequence.

An additional topic in extreme value theory is the distribution of scaled excesses over high thresholds. The following definition makes this notion precise.

Definition 2.2.6. *Let X be a r.v. with d.f. F and a right endpoint x_F ($x_F = \infty$ is allowed). For a fixed $u < x_F$*

$$F_u(x) = P(X - u \leq x \,|\, X > u), \;\; x > 0$$

is called the excess *distribution of X over the threshold u.*   □

It turns out that the limit behaviour of $F_u(x)$ for u tending to x_F can be described by the following d.f.

Definition 2.2.7. *The* standardized generalized Pareto distribution (GPD) *is defined as*

$$G_\xi(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-x) & \xi = 0 \,, \end{cases}$$

where the support is

$$\begin{array}{ll} x \geq 0 & \xi \geq 0 \\ 0 \leq x \leq -\frac{1}{\xi} & \xi < 0 \,. \end{array}$$

One can introduce the related location-scale family $G_{\xi;\mu,\sigma}$ by replacing the argument x above with $\frac{x-\mu}{\sigma}$ for some $\mu \in \mathbb{R}$ and some $\sigma > 0$. The support has to be adjusted accordingly. We will refer to $G_{\xi;\mu,\sigma}$ as GPD.   □

Theorem 2.2.8. *[Embrechts et al. [46], Theorem 3.4.13] Let X be a r.v. with distribution F. Then F is in the domain of attraction of the extreme value distribution H_ξ for some $\xi \in \mathbb{R}$ if and only if*

$$\lim_{u \to x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi;0,\sigma(u)}(x)| = 0$$

for some positive function $\sigma(u)$. $\square$

The extreme value theory is a tool for the analysis and extrapolation of extreme events. An immense amount of research has been done in the development of the corresponding statistical methods for extreme events (see Embrechts et al. [46], Chapter 6 for a detailed overview). One popular suggestion is the Hill estimator for the parameter ξ of the GEV distribution.

Definition 2.2.9. *Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with d.f. in the domain of attraction of the Frechet distribution H_ξ, $\xi > 0$. Let $X_{1,n} \geq X_{2,n} \geq \ldots \geq X_{n,n}$ be the ordered sample. The* Hill *estimator of ξ takes the form*

$$\widehat{\xi}_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n} \tag{2.8}$$

where $k = k(n)$ is some positive function of the number of observations n, such that $k(n) \to \infty$ and $\frac{k(n)}{n} \to 0$, $n \to \infty$. $\square$

Theorem 6.4.6. in Embrechts et al. [46] provides conditions for consistency and asymptotic normality of $\widehat{\xi}$. Note that, to apply the estimator we need to select k from the upper order statistics. Thus, if we select a small k, the estimator has a large variance; otherwise, if we select a large k, a bias may enter. Methods to find a bias-variance trade-off are given in e.g. de Haan and Peng [32].

Another popular statistical method in extreme value theory, known as the Peaks-over-Threshold method (POT), is based on Theorem 2.2.8. Note that the density of the GPD with parameters ξ and σ is explicitly available:

$$g_{\xi;\sigma}(x) = \frac{1}{\sigma} \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi - 1} , \quad x \in \text{supp}(G_{\xi;\sigma}).$$

Therefore, given a sample $Y_1, \ldots, Y_n$ of i.i.d. r.v.s with GPD, the log-likelihood function equals

$$l(\xi, \sigma; Y) = -n \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{n} \ln \left(1 + \xi \frac{Y_i}{\sigma}\right) . \tag{2.9}$$

Let now $X_1, \ldots, X_n$ be a sample of i.i.d. r.v.s with d.f. $F(x) = P(X \leq x)$ in the domain of attraction of the Frechet distribution Φ_α. We select a high threshold u and denote

$N_u = \#\{i : X_i > u\}$ and the corresponding excesses $Y_1, \ldots, Y_{N_u}$. For every $x > u$ we have

$$P(X > x) = P(X > u)P(X > x \,|\, X > u).$$

The Glivenko-Cantelli theorem suggests to approximate $P(X > u)$ with the empirical d.f., i.e. with $\frac{N_u}{n}$, for a range where there are still enough data to guarantee good approximation. Theorem 2.2.8 suggests to approximate $P(X > x \,|\, X > u)$ with the GPD $G_{\xi;\sigma}(x-u)$ with the appropriate parameters ξ and σ, introducing a semiparametric model, where data are sparse. Maximum likelihood (ML) estimates for ξ and σ can be found by using the excesses $Y_1, \ldots, Y_{N_u}$ from the sample and maximizing numerically (2.9).

For the asymptotic theory behind this method see Smith [132]. As with the Hill estimator, a key point is the selection of the threshold u, where a bias-variance trade-off is sought, see the discussion on p.355 in Embrechts et al. [46].

2.2.2   Multivariate regular variation

In this section we give some basic results on multivariate regular variation, for details see Kallenberg [86]. To prepare for the definition of regular variation for random vectors, we recall the concept of vague convergence. Let χ be a separable metric space. A set $B \subset \chi$ is said to be relatively compact if its closure $\overline{B}$ is compact. Let $\sigma(\chi)$ be the Borel σ-algebra on χ. A measure μ on $(\chi, \sigma(\chi))$ is called a Radon measure if $\mu(B) < \infty$ for all relatively compact sets $B \in \sigma(\chi)$.

Definition 2.2.10. *Let $\mu, \mu_1, \mu_2, \ldots$ be Radon measures on $(\chi, \sigma(\chi))$. We say that μ_n converges to μ* vaguely*, if*

$$\lim_{n \to \infty} \int_\chi f(s)\mu_n(ds) = \int_\chi f(s)\mu(ds)$$

for all continuous functions $f : \chi \to \mathbb{R}^+$ with compact support. We denote $\mu_n \overset{v}{\to} \mu$.   □

A useful equivalent formulation of vague convergence is given in the following theorem.

Theorem 2.2.11. *Let $\mu, \mu_1, \mu_2, \ldots$ be Radon measures on $(\chi, \sigma(\chi))$. Then μ_n converges to μ vaguely, if and only if*

$$\lim_{n \to \infty} \mu_n(B) = \mu(B)$$

for all relatively compact $B \in \sigma(\chi)$ with $\mu(\partial B) = 0$.   □

For a proof see Kallenberg [86], p.169.

Definition 2.2.12. *The d-dimensional random vector Y is said to be* regularly varying *with* tail index $\alpha > 0$ *if there exists a random vector Θ with values in $\mathbb{S}^{d-1} = \{y \in \mathbb{R}^d : |y| = 1\}$ such that for all $y > 0$*

$$\frac{P(|Y| > ty, \frac{Y}{|Y|} \in \cdot)}{P(|Y| > t)} \xrightarrow{v} y^{-\alpha} P(\Theta \in \cdot), \quad t \to \infty. \tag{2.10}$$

The distribution of Θ is referred to as the spectral measure *of Y.*   □

Remark 2.2.13. In Definition 2.2.12 we do not specify the choice of a norm $|\cdot|$. The reason for this is that whether a random vector is regularly varying or not does not depend on the choice of norm, see Hult and Lindskog [78]. In particular, the tail index α in (2.10) remains the same irrespective of the selected norm.   □

Although the property of multivariate regular variation is norm-independent, the form of the spectral measure is certainly not. Frequently encountered norms include the Euclidean L^2-norm

$$|Y| = \sqrt{\sum_{i=1}^{d} Y_i^2}$$

and the max-norm

$$|Y|_{max} = \max(|Y_1|, \ldots, |Y_d|).$$

In what follows $|\cdot|$ denotes an arbitrary norm, unless stated otherwise. The choice of a norm for a given situation is a problem in its own right. Clearly, the max-norm is large as soon as at least one of the components of Y is large. Whether the other components are large or not is not of interest. In the Euclidean norm, on the other hand, we need not have any component of Y extremely large in order for $|Y|$ to be large, since the components of Y jointly determine $|Y|$. The max-norm might therefore be useful in models where always the largest component outcome as such is of interest, while the Euclidean norm pays attention to all components of Y. Before starting to model multivariate extremes it might be worthwhile giving some thought to the choice of norm.

A useful fact contained in the next proposition concerns the effect of additive constants in the multivariate regular variation settings.

Proposition 2.2.14. *Let Y be a d-dimensional regularly varying random vector with tail index $\alpha > 0$ and spectral measure P_Θ w.r.t. some norm $|\cdot|$ and let $b \in \mathbb{R}^d$ be a constant vector. Then $Y + b$ is regularly varying with the same tail index α and the same spectral measure w.r.t. the norm $|\cdot|$.*   □

For a proof see Hult and Lindskog [78].

The next proposition presents a link between regularly varying vectors and their norms.

Proposition 2.2.15. *Let Y be a d-dimensional regularly varying random vector with tail index α. Then $|Y|$ is a regularly varying r.v. in the sense of Definition 2.2.1 with tail index α.*   □

For a proof see Wendin [135], Proposition 1.3. Theorem 1.1 in the same work states the result to follow, see also Theorem 2.2.11 and Basrak [10].

Proposition 2.2.16. *Let Y be a d-dimensional random vector with support on the whole space of $\mathbb{R}^d$. Then Y is regularly varying with tail index α if and only if there exists a sequence of norming constants $a_n > 0$ and a non-zero Radon measure μ such that*

$$nP(a_n^{-1}Y > \cdot) \xrightarrow{v} \mu(\cdot). \tag{2.11}$$

Moreover, for every $t > 0$ and for every relatively compact set $B \in \sigma(\chi)$ with $\mu(\partial B) = 0$, $\mu(tB) = t^{-\alpha}\mu(B)$ □

Remark 2.2.17. The Radon measure in (2.11) is unique up to a multiplicative constant. □

Proposition 2.2.16 implies also the existence of constants $a_n > 0$, $n \in \mathbb{N}$, such that for the marginals of the random vector Y holds

$$nP(a_n^{-1}Y_i > x) \to c_i x^{-\alpha}, \quad n \to \infty,$$

where $c_i \geq 0$, $i = 1, \ldots, d$. Unfortunatelly nothing guarantees that the constants c_i are strictly positive. Non-positive constants may occur, for instance, when the marginals of Y are regularly varying with different tail indices, see p.11 in Mikosch [116].

Our next aim is to present a consistent estimator of the Radon measure μ in Proposition 2.2.16.

Let $|\cdot|$ be an arbitrary norm and let Y be a regularly varying random vector with tail index α. Denote by F the d.f. of $|Y|$ and by F^{-1} its generalized inverse function. A possible choice of norming constants in (2.11) is $a_n = F^{-1}(1/n)$, see Proposition 2.2.15 and Resnick [124], Proposition 1.11. Then we introduce the auxiliary sequence $k = k(n)$ with the properties $k(n) \to \infty$ and $\frac{k(n)}{n} \to 0$ as $n \to \infty$. Because of (2.11),

$$\frac{n}{k}P(\frac{1}{F^{-1}(k/n)}Y \in \cdot) \xrightarrow{v} \mu(\cdot), \quad n \to \infty.$$

Consequently, with $Y^{(1)} \ldots, Y^{(n)}$ being i.i.d copies of Y, we obtain

$$\widehat{\mu}_n(\cdot) = \frac{1}{k}\sum_{i=1}^n 1_{\{Y^{(i)}/F_E^{-1}(k/n)\in\cdot\}} \xrightarrow{v} \mu(\cdot), \quad n \to \infty, \tag{2.12}$$

where $F_E^{-1}(x)$ denotes the inverse of the empirical d.f. of $|Y|$. For details of the proof see e.g. Wendin [135].

2.3   Modelling dependence by elliptical distributions

The main topic of this section is to understand the various measures of dependence through elliptical distributions. First we introduce the class of elliptically distributed random vectors and give some of their properties. For further details about elliptical distributions we refer to Fang et al. [50].

Definition 2.3.1. *If Y is a d-dimensional random vector and, for some vector $\mu \in \mathbb{R}^d$, some $d \times d$ non-negative definite symmetric matrix Σ and some function $\phi : [0, \infty) \to \mathbb{R}$, the characteristic function of $Y - \mu$ is of the form $\varphi_{Y-\mu}(t) = \phi(t'\Sigma t)$, we say that Y has an* elliptical *distribution with parameters μ, Σ and ϕ. The function ϕ is referred to as the* characteristic generator *of Y.* $\square$

When $d = 1$, the class of elliptical distributions coincides with the class of one-dimensional symmetric distributions.

For elliptically distributed random vectors, we have the following general representation theorem see Fang et al. [50], Theorem 1.3.

Theorem 2.3.2. *The d-dimensional random vector Y is elliptically distributed with parameters μ, Σ (rank(Σ) = k) and ϕ, if and only if there exist a non-negative r.v. R, independent of U, a k-dimensional random vector uniformly distributed on the unit hypersphere $S_k = \left\{ z \in \mathbb{R}^k : z'z = 1 \right\}$, and a $d \times k$ matrix A with $AA' = \Sigma$, such that*

$$Y = \mu + RAU.$$ (2.13)

$\square$

Remark 2.3.3. Uniqueness of the representation:

(1) Representation (2.13) is not unique: if B is an orthogonal $k \times k$ matrix, then (2.13) holds also with $A_{new} = AB$ and $U_{new} = B'U$.

(2) Elliptical distributions with different parameters can in fact be equal: for every $c > 0$, $Y \overset{d}{=} Y_{new}$, where Y_{new} has parameters μ, $c\Sigma$ and $\phi(\frac{\cdot}{c})$, or, alternatively, $R_{new} = R/\sqrt{c}$ in (2.13) $\square$

In this work we are going to use frequently the class of normal variance mixture distributions (NVM).

Definition 2.3.4. *Let Z be a d-dimensional normal random vector with zero mean and covariance matrix Σ with rank(Σ) = d ($Z \in N_d(0, \Sigma)$). Let $W > 0$ be a r.v., independent of Z and $\mu \in \mathbb{R}^d$ be a constant vector. Then we say that $Y = \mu + WZ$ has a* normal variance mixture *distribution ($Y \in N_d(\mu, W^2\Sigma)$).* $\square$

Remark 2.3.5. Note that the NVM distributions are a subclass of the elliptical distributions. To see this we take an arbitrary NVM $Y = \mu + WZ$ and we let $R = W\sqrt{\chi_d^2}$, where χ_d^2 is a chi-square distributed r.v., independent of W. We obtain for the NVM distribution $Y = \mu + WZ \overset{d}{=} \mu + RAU$, where A is a matrix such that $AA' = \Sigma$ and U is a d-dimensional random vector uniformly distributed on the unit hypersphere $S_d = \{z \in \mathbb{R}^d : z'z = 1\}$ – i.e. Y is elliptical by means of (2.13). In particular, Y is t-distributed with ν degrees of freedom if $R^2/d \in F(d, \nu)$ (F-distribution with d and ν degrees of freedom), or, equivalently, $W = \sqrt{\frac{\nu}{\chi_\nu^2}}$. Most of the known elliptical distributions can be represented as NVM, see Fang et al. [50], Theorem 2.21 for details.   □

Remark 2.3.6. If the NVM Y has finite second moments, then we can always find a representation such that $\text{cov}(Y) = \Sigma$. To see this we observe that $\text{cov}(Y) = \text{cov}(\mu + WZ) = E[W^2]\text{cov}(AZ_I)$, where A is a matrix such that $AA' = \Sigma$ (Choleski decomposition) and Z_I is a vector of independent standard normals. Then $\text{cov}(Y) = E[W^2]\Sigma$. So, as in Remark 2.3.3 (2) we set $c = E[W^2]$.   □

The following result provides the basis of most applications of elliptical distributions.

Lemma 2.3.7. *Let Y be d-dimensional elliptically distributed random vector with parameters μ, Σ and ϕ and let B be a $q \times d$ matrix and $b \in \mathbb{R}^q$. Then $b + BY$ is a q-dimensional elliptically distributed random vector with parameters $\mu_{new} = b + B\mu$, $\Sigma_{new} = B\Sigma B'$ and ϕ.*

Proof. By Theorem 2.3.2, $b + BY \overset{d}{=} b + B\mu + RBAU$.   □

If we partition Y, μ and Σ into $Y = (Y_1, Y_2)'$, $\mu = (\mu_1, \mu_2)'$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where Y_1 and μ_1 are r-dimensional vectors ($r < d$) and Σ_{11} is a $r \times r$ matrix, we have the following consequence of Lemma 2.3.7.

Corollary 2.3.8. *If Y is a d-dimensional elliptically distributed random vector with parameters μ, Σ and ϕ, then Y_1 and Y_2 are resp. r and $d - r$ dimensional elliptically distributed vectors with parameters resp. μ_1, Σ_{11}, ϕ and μ_2, Σ_{22}, ϕ.*   □

Hence, marginal distributions of the elliptical distributions are elliptical and of the same type (with the same characteristic generator).

We start to investigate the dependence structure of the elliptical distributions.

Definition 2.3.9. *Let Y be a d-dimensional elliptically distributed random vector with parameters μ, Σ and ϕ. We call*

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}, \ i, j = 1, \ldots, d, \tag{2.14}$$

correlation coefficient.   □

We observe that if $\text{var}(Y_i) < \infty$ and $\text{var}(Y_j) < \infty$, then ρ_{ij} in the above definition is the usual linear correlation, i.e. $\rho_{ij} = \frac{\text{cov}(Y_i, Y_j)}{\sqrt{\text{var}(Y_i)\text{var}(Y_j)}}$. However, for elliptical distributions ρ_{ij} is well defined even when linear correlation is not.

Corollary 2.3.10. *Let Y be d-dimensional elliptically distributed random vector with parameters μ, Σ and ϕ. If the marginals of Y are continuous, then the copula of Y is uniquely determined by ϕ and $[\rho_{ij}]_{i,j=1,\dots,d}$.*

Proof. We observe that due to Proposition 2.1.5 Y has the same copula as the vector Y_{new} obtained from Y by substracting the means μ_i and multiplying by $(\Sigma_{ii})^{-1/2}$ component-wise for $i = 1, \dots, d$. By Lemma 2.3.7 we have that Y_{new} has elliptical distribution with parameters $\mu_{new} = 0$, $\Sigma_{new} = [\rho_{ij}]_{i,j=1,\dots,d}$ and $\phi_{new} = \phi$. $\qquad\square$

The relation between the Kendall's tau (2.1) and the linear correlation coefficient is well known for bivariate normally distributed random vectors. In the next theorem we state a more general relation between the Kendall's tau and the correlation ρ_{ij} for elliptically distributed random vectors.

Theorem 2.3.11. *[Lindskog et al. [104]] Let Y be a d-dimensional elliptically distributed random vector with absolutely continuous marginals with support on the whole of $\mathbb{R}$. Then the following relation holds:*

$$\tau(Y_i, Y_j) = \frac{2}{\pi} \arcsin \rho_{ij}, \quad i, j = 1, \dots, d. \qquad (2.15)$$

$\qquad\square$

The next theorem clarifies the relation between regular variation and tail dependence for an elliptically distributed random vector. Recall Definitions 2.1.12, 2.2.1 and 2.2.12.

Proposition 2.3.12. *Let Y be a d-dimensional elliptically distributed random vector with absolutely continuous marginals with support on the whole of $\mathbb{R}$ and representation $Y = \mu + RAU$ as in (2.13). Then the following statements are equivalent:*

(1) R is regularly varying with tail index $\alpha > 0$;

(2) Y is regularly varying with tail index $\alpha > 0$;

(3) for all $i, j = 1, \dots, d$, $i \neq j$, the vector (Y_i, Y_j) is tail dependent. Moreover, the coefficients of tail dependence are given by

$$\lambda_U(Y_i, Y_j) = \lambda_L(Y_i, Y_j) = \frac{\int_{g(\rho_{ij})}^{\frac{\pi}{2}} \cos^\alpha t \, dt}{\int_0^{\frac{\pi}{2}} \cos^\alpha t \, dt}, \qquad (2.16)$$

where $g(\rho_{ij}) = \frac{\pi}{4} - \frac{\arcsin \rho_{ij}}{2} = \frac{\pi}{4}(1 - \tau_{ij})$ and ρ_{ij} and τ_{ij} are the correlation coefficient in (2.14) and Kendall's tau in (2.1), respectively.

Proof. It has been shown by Hult and Lindskog [78], that (1) and (2) are equivalent, as well as that from (1) follows (3). The same paper states also the result, that if for all $i, j = 1, \ldots, d$, $i \neq j$, the vector (Y_i, Y_j) is tail dependent, then R is regularly varying with some index $\beta > 0$ (possibly $\beta \neq \alpha$), such that

$$\frac{\int_{g(\rho_{ij})}^{\frac{\pi}{2}} \cos^\alpha t\, dt}{\int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt} = \frac{\int_{g(\rho_{ij})}^{\frac{\pi}{2}} \cos^\beta t\, dt}{\int_0^{\frac{\pi}{2}} \cos^\beta t\, dt}.$$

However, if $\beta \neq \alpha$, then this contradicts to Lemma 2.3.13 (3). $\square$

In the next lemma we analyze further the function on the right-hand side of (2.16).

Lemma 2.3.13. *Setting* $x = \frac{\pi}{4}(1 - \tau)$ *in (2.16), we define*

$$\lambda(\alpha, x) = \frac{\int_x^{\frac{\pi}{2}} \cos^\alpha t\, dt}{\int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt}, \ \alpha \geq 0, \ x \in [0, \frac{\pi}{2}].$$

It satisfies the following properties.

(1) $\lambda(\alpha, x)$ *is continuous and differentiable in* $\alpha > 0$, $x \in (0, \frac{\pi}{2})$.

(2) $0 < \lambda(\alpha, x) < 1$.

(3) Let $x \in (0, \frac{\pi}{2})$ *be fixed. Then* $\lambda(\alpha, x)$ *is strictly decreasing in* $\alpha > 0$. *Furthermore,*

$$\lim_{\alpha \to 0} \lambda(\alpha, x) = 1 - \frac{2x}{\pi}, \quad \lim_{\alpha \to \infty} \lambda(\alpha, x) = 0.$$

(4) Let $\alpha > 0$ *be fixed. Then* $\lambda(\alpha, x)$ *is strictly decreasing in* $x \in (0, \frac{\pi}{2})$.

Proof. The function $\cos^\alpha t$ for $t \in (0, \frac{\pi}{2})$ is continuous and differentiable, and so are $\int_x^{\frac{\pi}{2}} \cos^\alpha t\, dt$ and $\int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt$ as functions of $\alpha > 0$ and of $x \in (0, \frac{\pi}{2})$. Furthermore, $\cos^\alpha t > 0$ for $t \in (0, \frac{\pi}{2})$, hence $0 < \int_x^{\frac{\pi}{2}} \cos^\alpha t\, dt < \int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt$, therefore we obtain (2).

To prove (3) we differentiate with respect to α

$$\frac{\partial}{\partial \alpha} \lambda(\alpha, x) = \frac{\int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt \int_x^{\frac{\pi}{2}} \log(\cos t) \cos^\alpha t\, dt - \int_x^{\frac{\pi}{2}} \cos^\alpha t\, dt \int_0^{\frac{\pi}{2}} \log(\cos t) \cos^\alpha t\, dt}{\left(\int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt\right)^2}$$

$$= \frac{D(\alpha, x)}{\left(\int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt\right)^2}.$$

We will prove that $D(\alpha, x) < 0$ for every $x \in (0, \frac{\pi}{2})$ and $\alpha > 0$.

First we note that $D(\alpha, 0) = D(\alpha, \frac{\pi}{2}) = 0$. Then we differentiate with respect to x:

$$\frac{\partial}{\partial x} D(\alpha, x) = -\log(\cos x) \cos^\alpha x \int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt + \cos^\alpha x \int_0^{\frac{\pi}{2}} \log(\cos t) \cos^\alpha t\, dt$$

$$= \cos^\alpha x \left(-\log(\cos x) \int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt + \int_0^{\frac{\pi}{2}} \log(\cos t) \cos^\alpha t\, dt\right)$$

$$= C(\alpha, x) \cos^\alpha x.$$

Note that

$$C(\alpha, 0) = \int_0^{\frac{\pi}{2}} \log\left(\cos t\right) \cos^\alpha t \, dt < 0$$

and that

$$\lim_{x \to \frac{\pi}{2}} C(\alpha, x) = \infty$$

and that $C(\alpha, x)$ is strictly increasing in x for $x \in (0, \frac{\pi}{2})$, as $-\log\left(\cos x\right)$ is strictly increasing. Therefore there exists a unique point y, $0 < y < \frac{\pi}{2}$, such that $C(\alpha, y) = 0$. Furthermore, $\frac{\partial}{\partial x} D(\alpha, x) = C(\alpha, x) \cos^\alpha x < 0$ for $x \in (0, y)$ and $\frac{\partial}{\partial x} D(\alpha, x) = C(\alpha, x) \cos^\alpha x > 0$ for $x \in (y, \frac{\pi}{2})$, so $D(\alpha, x)$ is strictly decreasing for $x \in (0, y)$ (i.e decreasing from $D(\alpha, 0) = 0$ to $D(\alpha, y) < 0$) and $D(\alpha, x)$ is strictly increasing for $x \in (y, \frac{\pi}{2})$ (i.e. increasing from $D(\alpha, y) < 0$ to $D(\alpha, \frac{\pi}{2}) = 0$). Therefore $D(\alpha, x) < 0$ for $x \in (0, \frac{\pi}{2})$ and $\alpha > 0$. Therefore $\frac{\partial}{\partial \alpha} \lambda(\alpha, x) < 0$ for $x \in (0, \frac{\pi}{2})$ and $\alpha > 0$, which proves that $\lambda(\alpha, x)$ is strictly decreasing. Furthermore,

$$\lim_{\alpha \to 0} \lambda(\alpha, x) = \lim_{\alpha \to 0} \frac{\int_x^{\frac{\pi}{2}} \cos^\alpha t \, dt}{\int_0^{\frac{\pi}{2}} \cos^\alpha t \, dt} = 1 - \frac{2x}{\pi} \,.$$

Taking some $0 < \epsilon < x$ and using the fact that $\cos^\alpha t$ is strictly decreasing in t for every $\alpha > 0$ we obtain

$$
\begin{aligned}
\frac{1}{\lambda(\alpha, x)} &= \frac{\int_0^{\frac{\pi}{2}} \cos^\alpha t \, dt}{\int_x^{\frac{\pi}{2}} \cos^\alpha t \, dt} \\
&= \frac{\int_0^{\epsilon} \cos^\alpha t \, dt}{\int_x^{\frac{\pi}{2}} \cos^\alpha t \, dt} + \frac{\int_{\epsilon}^{\frac{\pi}{2}} \cos^\alpha t \, dt}{\int_x^{\frac{\pi}{2}} \cos^\alpha t \, dt} \\
&\geq \frac{\epsilon \cos^\alpha \epsilon}{\left(\frac{\pi}{2} - x\right) \cos^\alpha x} + 1 \,.
\end{aligned}
$$

Since $\left(\frac{\cos \epsilon}{\cos x}\right)^\alpha \to \infty$ as $\alpha \to \infty$, we obtain

$$\lim_{\alpha \to \infty} \lambda(\alpha, x) = 0 \,.$$

As $\cos^\alpha t > 0$ for $t \in (0, \frac{\pi}{2})$, we have also the monotonicity of $\int_x^{\frac{\pi}{2}} \cos^\alpha t \, dt$, i.e. (4). $\qquad \square$

From Proposition 2.3.12 we may conclude that the bivariate marginals of an elliptically distributed random vector Y are tail-dependent if and only if the spectral r.v. R in (2.13) is regularly varying. Kendall's tau τ_{ij} only affects the magnitude of the tail dependence. As a consequence of the proposition, r.v.s with a Gaussian copula are tail-independent, whereas the t-copula with ν degrees of freedom leads to a tail dependence with $\alpha = \nu$. Furthermore, due to Lemma 2.3.13 (3) we find that the smaller the tail index α of the spectral r.v. R is (i.e. the heavier the tails of $Y_1, \dots, Y_d$ are), the higher are the tail dependence coefficients. Note that the tail dependence coefficients depend on the distribution of R only through the tail index α. In the next corrolary we state the result for NVM random vectors.

Proposition 2.3.14. *Let* $Y = \mu + WZ$ *be a d-dimensional NVM random vector. Then the following statements are equivalent:*

(1) W *is regularly varying with tail index* $\alpha > 0$;

(2) Y *is regularly varying with tail index* $\alpha > 0$;

(3) for all $i, j = 1, \ldots, d$, $i \neq j$, *the vector* (Y_i, Y_j) *is tail dependent. Moreover, the coefficients of tail dependence are given by*

$$\lambda_U(Y_i, Y_j) = \lambda_L(Y_i, Y_j) = \frac{\int_{g(\rho_{ij})}^{\frac{\pi}{2}} \cos^\alpha t \, dt}{\int_0^{\frac{\pi}{2}} \cos^\alpha t \, dt} , \tag{2.17}$$

where $g(\rho_{ij}) = \frac{\pi}{4} - \frac{\arcsin \rho_{ij}}{2} = \frac{\pi}{4}(1 - \tau_{ij})$ *and* ρ_{ij} *and* τ_{ij} *are the correlation coefficient in (2.14) and Kendall's tau in (2.1), respectively.*

Proof. By Breimann's classical result (see Breimann [17]), the r.v. $R = W\sqrt{\chi_d^2}$ is regularly varying with tail index α if and only if W is (since $\sqrt{\chi_d^2}$ has moments of every order). This, together with Proposition 2.3.12 and Remark 2.3.5 leads to the required result. $\square$

Next we derive a formula for the spectral measure of a regularly varying elliptical random vector. Recall that the spectral measure as in (2.10) depends on the choice of the norm $|\cdot|$.

Proposition 2.3.15. *Let* Y *be a d-dimensional regularly varying elliptical random vector with tail index* α *and representation* $Y = RAU + \mu$ *as in (2.13). Then its spectral measure with respect to the norm* $|\cdot|$ *is given by*

$$P(\Theta \in \cdot) = \frac{E\left[|AU|^\alpha \mid \frac{AU}{|AU|} \in \cdot\right]}{E\left[|AU|^\alpha\right]} . \tag{2.18}$$

Proof. Due to Proposition 2.2.14 we may assume without loss of generality that $\mu = 0$. Set $y = 1$ in Definition 2.2.12 and note that

$$\frac{P(|RAU| > x, \frac{RAU}{|RAU|} \in \cdot)}{P(|RAU| > x)} = \frac{P(R|AU| > x, \frac{AU}{|AU|} \in \cdot)}{P(R|AU| > x)} .$$

By Theorem 2.2.15 R is regularly varying with tail index α. Therefore, by means of Proposition 2.2.4, we have that $P(R > x) = x^{-\alpha}L(x)$, where $L(x)$ is some slowly varying function. Conditioning on AU we obtain

$$\lim_{x \to \infty} \frac{P(R|AU| > x, \frac{AU}{|AU|} \in \cdot)}{P(R|AU| > x)} = \lim_{x \to \infty} \frac{E\left[x^{-\alpha}|AU|^\alpha L(x/|AU|) \mid \frac{AU}{|AU|} \in \cdot\right]}{E\left[x^{-\alpha}|AU|^\alpha L(x/|AU|)\right]} .$$

Since $\lim_{x \to \infty} \frac{L(xt)}{L(x)} = 1$ for all $t > 0$, we obtain the required result. $\square$

Proposition 2.3.15 enables us to to compute explicit formulas for the spectral measure with respect to a given norm $|\cdot|$ of a bivariate elliptical random vector. As stated in Hult and Lindskog [78], Example 5.1, every elliptical random vector (Y_1, Y_2) has representation

$$(Y_1, Y_2) = \left(\sqrt{\Sigma_{1,1}} R \cos(\varphi), \sqrt{\Sigma_{2,2}} R \sin(\varphi + \arcsin \rho_{1,2})\right),$$

where φ is uniformly distributed on $(-\pi/2, 3\pi/2)$. Then

$$RAU = R|AU|\frac{AU}{|AU|} = Rf(\varphi)(\cos g(\varphi), \sin g(\varphi)),$$

where

$$g(t) = \begin{cases} -\frac{\pi}{2} & t = -\frac{\pi}{2} \\ \arctan\left(\frac{\sqrt{\Sigma_{11}}}{\sqrt{\Sigma_{22}}}(\rho_{12} - \sqrt{1 - \rho_{12}}\tan t)\right) & t \in (-\frac{\pi}{2}, \frac{\pi}{2}) \\ g(t - \pi) + \pi & t \in (\frac{\pi}{2}, \frac{3\pi}{2}) \end{cases}$$

and $f(\varphi) \overset{d}{=} |AU|$. Denote $S(\theta_1, \theta_2) = \{(\cos t, \sin t) : \theta_1 < t < \theta_2\}$. Substituting this into (2.18) and making use of a symmetry argument we obtain

$$P_\Theta(S(\theta_1, \theta_2)) = \frac{\int_{g^{-1}(\theta_1)}^{g^{-1}(\theta_2)} f^\alpha(t) dt}{\int_0^{2\pi} f^\alpha(t) dt}.$$

With respect to the L^2-norm we have

$$f_{L^2}(t) = \sqrt{\Sigma_{11} \cos^2(t) + \Sigma_{22} \sin^2(t + \arcsin \rho_{1,2})},$$

and with respect to the max-norm we have

$$f_{max}(t) = \max(\sqrt{\Sigma_{11}}|\cos(t)|, \sqrt{\Sigma_{22}}|\sin(t + \arcsin \rho_{1,2})|).$$

In many applications, where one is interested in the extremal behaviour of Y_1, given that Y_2 is extreme relative to its marginal d.f. In this context the weighted-max-norm appears to be particularly useful. It is given by

$$|Y|_{max,\infty} = \max\left(\frac{|Y_1|}{\sqrt{\Sigma_{11}}}, \frac{|Y_2|}{\sqrt{\Sigma_{22}}}\right).$$

In this case we have

$$f_{max,\infty}(t) = \max(|\cos(t)|, |\sin(\varphi + \arcsin \rho_{1,2})|).$$

We conclude with a generalisation of Proposition 2.3.12, which considers the tail copula of an elliptical random vector, recall Definition 2.1.13.

Proposition 2.3.16. *[Klüppelberg et al. [91]] Let (Y_1, Y_2) be a bivariate regularly varying elliptical random vector with tail index $\alpha > 0$, correlation ρ and tail copula $\lambda(x, y)$. Then*

$$\lambda(x, y) = \frac{\int_{g\left((x/y)^{1/\alpha}\right)}^{\frac{\pi}{2}} x \cos^\alpha t\, dt + \int_{-\arcsin\rho}^{g\left((x/y)^{1/\alpha}\right)} y \sin^\alpha(t + \arcsin\rho)\, dt}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^\alpha t\, dt}, \qquad (2.19)$$

where $g(t) = \arctan((t - \rho)/\sqrt{1 - \rho^2}) \in [-\arcsin\rho, \frac{\pi}{2}], \ \ t \in \mathbb{R}.$ $\qquad\qquad\square$

Chapter 3

The model

We consider a portfolio credit risk model in the spirit of CreditMetrics [74] and investigate the loss distribution over a fixed time horizon T. The dependence structure in the portfolio is given by a set of underlying risk factors which we model by a general multivariate elliptical distribution with heavy-tailed marginals, introducing tail-dependence. We present the main model in Section 3.1 and outline the key model parameters. Some of the common features and of the differences between the heavy-tailed model and the CreditMetrics model are discussed and illustrated by simple numerical examples in Section 3.2. In Section 3.3 we find an expression for the moment generating function of the portfolio loss and its first two moments.

3.1 Heavy-tailed risk factors

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space which carries all random objects in this thesis. For $m \in \mathbb{N}$ let $X = (X_1, \ldots, X_m)$ be a random vector with discrete marginals, all having the same range $\{1, 2, \ldots, K\}$ – the unknown rating (the credit quality) of the credits at the time horizon T (rating 1 means default and the credit quality increases with the rating). The loss of a portfolio of m credits (loans, bonds or credit derivatives) is modelled by the r.v.

$$L = \sum_{j=1}^{m} e_j L_j \,, \tag{3.1}$$

where for $j = 1, \ldots, m$:

   - e_j is a known positive constant - **the exposure**;
   - L_j is a real-valued r.v., defined on the probability space $(\Omega, \mathcal{F}, P(\cdot \,|\, X_j))$, where $P(\cdot \,|\, X_j)$ denotes the conditional probability measure - **the loss given rating**.

   We assume further:

   (A) L_j are conditionally independent, given X;

(B) given X_j, L_j is independent of X_s for $s = 1, \ldots, m$, $s \neq j$;

(C) $C_j \leq L_j \leq 1$ a.s. for every outcome X_j, $j = 1, \ldots, m$, where $C_j < 1$ are real constants.

Most of the credit risk models used in practice fit within model (3.1) with the above assumptions. For instance, when $K = 2$ (default and non-default rating) and $L_j = 1_{\{X_j=1\}}$, L is the loss of a credit portfolio under the so called 'actuarial valuation' (see Gordy [72], Section 1). With the actuarial valuation one takes care only of the default risk, and the uncertainty in the recovery of a credit in the event of default is ignored. An extension to random recovery rates has been considered by various authors, see for example Bluhm et al. [16], Section 1.1.3. A further extension to multiple ratings is necessary for the so called 'mark-to-market' valuation, see Gordy [72], Section 3, or CreditMetrics [74].

The complexity of model (3.1) is in the joint distribution of $X = (X_1, \ldots, X_m)$. We denote the marginal default and rating migration probabilities by $P(X_j = k) = p_{j,k}$ and

$$P\left(X_j \leq s\right) = \sum_{k=1}^{s} p_{j,k} = p_j^s, \ \ s = 1, \ldots, K, \ j = 1, \ldots, m. \tag{3.2}$$

In order to model the dependence structure of $X = (X_1, \ldots, X_m)$ we introduce the random vector $Y = (Y_1, \ldots, Y_m)$ with continuous marginal distributions G_j and a copula C, i.e. the multivariate d.f. of Y is given by

$$G_Y(y_1, \ldots, y_m) = C(G_1(y_1), \ldots, G_m(y_m)). \tag{3.3}$$

For $j = 1, \ldots, m$ the r.v. Y_j is interpreted as asset (log-)return of obligor j in the portfolio.

Following the approach in CreditMetrics [74], we set for $j = 1, \ldots, m$

$$X_j = k \iff G_j^{-1}(p_j^{k-1}) < Y_j \leq G_j^{-1}(p_j^k), \ k = 1, \ldots, K, \tag{3.4}$$

where we interpret $G_j^{-1}(p_j^0) = -\infty$ and $G_j^{-1}(p_j^K) = \infty$.

In the following proposition we show that the calibration of the distribution of $X = (X_1, \ldots, X_m)$ can be reduced to the calibration of the marginal default and migration probabilities and the copula of $Y = (Y_1, \ldots, Y_m)$ (see Frey and McNeil [59], Proposition 3.3 for a less general version).

Proposition 3.1.1. *The d.f. of X is uniquely determined by the marginal probabilities in (3.2) and the copula C of Y.*

Proof. Observe that

$$
\begin{aligned}
P(X_1 = x_1, \ldots, X_m = x_m) &= P\left(\bigcap_{j=1}^{m} \left\{G_j^{-1}(p_j^{x_j-1}) < Y_j \leq G_j^{-1}(p_j^{x_j})\right\}\right) \\
&= P\left(\bigcap_{j=1}^{m} \left\{p_j^{x_j-1} < G_j(Y_j) \leq p_j^{x_j}\right\}\right)
\end{aligned}
$$

From Sklar's Theorem 2.1.2 we have

$$P(X_1 = x_1, \ldots, X_m = x_m) = P\left(\bigcap_{j=1}^{m}\left\{p_j^{x_j-1} < U_j \leq p_j^{x_j}\right\}\right)$$

where $(U_1, \ldots, U_m)$ are uniformly distributed with copula C, i.e.

$$P(X_1 = x_1, \ldots, X_m = x_m) = \sum_{s_1=0}^{1} \cdots \sum_{s_m=0}^{1} (-1)^{\sum_{j=1}^{m} s_j} C\left(p_1^{x_1-s_1}, \ldots, p_m^{x_m-s_m}\right).$$

$\square$

We focus on the dependence structure of the asset returns $Y_1, \ldots, Y_m$, i.e. on their copula. We assume that they follow a linear factor model with a multiplicative random shock:

$$Y_j = \sum_{l=1}^{p} \alpha_{j,l} W Z_l + \sigma_j W \epsilon_j, \ j = 1, \ldots, m. \tag{3.5}$$

We have used the following notations.

- $Z = (Z_1, \ldots, Z_p)$ is p-dimensional multivariate normal with standard normal $N(0,1)$ marginals and correlation matrix Σ – **the regional and business sector common factors**.

- W is a positive r.v., independent of Z. It represents **a global shock** affecting all assets across regions and business sectors.

- ϵ_j, $j = 1, \ldots, m$, are i.i.d $N(0,1)$, independent of W and Z – **the obligor-specific risk factors**.

- the constants $\alpha_{j,l} \in \mathbb{R}$ and $\sigma_j > 0$, $j = 1, \ldots, m$, $l = 1, \ldots, p$ are **the factor loadings**. They are normalized so that var$[Y_j \,|\, W] = W^2$.

Given (3.5), $Y \in N_m(0, W^2 \Sigma_Y)$ – normal variance mixture (NVM) distribution with mixing variable W, or otherwise called, multivariate elliptical distribution, see Definition 2.3.4 and the remarks after it. The simplest special case of (3.5) is the one-factor Gaussian model, obtained when $W = 1$ a.s. and $p = 1$. This model has been investigated by various authors, see for instance Bluhm et al. [16], Section 2.5.1. The popular in practice model CreditMetrics [74] can be obtained from (3.5) by setting $W = 1$ a.s.

The marginal distributions of Y_j are all one-dimensional NVMs, i.e. $Y_j \overset{d}{=} W Z_0$, where W is defined as above and $Z_0 \in N(0,1)$, $Z_0 \perp W$. Since in our credit risk model the risk factors Y_j are needed only to introduce the dependence structure in the portfolio, of crucial importance is their copula C, see Proposition 3.1.1. Given (3.5), this copula is an *elliptical copula*. It is uniquely determined (see Corrolary 2.3.10) by the correlation matrix Σ_Y and the d.f. of the global shock W. The correlation matrix Σ_Y can be computed in terms of

the correlation matrix Σ of the common factors $(Z_1, \ldots, Z_p)$ and the factor loadings $\alpha_{j,l}$, $j = 1, \ldots, m$, $l = 1, \ldots, p$. It is the same as in the Gaussian case ($W = 1$ a.s.), i.e.

$$\Sigma_{Y_i, Y_j} = \sum_{l=1}^{p} \sum_{q=1}^{p} \alpha_{j,l} \alpha_{i,q} \Sigma_{l,q} \,,$$

see Lemma 2.3.7. However, model (3.5) has certain important differences from the Gaussian model, which we investigate in the next section.

3.2   Heavy tails vs CreditMetrics – a first look

The part $\sum_{l=1}^{p} \alpha_{j,l} W Z_l$ in (3.5) is frequently referred to as the systematic part of the risk and $\sigma_j W \epsilon_j$ as the specific part of the risk for credit j. Note that, in contrast to CreditMetrics, the specific parts in our case are no longer independent of the systematic part, nor between each other. They are uncorrelated, but depend through the r.v. W. Recall that we interpret W as a global shock affecting simultaneously all obligors across countries and industries.

We are particularly interested in model (3.5), when W is regularly varying, i.e. for all $t > 0$

$$\lim_{w \to \infty} \frac{P(W > tw)}{P(W > w)} = t^{-\alpha} \tag{3.6}$$

for some $\alpha > 0$. As shown in Theorem 2.3.14, only in this case Y_i and Y_j, $i \neq j$, exhibit tail dependence, i.e.

$$\lim_{p \to 0} \frac{P\left(Y_i < G_i^{-1}(p), Y_j < G_j^{-1}(p)\right)}{p} > 0 \,.$$

Note that by (3.4) the probability of joint default of credits i and j is given by

$$P(X_i = 1, X_j = 1) = P\left(Y_i < G_i^{-1}(p_i^1), Y_j < G_j^{-1}(p_j^1)\right) \,.$$

Taking into account that usually the default probabilities p_i^1 and p_j^1 are small, the pairwise tail dependence of assets Y_i and Y_j results in an increased likelihood for simultaneous defaults in the credit portfolio. This has an important impact on the credit loss distribution, in particular on its tail (see Frey and McNeil [59] and Section 5.3.2 for some numerical examples).

Example 3.2.1. [t-model]
The most frequently used model including a r.v. W satisfying (3.6) is the *t-model* ($p \geq 1$, $W = \sqrt{\frac{\nu}{S_\nu}}$, where $S_\nu \in \chi_\nu^2$ – chi-square distribution with ν degrees of freedom). Then $Y \in T_m(0, \Sigma_Y, \nu)$ (multivariate *t*-distribution with ν degrees of freedom). This means that $\alpha = \nu$ in (3.6).

In what follows we investigate further the impact of the heavy-tailed assumption on the dependence structure of a credit portfolio. For the sake of simplicity we take $K = 2$ in (3.1) (default and non-default ratings only), and consider the implied by the model dependence of the binary r.v.s $L_j = 1_{\{X_j=1\}}$, $j = 1, \ldots, m$. The dependence among these default indicators is frequently measured by their correlation coefficient, see e.g. CreditMetrics [74], Chapter 2. However, the odds ratio defined in (2.5) could be more appropriate in such cases, since we are dealing with discrete binary events. In our case, this quantity has the interpretation of relative risk, see the comments after Definition 2.1.14. If $\mathrm{OR}(L_i, L_j)$ is close to 1, this would mean that the credits are close to being independent. Otherwise, if $\mathrm{OR}(L_i, L_j)$ is large, this would indicate a high dependence in the portfolio. In other words, if, under a reasonable parametrization, model (3.5) implies through (3.4) odds ratios close to 1, we would be able to conclude, that the dependence in the assets Y has a small impact on the portfolio loss distribution, and vise versa.

We start by computing a general formula for the odds ratios in our model.

Proposition 3.2.2. *Assume that $K = 2$ and $L_j = 1_{\{X_j=1\}}$, $j = 1, \ldots, m$, in model (3.1) with (3.4). Denote by $C_{i,j}(u, v)$ the copula of the bivariate random vector (Y_i, Y_j), for $i, j = 1, \ldots, m$. Then for the odds ratio defined in (2.5) we have*

$$\mathrm{OR}(L_i, L_j) = \frac{C_{i,j}(p_i^1, p_j^1)(1 - p_i^1 - p_j^1 + C_{i,j}(p_i^1, p_j^1))}{(p_i^1 - C_{i,j}(p_i^1, p_j^1))(p_j^1 - C_{i,j}(p_i^1, p_j^1))}, \quad i, j = 1, \ldots, m.$$

Proof. Due to (3.4) we have

$$P(L_i = 1, L_j = 1) = P(Y_i < G_i^{-1}(p_i^1), Y_j < G_i^{-1}(p_i^1)) = C_{i,j}(p_i^1, p_j^1).$$

Then

$$P(L_i = 1, L_j = 0) = P(L_i = 1) - P(L_i = 1, L_j = 1) = p_i^1 - C_{i,j}(p_i^1, p_j^1)$$

and

$$P(L_i = 0, L_j = 0) = P(L_i = 0) - P(L_i = 0, L_j = 1) = 1 - p_i^1 - p_j^1 + C_{i,j}(p_i^1, p_j^1).$$

Substituting in (2.5) we obtain the required result. □

Remark 3.2.3. From Proposition 3.2.2 we see that the odds ratios in our model are directly related to other dependence measures. For instance, the linear correlation is given by

$$\rho(L_i, L_j) = \frac{C_{i,j}(p_i^1, p_j^1) - p_i^1 p_j^1}{\sqrt{(1 - p_i^1)(1 - p_j^1)p_i^1 p_j^1}}, \quad i, j = 1, \ldots, m.$$

The Kendalls tau is given by

$$\tau(L_i, L_j) = C_{i,j}(p_i^1, p_j^1)(1-p_i^1-p_j^1+C_{i,j}(p_i^1, p_j^1))-(p_i^1-C_{i,j}(p_i^1, p_j^1))(p_j^1-C_{i,j}(p_i^1, p_j^1)) ,\ \ i,j = 1,\ldots, m .$$

In all cases of key importance is the object $C_{i,j}(p_i^1, p_j^1),\ i, j = 1, \ldots, m$.   □

We continue with an example demonstrating the impact of the heavy-tailed model on the odds ratio.

Example 3.2.4. [Odds ratios, heavy tails vs CreditMetrics]
Consider model (3.1) with (3.4) and (3.5). Assume that:

- K=2 and $L_j = 1_{\{X_j=1\}}$, $j = 1, \ldots, m$ (default-only model).

- The credits are split into 6 groups $R_1, \ldots, R_6$ with $p_j^1 = 0.001\%, j \in R_1$; $p_j^1 = 0.0055\%, j \in R_2$; $p_j^1 = 0.0288\%, j \in R_3$; $p_j^1 = 0.151\%, j \in R_4$; $p_j^1 = 0.7916\%, j \in R_5$; $p_j^1 = 4.1996\%, j \in R_6$. Such parametrization is typical in practice, see e.g. Bluhm et al. [16], Section 2.7. More precisely, in Moody's rating notation, we have the group R_1 corresponding to the rating category Aaa, R_2 to Aa and so on up to R_6 corresponding to the rating category B.

- There is one common factor in (3.5) ($p = 1$) and the factor loadings are

$$\alpha_{j,1} = \sqrt{1 - \sigma_j^2} = \sqrt{\rho_k},\ j \in R_k,\ k = 1, \ldots, 6.$$

This means that the assets Y_i and Y_j have correlation $\rho_{ij} = \sqrt{\rho_k \rho_s}$ when $i \in R_k$ and $j \in R_s$.

In Figure 3.1 we plot the odds ratios of the default indicators (L_i, L_j), for i, j being in various groups, as a function the correlation ρ_{ij} of the assets (Y_i, Y_j). The left column corresponds to the odds ratios for the CreditMetrics model ($W = 1$ a.s. in (3.5)) and the right column corresponds to the heavy-tailed t-model with $\nu = 4$ degrees of freedom (Example 3.2.1). The selected correlation ranges are also typical in practice, see Bluhm et al. [16], Section 2.7.

We observe that in both models and in all groups the odds ratios are quite high. This means that in any case, the dependence in the assets Y induces a significant dependence between the default indicators. Hence, we may reasonably expect, that this dependence is important for the portfolio loss distribution. Further, we note that the odds ratios increase significantly with the increase of the credit quality (notice the difference in the scales of the vertical axes in row 1 (good quality obligors from groups R_1, R_2) to row 3 (low quality obligors from groups R_5, R_6).

By comparing the left and the right column in Figure 3.1, we observe that in the heavy-tailed t-model, the odds ratios are less sensitive to the correlation parameter than in the CreditMetrics model. However, in all cases the t-model induces higher odds ratios

than the CreditMetrics model. Therefore, the heavy tails have significant impact on the portfolio loss distribution. □

The fact that the default probabilities are typically quite small allows for an approximation of the odds ratios $\mathrm{OR}(L_i, L_j)$ in the heavy-tailed case. Recall Proposition 2.3.16 which gives an explicit formula for the tail copula of a regularly varying elliptical random vector. Furthermore, note that by Proposition 3.2.2, the odds ratio can be computed through the copula function $C_{i,j}(p_i^1, p_j^1)$, where $p_j^1, j = 1, \ldots, m$, are the (small) marginal default probabilities. Denote

$$p_{i,j} = \frac{p_i^1 + p_j^1}{2}, i, j = 1, \ldots, m,$$

and note that $p_{i,j} \to 0$ when both p_i^1 and p_j^1 tend to 0. Set $x_{i,j} = p_{i,j}/p_i^1$ and approximate $C_{i,j}(p_i^1, p_j^1)$ by $p_{i,j}\lambda(x_{i,j}, x_{j,i})$, where $\lambda(x, y)$ is the tail copula of an elliptical random vector given in Proposition 2.3.16. Substituting in the formula in Proposition 3.2.2 we obtain an approximation for the odds ratios. In the next example we investigate the accuracy of the suggested approximation.

Example 3.2.5. [Approximation of odds ratios, Example 3.2.4 continued]
Consider model (3.1) with (3.4) and (3.5) and the parameters from Example 3.2.4. Recall that the t-model satisfies (3.6) with $\alpha = \nu = 4$.

In Figure 3.2 we compare the true odds ratios to the above suggested approximation. In all cases, the approximation is reasonably accurate, and in particular for the credits from the top quality groups R_1 and R_2 it cannot be distinguished from the true odds ratio. Note that the approximation depends (apart from the marginal default probabilities) only on the correlation coefficients ρ_{ij} and on the tail index α of the assets Y. Hence, we may conclude that these are the parameters of the copula of Y with important impact on the portfolio loss distribution. □

3.3 Moment generating functions

The moment generating function (m.g.f.) of a r.v. (provided it exists) determines completely the distribution of the r.v. Therefore, knowing the m.g.f. in an explicit form is sufficient to compute the d.f. by numerical methods, see for instance Duhamel and Vetterli [42]. Furthermore, the m.g.f. is the basis for providing various tail approximations, incl. the saddlepoint approximation (see Jensen [82]) or the classical Cramer-Lundberg upper bound for the ruin probability in insurance mathematics (see e.g. Asmussen [7], Section 3.5). These techniques have been extensively applied in the analysis of another class of credit risk models - the so called actuarial approach, see CSFB [25] or Gordy [73].

Unfortunately, even in the most simple special cases of the CreditMetrics model, the m.g.f. of the portfolio loss cannot be computed explicitly, see Finger [52]. An application of the saddlepoint approximation or the FFT is, therefore, subject to complicated numerical procedures, and is feasible only in some special cases of our general model (3.1) with (3.4) and (3.5), see Martin et al. [109]. However, in this section we derive a (non-explicit) expression for the m.g.f. of L which we use extensively in Section 6.2. The expression is essentially based on the fact that, given the global shock W and the common factors Z, the individual credits in the portfolio are independent, see the proofs below.

We note first that due to assumption (C) in model (3.1), the portfolio loss L has bounded support

$$L_{min} \leq \sum_{j=1}^{m} C_j e_j \leq L = \sum_{j=1}^{m} e_j L_j \leq \sum_{j=1}^{m} e_j = L_{max} \, . \tag{3.7}$$

From now on we exclude some degenerate cases and we suppose that for every $x < L_{max}$ we have $P(L > x) > 0$. This implies, for instance, that if the distribution of L is discrete, then $P(L = L_{max}) > 0$.

Proposition 3.3.1. *Assume model (3.1) with (3.4) and (3.5). Then the m.g.f $\varphi(\theta) = E[\exp(\theta L)]$ exists for every $\theta \in \mathbb{R}$ and is given by*

$$\varphi(\theta) = E[\exp(H(W, Z, \theta))] \tag{3.8}$$

with

$$H(W, Z, \theta) = \sum_{j=1}^{m} \log H_j(W, Z, \theta) \, , \tag{3.9}$$

where for $j = 1, \ldots, m$,

$$H_j(W, Z, \theta) = E[\exp(\theta e_j L_j) \mid W, Z] = \sum_{k=1}^{K} g_{j,k}(W, Z)\varphi_{j,k}(e_j \theta). \tag{3.10}$$

Furthermore, for all $j = 1, \ldots, m$, $k = 1, \ldots, K$

$$g_{j,k}(w, z) = \Phi\left(\frac{G_j^{-1}(p_j^k)}{\sigma_j}\frac{1}{w} - \sum_{l=1}^{p}\frac{\alpha_{j,l}}{\sigma_j}z_l\right) - \Phi\left(\frac{G_j^{-1}(p_j^{k-1})}{\sigma_j}\frac{1}{w} - \sum_{l=1}^{p}\frac{\alpha_{j,l}}{\sigma_j}z_l\right), \tag{3.11}$$

where we interpret the second term as 0 for $k = 1$ and the first term as 1 for $k = K$; and, finally

$$\varphi_{j,k}(\theta) = E[\exp(\theta L_j) \mid X_j = k] \, . \tag{3.12}$$

Proof. Due to (3.7) we obtain immediately for $\theta > 0$

$$\varphi(\theta) = E\left[\exp(\theta L)\right] \leq \exp(\theta L_{max}),$$

therefore $\varphi(\theta)$ exists for every $\theta \geq 0$. Also we have by (3.7) for $\theta < 0$

$$\varphi(\theta) = E\left[\exp(\theta L)\right] \leq \exp(\theta L_{min}),$$

therefore $\varphi(\theta)$ exists for every $\theta \in \mathbb{R}$.

By conditioning on X we have

$$\varphi(\theta) = E\left[\exp(\theta L)\right] = E_X\, E\left[\exp(\theta L) \mid X\right].$$

By assumption (A) in (3.1) L_j, $j = 1, \ldots, m$, are independent, given X. Therefore

$$\varphi(\theta) = E_X\left[\prod_{j=1}^{m} E\left[\exp(\theta e_j L_j) \mid X\right]\right].$$

Since by assumption (B) in (3.1) L_j, given X_j, is independent of X_s for $j = 1, \ldots, m$ and $s = 1, \ldots, m$, $s \neq j$, we get

$$\varphi(\theta) = E_X\left[\prod_{j=1}^{m} E\left[\exp(\theta e_j L_j) \mid X_j\right]\right].$$

Due to (3.5), given W and Z, the r.v.s Y_j, $j = 1, \ldots, m$, are independent (inherited by the independence of ϵ_j). Therefore X_j, $j = 1, \ldots, m$, are conditionally independent by means of (3.4). Hence, by conditioning on W and Z we get

$$\begin{aligned}
\varphi(\theta) &= E_{W,Z}\left[E_X\left[\prod_{j=1}^{m} E\left[\exp(\theta e_j L_j) \mid X_j\right] \mid W, Z\right]\right] \\
&= E_{W,Z}\left[\prod_{j=1}^{m} E_{X_j}\left[E\left[\exp(\theta e_j L_j) \mid X_j\right] \mid W, Z\right]\right].
\end{aligned}$$

Given (3.5),

$$P\left(X_j = k \mid W = w, Z = z\right) = P\left(G_j^{-1}(p_j^{k-1}) \leq Y_j < G_j^{-1}(p_j^{k}) \mid W = w, Z = z\right)$$

$$= P\left(G_j^{-1}(p_j^{k-1}) \leq \sum_{l=1}^{p} \alpha_{j,l} W Z_l + \sigma_j W \epsilon_j < G_j^{-1}(p_j^{k}) \mid W = w, Z = z\right)$$

$$= P\left(\frac{G_j^{-1}(p_j^{k-1})}{\sigma_j}\frac{1}{w} - \sum_{l=1}^{p}\frac{\alpha_{j,l}}{\sigma_j}z_l \leq \epsilon_j < \frac{G_j^{-1}(p_j^{k})}{\sigma_j}\frac{1}{w} - \sum_{l=1}^{p}\frac{\alpha_{j,l}}{\sigma_j}z_l\right)$$

$$= g_{j,k}(w, z).$$

Therefore

$$E_{X_j}\left[E\left[\exp\left(\theta e_j L_j\right)\mid X_j\right]\mid W,\,Z\right] = \sum_{k=1}^{K} g_{j,k}(W,Z)\varphi_{j,k}(e_j\theta)$$

and hence

$$\varphi\left(\theta\right) = E_{W,Z}\left[\prod_{j=1}^{m}\sum_{k=1}^{K} g_{j,k}(W,Z)\varphi_{j,k}(e_j\theta)\right].$$

By assumption (C) in (3.1) $C_j \leq L_j \leq 1$ and $\varphi_{j,k}(e_j\theta)$ is finite for every $\theta \in \mathbb{R}$, $j = 1,\ldots,m$, $k = 1,\ldots,K$. Therefore we get the required result. $\square$

Next we derive formulas for the mean and the variance of L.

Proposition 3.3.2. *Denote $\mu_{j,k} = E\left[L_j \mid X_j = k\right]$. Under the assumptions of model (3.1) with (3.4) and (3.5), the mean of L is*

$$E\left[L\right] = \sum_{j=1}^{m}\sum_{k=1}^{K} e_j p_{j,k}\mu_{j,k}\,,$$

and the variance is

$$\mathrm{var}(L) = \sum_{j,l=1}^{m}\sum_{k,s=1}^{K} e_j e_l E\left[g_{j,k}\left(W,Z\right) g_{l,s}\left(W,Z\right)\right]\left(\mu_{j,k} - E[L_j]\right)\left(\mu_{j,s} - E[L_j]\right),$$

where $g_{j,k}\left(W,Z\right)$, $j = 1,\ldots,m$, $k = 1,\ldots,K$ are defined in (3.11). Furthermore, the conditional mean, given the global shock W and the common factors $Z = (Z_1,\ldots,Z_p)$, is

$$E\left[L\mid W,Z\right] = \sum_{j=1}^{m}\sum_{k=1}^{K} e_j g_{j,k}(W,Z)\mu_{j,k}\,.$$

Proof. We have

$$E\left[L\right] = E\left[E\left[L\mid X\right]\right] = \sum_{j=1}^{m} e_j\, E\left[E\left[L_j\mid X\right]\right] = \sum_{j=1}^{m}\sum_{k=1}^{K} e_j p_{j,k} E\left[L_j\mid X_j = k\right]. \qquad (3.13)$$

For the variance of L we have

$$\mathrm{var}\left(L\right) = \sum_{j,l=1}^{m} e_j e_l \mathrm{cov}(L_j,L_l)\,. \qquad (3.14)$$

By assumptions (A) and (B) in (3.1) we obtain

$$
\begin{aligned}
\mathrm{cov}(L_j,L_l) &= E\left[E[(L_j - E[L_j])(L_l - E[L_l])\mid X]\right]\\
&= E\left[E[L_j - E[L_j]\mid X_j]\, E[L_l - E[L_l]\mid X_l]\right]\\
&= \sum_{k,s=1}^{K} P(X_j = k, X_l = s)(E[L_j|X_j = k] - E[L_j])(E[L_l|X_l = s] - E[L_l]).
\end{aligned}
$$

Using (3.4) and (3.5) and similar arguements as in the proof of Proposition 3.3.1 we have for $k, s = 1, \ldots, K,\ j, l = 1, \ldots, m$

$$
\begin{aligned}
P(X_j = k, X_l = s) &= E[P(X_j = k, X_l = s \,|\, W, Z)] \\
&= E\left[g_{j,k}\left(W, Z\right) g_{l,s}\left(W, Z\right)\right]
\end{aligned}
$$

and

$$
P(X_j = k \,|\, W, Z) = g_{j,k}(W, Z)
$$

which lead to the required results. $\qquad\square$

Note that the mean is available explicitly, while for the variance one needs to compute numerically the integrals $E\left[g_{j,k}\left(W, Z\right) g_{l,s}\left(W, Z\right)\right],\ k, s = 1, \ldots, K,\ j, l = 1, \ldots, m$. In the CreditMetrics model ($W = 1$ a.s.), using the properties of the normal distribution, the computation of such an integral turns into a relatively simple integration w.r.t. a two-dimensional normal distribution, even when $p > 2$. This happens because in (3.11), only a simple linear combination of the common factors $Z_1, \ldots, Z_p$ appears. In the heavy-tailed case, we have an additional integration w.r.t. the distribution of the global shock W.
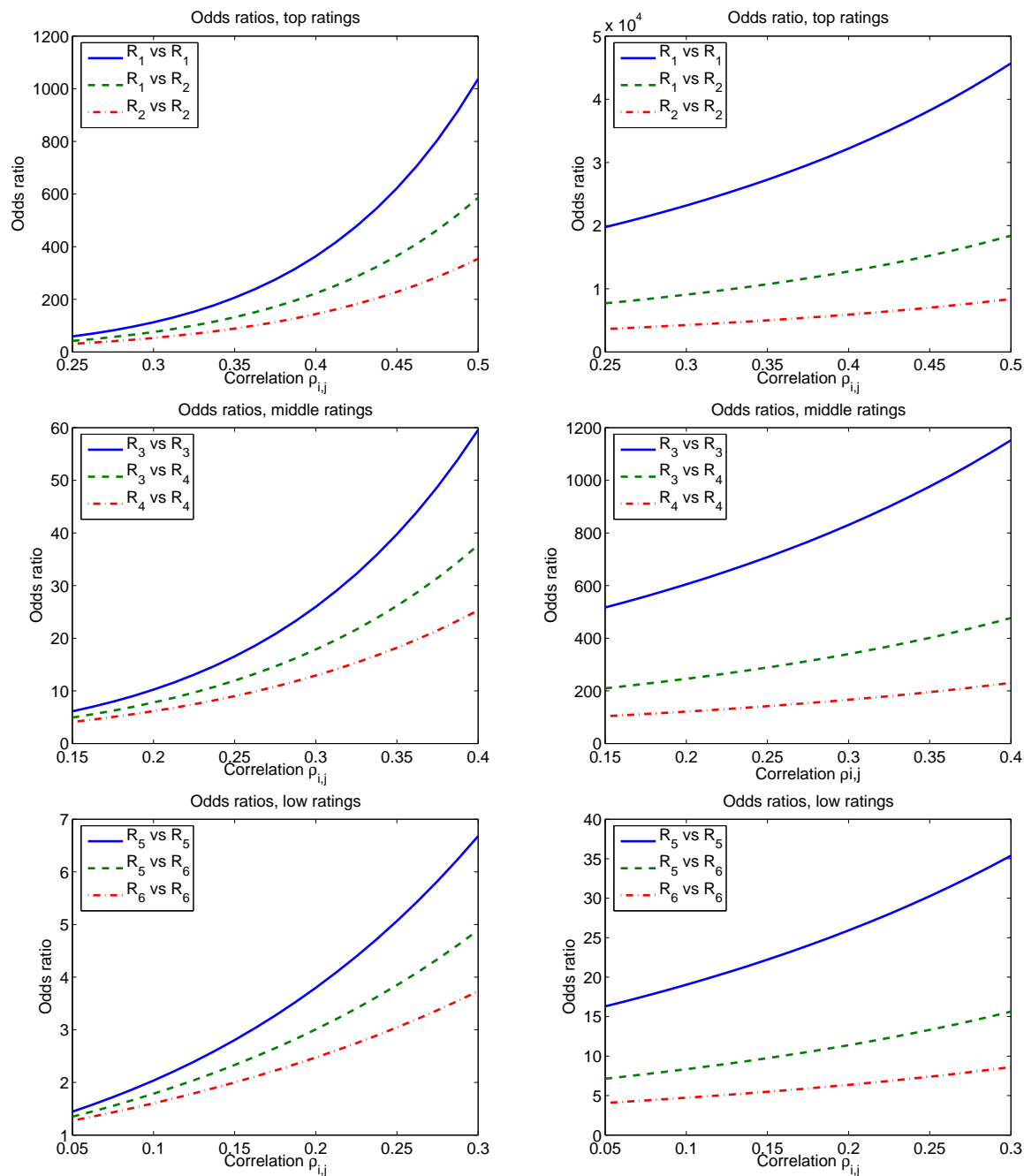
Figure 3.1: The parameters are given in Example 3.2.4

Left column: The odds ratios for credits of various groups (rating categories) as a function of the asset correlation $\rho_{i,j}$ for the standard CreditMetrics model. In all cases the odds ratio increases with the correlation. Furthermore, it is significantly large, in particlular for the high-quality credits on the first row.

Right column: The odds ratios for credits of various groups (rating categories) as a function of the asset correlation $\rho_{i,j}$ for the heavy-tailed t-model. In all cases the odds ratio increases with the correlation, however, it is less sensitive to it (compare the right to the left column). The t-model implies significantly higher odds ratios than the CreditMetrics model for all groups.
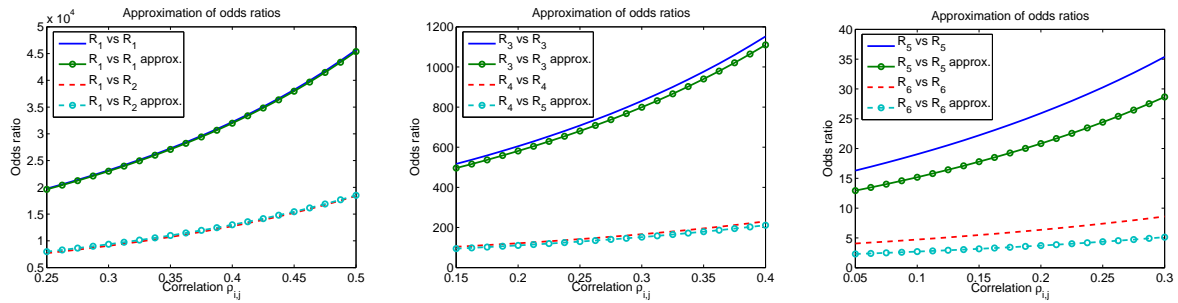
Figure 3.2: The parameters are given in Example 3.2.5 (the same as in Figure 3.1)
The odds ratios for credits of various rating categories as a function of the asset correlation $\rho_{i,j}$ for the t-model, compared to the approximation using tail dependence. The approximation is accurate, in particular for the high quality credits in the left plot.

Chapter 4

Calibration

The practical implementation of any portfolio credit risk model depends crucially on the effectiveness of the statistical procedures for estimation of the model parameters. In model (3.1) with (3.4) and (3.5) the parameter space includes:

(1) the marginal distributions of the ratings X_j, $j = 1, \ldots, m$;

(2) the marginal distributions of the losses L_j, given X_j, $j = 1, \ldots, m$;

(3) the copula of the assets Y (see Proposition 3.1.1). By means of (3.5), this copula is determined by the distribution of the global shock W, the correlation matrix Σ of the common factors $Z_1, \ldots, Z_p$ and the factor loadings $\alpha_{j,l}, \sigma_j$, $j = 1, \ldots, m$, $l = 1, \ldots, p$.

In the next section we describe briefly the available approaches regarding the one-dimensional marginal parameters (1) and (2). We continue with the bivariate case and in Section 4.2 we introduce several estimation methods for important dependence measures for two-dimensional random vectors, including the linear correlation, Kendall's tau and the tail dependence coefficients. Particular attention is paid to the estimation of the tail dependence coefficient, where a new method aiming at a lower variance of the estimates is suggested. In Section 4.3 we consider the calibration of a multidimensional elliptical copula, based on i.i.d. observations of a random vector with arbitrary continuous marginals. We start with several semi-parametric and non-parametric methods for dealing with the marginals of the observed random vector. We continue with two classical methods for copula estimation. The main result in this section is a new calibration procedure for an elliptical copula, which makes extensive use of the information contained in the joint extreme observations. In Section 4.4 we apply the procedure to calibrate model (3.5). An additional result regarding the estimation of the factor loadings $\alpha_{j,l}, \sigma_j$, $j = 1, \ldots, m$, $l = 1, \ldots, p$ in (3.5) is provided. We conclude this chapter with an extended simulation study and some real data examples.

4.1   Marginal parameters

4.1.1   Rating migration probabilities

The possibility that major rating agencies will change the credit rating of an obligor is an important source of credit risk, above and beyond its implications for the direct risk of default. Changes in the credit ratings may have an immediate effect on the values of defaultable bonds portfolios, may indicate that certain credits are no longer admissible for investors, who are subject to restrictions on the ratings of their credits, and may lead even to mandatory termination of some financial contracts. Some corporate bonds have coupon rates linked explicitly to the credit rating. The BIS capital accord determines the capital requirements of the regulated banks based in part on the credit ratings of the obligors they have in their portfolios. For these and related reasons, a model for the risk of a rating change is a key ingredient of the credit risk management system.

In what follows we model the credit rating of an obligor as a continuous-time stochastic process $(X(t))_{t \geq 0}$ with state space $\{1, \ldots, K\}$, where state 1 means default (absorbing state), and the credit quality increases with the rating. This numeric notion is chosen only for notational convenience; if we use e.g. Moody's rating categories, we would set $K = 8$, $X = 8$ corresponding to rating Aaa, $X = 7$ to Aa and so on until $X = 2$ Caa and $X = 1$ default (D).

Furthermore, we assume that the rating process follows the dynamics

$$dX(t) = (V_{X(t)} - X(t))dN_{X(t)}(t)\,, \ t \geq 0\,,$$

where $N = (N_1, \ldots, N_K)$ is a set of independent time-homogeneous Poisson processes with intensities $\lambda_1, \ldots, \lambda_K$ and $V_1, \ldots, V_K$ is a set of mutually independent and independent of N discrete r.v.s with d.f.s, for $k = 1, \ldots, K$,

$$P(V_k = j) = \frac{\lambda_{kj}}{\lambda_k} \ j = 1, \ldots, K, \, j \neq k\,,$$

with λ_{kj} being some positive constants summing up to λ_k.

Under these assumptions, the credit rating process $X(t)$ is a time-homogeneous Markov process, see e.g. Schönbucher [129], Section 8.2.3. More precisely, we have that for all $T > t$

$$P(X(T) = j \,|\, \mathcal{F}_t) = P(X(T) = j \,|\, X(t))\,, \ \ j = 1, \ldots, K\,, \tag{4.1}$$

and, denoting by $Q(t, T) = [q_{kj}(t, T) = P(X(T) = j \,|\, X(t) = k)]_{k, j = 1, \ldots, K}$, we have also

$$Q(t, T) = Q(T - t, 0) := Q(T - t)\,. \tag{4.2}$$

In particular, $Q(t) = \exp(\Lambda t)$, $t \geq 0$, where $\Lambda = [\lambda_{jk}]_{j, k = 1, \ldots, K}$ with $\lambda_{kk} = -\lambda_k$, $k = 1, \ldots, K$. The matrix Λ is often called *generator matrix*. It describes completely the stochastic behaviour of the rating process X.

The calibration of the generator matrix Λ is complicated by the fact that usually the rating agencies publish only rating migrations over a given time period like 1 year. The most common method for estimation in this case is based upon the observed behaviour of groups of obligors within the same initial rating, the cohorts. Denote by $n_k(t)$, $k = 1, \ldots, K$, the total number of obligors of rating k at time t, and by $n_{kj}(t_1, t_2)$ the number of observed migrations of obligors with rating k at time t_1 to rating j at time $t_2 > t_1$. Due to the Markov chain dynamics of the rating process, these migration events can be viewed as outcomes of $n_k(t_1)$ independent multinomial trials. With rating migration observations over T periods, the maximum likelihood estimator for the one-period migration probabilities $q_{kj}(1)$ is

$$\widehat{q}_{kj}(1) = \frac{1}{T} \sum_{t=1}^{T} \frac{n_{kj}(t-1, t)}{n_k(t-1)} \, ,$$

see e.g. Johnson and Kotz [84], Chapter 3 for confidence intervals.

This standard estimator has certain weak points. First, it does not utilize the full continuous-time information on the rating migrations which could be available at a rating agency or at a bank. In particular, when it comes to the rare events like a default of a high-rated obligor, the method usually produces 0 as an estimate. This is due to the fact that in practice a default of such an obligor typically happens not immediately within one period, but through a sequence of rating downgrades. This information cannot be captured by the standard estimator.

In addition, estimating a one-period migration probability matrix $Q(1)$ by $\widehat{Q}(1)$ does not automatically imply an estimate $\widehat{\Lambda}$ for the generator matrix Λ. Israel et al. [79] consider the problem of finding a generator matrix $\widehat{\Lambda}$ consistent with $\widehat{Q}(1)$. It is well-known that for certain stochastic matrices Q, there exists no generator matrix. Israel et al. [79] provide sufficient conditions on $\widehat{Q}(1)$ for the existence of a generator. Furthermore, there can be distinctly different generator matrices consistent with the same migration matrix $\widehat{Q}(1)$, implying different estimates for migration probabilities $\widehat{Q}(t)$, for all $t \neq 1$. The problem is known as embedding problem for continuous-time Markov chains. It is studied in detail by Kreinin and Sidelnikova [95], who give also a survey on the available numerical methods for it.

To avoid these problems, Lando and Skodeberg [98] propose a direct maximum likelihood estimator for the generator matrix. It requires knowledge of the precise points in time (the exact dates) at which rating migrations take place.

For $k, j = 1, \ldots, K$, denote by $m_{kj}(t)$ the total number of migrations from rating k to rating j *throughout* the period $[0, t]$. Recall the notation $n_k(t)$ as above. Then a maximum

likelihood estimator for the migration intensities in the generator matrix Λ is

$$\widehat{\lambda}_{kj} = \frac{m_{kj}(T)}{\int_0^T n_k(t)dt}\,, \ k, j = 1, \ldots, K\,,$$

see Lando and Skodeberg [98] for asymptotic properties and confidence bounds.

To conclude this section, we note that with respect to the rating modelling, the time-homogeneos Markov chain assumptions are quite strong and cannot be confirmed by empirical data. The default and migration probabilities vary significantly with the business cycle, as documented by Nickell et al. [120] and Kavvathas [87]. This contradicts to the time-homogenuity assumption (4.2). In addition, Behar and Nagpal [12], Lando and Skodeberg [98], and Kavvathas [87] all find that, for obligors of certain ratings, the prior rating is an important determinant of the likelihood of a downgrade (and default) v.s that of an upgrade, over a given time horizon. There is indeed an apparent momentum in rating migration data, which contradicts to (4.1). Also, there is a significant aging effect (dependence of the migration probabilities on the duration in a rating category and/or the age of the credit), documented by Carty and Fons [20], Lando and Skodeberg [98], and Kavvathas [87]. Further information for the obligors available from the rating agencies like watch-lists and outlooks could also be important for accurate estimation of the default and migration probabilities, see Cantor and Hamilton [19]. To summarize, different obligors of the same rating have different default and migration probabilities, and a given obligor of a fixed rating has default and migration probabilities which change over time. For this reason, in the thesis we consider the marginal default and migration probabilities in (3.2) as attributes of the individual credit and not of the rating category to which the credit belongs. This gives a sufficient flexibility to the model and allows for statistical estimation of the credit risk based on every kind of relevant information for the marginal credits that could be available.

4.1.2   Loss given rating

Knowing the credit quality (the rating) of a given obligor at the time horizon is not sufficient in itself to determine completely the profit/loss, which the bank will generate for holding the credit. In case of default, it is not clear whether the collateral of the credit will be sufficient to cover at least the expected part of the losses. In all other cases, the change in the market value of the credit may not be completely determined by the credit rating.

An important ingredient of any credit risk model is the *loss given default (LGD)*, i.e. the d.f. of L_j, given $X_j = 1$, for $j = 1, \ldots, m$ in (3.1). Gupton and Stein [75] and Hamilton et al. [76] provide comprehensive studies and point out the most powerful determinants of LGD. Their key findings are:

- The type of the obligor and the type of the credit are the most powerful determinants.

- The initial rating category of the obligor can also be used as a predictor: the higher the rating, the lower the LGD.

- The LGD moves with the business cycle and thus is positively correlated with the default probabilities: the higher the default probabilities, the higher the LGD.

These and some additional factors are summarized in Moody's Loss-Calc – the industrial model for prediction of LGD, see Gupton and Stein [75]. Further studies of historical LGD are Eberhart and Sweeny [44], Altman [3], Altman and Kishore [4] and many others.

Apart from the case of default, a lot of research has also been done in the development of rating-based forward pricing models for single-obligor credit risks (bonds and derivatives). A general pricing framework can be found in Lando [97]. Jarrow et al. [81] consider a very simple case leading to deterministic prices, given the rating. CreditMetrics [74] takes also a similar approach. More complicated, but more realistic, are the models developed for instance by Das and Tufano [28], and Li [103]. We do not go into detail since our goal is the portfolio view of credit risk. We adopt a general formulation. In this thesis the profit/loss of the individual credit, given its credit rating, is modelled by an arbitrary r.v. L_j, see model (3.1). Specifying the distribution of this r.v., for all possible rating categories at the risk horizon, and the corresponding default and migration probabilities as in Section 4.1.1 is sufficient to calibrate the marginal loss distribution in our model.

More problematic are assumptions (A) and (B) in model (3.1). Sytematic dependencies between the losses given rating are observed in various empirical studies, see e.g. Frye [63, 64]. In particular, for some credit types LGDs are positively correlated with certain market factors. The effect of this dependence on the portfolio level is studied by Oda and Muranaga [121] and by Frey and McNeil [61], Section 5.1 (systematic recovery risk). For the sake of simplicity, we disregard this dependence and keep assumptions (A) and (B).

4.2 Bivariate dependence measures

We focus on the dependence structure of a two-dimensional random vector (Y_1, Y_2) with continuous marginals. Throughout this section we assume that we are given a sample $(Y_1^{(i)}, Y_2^{(i)})$, $i = 1, \ldots, n$, of i.i.d. copies of (Y_1, Y_2).

In Section 4.2.1 we consider the estimation of some classical dependence measures for the **whole range** of Y_1 and Y_2. In Section 4.2.2 we consider the estimation of the tail dependence coefficient which measures the dependence **in the extremes** of Y_1 and Y_2.

4.2.1   Kendall's tau and correlation

Recall Definition 2.1.6 of the linear correlation coefficient. Given a sample $(Y_1^{(i)}, Y_2^{(i)})$, $i = 1, \ldots, n$, of i.i.d. copies of the random vector (Y_1, Y_2) with finite variances, it is straightforward to estimate the variance of Y_j, $j = 1, 2$, by

$$\widehat{\sigma}(Y_j) = \frac{1}{n-1} \sum_{i=1}^{n} (Y_j^{(i)} - \overline{Y}_{j,n})^2 \,,$$

where

$$\overline{Y}_{j,n} = \frac{1}{n} \sum_{i=1}^{n} Y^{(i)} \,.$$

In order to estimate the linear correlation coefficient, we may compute the sample covariance

$$\widehat{\mathrm{cov}}(Y_1, Y_2) = \frac{1}{n-1} \sum_{i=1}^{n} (Y_1^{(i)} - \overline{Y}_{1,n})(Y_2^{(i)} - \overline{Y}_{2,n}) \,,$$

and then use

$$\widehat{\rho}(Y_1, Y_2) = \frac{\widehat{\mathrm{cov}}(Y_1, Y_2)}{\sqrt{\widehat{\sigma}(Y_1)}\sqrt{\widehat{\sigma}(Y_2)}} \,. \tag{4.3}$$

This estimator is classical and works quite satisfactory in the case when (Y_1, Y_2) are bivariate normal. However, it has certain weak points. In particular, it depends crucially on the marginals of Y_1 and Y_2. When they are asymmetric or heavy-tailed, the estimator lacks robustness. For instance, when $\mathrm{var}(Y_1) = \infty$, it may not even converge. Recall that (2.14) provides a definition of the correlation coefficient in the framework of elliptical copulas, which does not rely on marginal assumptions, and in particular allows for marginal heavy-tails and infinite variances. Since model (3.5) is essentially based on elliptical copulas and does not exclude heavy-tailed marginals, we need an alternative method to estimate $\rho(Y_1, Y_2)$.

Various covariance and correlation estimators have been proposed including M-estimators, estimators based on multivariate trimming and estimators based on variances of sums and differences of standardized variables (see Devlin et al. [35] for an overview). Particularly useful in the framework of elliptical distributions are the methods suggested by Frahm and Junker [57] and by Lindskog et al. [104].

The estimator of Frahm and Junker [57] is in principle applicable to a random vector Y of arbitrary dimension $d \geq 2$ with elliptical distribution. It uses extensively the repesentation of such a random vector $Y = RAU + \mu$ as in (2.13). First one needs to estimate the mean vector $\mu \in \mathbb{R}^d$. This can be done easily by the sample mean vector $(\overline{Y}_{1,n}, \ldots, \overline{Y}_{d,n})$. Alternatively, under some technical conditions detailed in Bünning and Trenkel [18], Chapter 3, the sample median vector is also a consistent estimate for μ (recall that the marginals of the elliptical distributions are symmetric). Once we have an

estimate for μ, we substract it from the data and assume that we have $Y = RAU$ in (2.13). Denote by $|\cdot|$ the L^2 norm on $\mathbb{R}^d$. We have that

$$S := \frac{Y}{|Y|} \overset{d}{=} \frac{AU}{|AU|}.$$

The vector S is distributed on the unit hypersphere (since U is), and, given the matrix A, its density can be computed explicitly. In order to estimate A, one maximizes the likelihood function

$$\widehat{\Gamma} = \arg \max_{\Gamma} \log(L(A; S^{(1)}, \dots, S^{(n)})) = \arg \max_{\Gamma} n \log(\det(\Gamma'\Gamma)) - d \sum_{i=1}^{n} n S^{(i)} \Gamma' \Gamma S^{(i)},$$

where $S^{(1)}, \dots, S^{(n)}$ are the observed i.i.d. copies of S and $\Gamma = \delta A^{-1}$ with $\delta > 0$ being a design constant, see Frahm and Junker [57] for more details. Once the estimate $\widehat{\Gamma}$ is available, we get directly an estimate for the correlation coefficients as defined in (2.14) (recall that $\Sigma = A'A$ and that the correlation is invariant under scaling with a constant δ).

In contrast to the above method, the estimator suggested in Lindskog et al. [104] is applicable to the more general situation when Y is a random vector with arbitrary absolutely continuous marginals and elliptical copula. It is essentially based on Theorem 2.3.11 which shows that the elegant relationship

$$\rho(Y_1, Y_2) = \sin(\frac{\pi}{2} \tau(Y_1, Y_2)) \tag{4.4}$$

between the linear correlation coefficent defined in (2.14) and Kendall's tau defined in (2.1) holds when (Y_1, Y_2) have an elliptical distribution with absolutely continuous marginals. The result is not only of theoretical interest; it is also extremely useful for statistical purposes. It can be used to build a robust estimator of the linear correlation coefficient for data coming from a distribution with elliptical copula. Formula (4.4) provides an appealing bivariate method; we estimate the Kendall's tau using the standard textbook estimator and then plug it in relationship (4.4) to get the Kendall's tau transform estimate of ρ. More precisely, one may consistently estimate the Kendall's tau by

$$\widehat{\tau}(Y_1, Y_2) = \binom{n}{2}^{-1} \sum_{i>k}^{n} \text{sign}[(Y_1^{(k)} - Y_1^{(i)})(Y_2^{(k)} - Y_2^{(i)})],, \tag{4.5}$$

and then use

$$\widehat{\rho}(Y_1, Y_2) = \sin(\frac{\pi}{2} \widehat{\tau}(Y_1, Y_2)). \tag{4.6}$$

The estimator (4.5) is consistent and asymptotically normal, see e.g. Höffding [77], which implies the same for the correlation estimate (4.6) by the continuous mapping theorem and the delta method. Simulation studies in Lindskog et al. [104] suggest that this simple

method performs better than most of its competitors. Further results in Daul et al. [30] show that (4.4) provides also a quite accurate approximation of the correlation coefficient in more general situations than an elliptical copula, hence (4.6) is also quite robust. We present some examples in Section 4.5, which investigate the accuracy of the estimator for modest sample size.

Note that, unlike other methods of correlation estimation, the Kendall's tau transform method directly exploits the geometry of the elliptical distributions and does not require to estimate variances and covariances. This is advantageous when the interest focusses explicitly on correlations, as it often does in financial applications, and in particular in our credit risk model. Recall that by means of Proposition 3.1.1 the copula of the assets Y determines the distribution of the ratings X (which is of central interest in our work) regardless of the assets' marginals, and that the correlations $\rho(Y_i, Y_j)$, $i, j = 1, \ldots, m$, are essential parameters of this copula.

More generally, (4.6) can be used to calibrate the correlation matrices of higher di- mensional elliptical copulas. However, in some cases the matrix of the pairwise correlation estimates has to be adjusted numerically to ensure that the resulting matrix is positive definite; see Rousseeuw and Molenberghs [125] for details. Note that the factor model (3.5) guarantees that the (high-dimensional) correlation matrix $\Sigma_Y \in \mathbb{R}^{m \times m}$ of the assets is positive definite as long as (the lower-dimensional) correlation matrix of the common factors $\Sigma \in \mathbb{R}^{p \times p}$ is. This reduces the problem significantly, and in most practical cases no or just minor adjustments of the correlation matrices are necessary (see Schwarz [131]).

4.2.2   Tail dependence

Recall Definition 2.1.12 of the tail dependence coefficient. In this section, we consider sev- eral methods for estimating this dependence measure, based on a sample $(Y_1^{(i)}, Y_2^{(i)})$, $i = 1, \ldots, n$, of i.i.d copies of the random vector (Y_1, Y_2). We focus on the lower tail dependence coefficient; the same methods can be applied to the upper tail dependence, by considering the vector $(-Y_1, -Y_2)$. We assume also that (Y_1, Y_2) are tail-dependent. Statistical meth- ods for testing of tail dependence or tail independence go beyond the scope of this work, for that we refer the reader to Ledford and Tawn [99] and Draisma et al. [36].

The main difficulty in estimating tail dependence arises in the limited availability of extreme data. According to the assumptions on the distribution of the observed random vector, the methods for estimating the tail dependence can be classified as parametric, semi-parametric and non-parametric methods. According to the usage of data, the ap- proaches are ascribed either to the entire dataset or to a subset of extreme data.

Within the framework of the methods based on extreme data, the extreme value theory introduced in Section 2.2.1 contains the natural tools for inferences on the tail behavior

of (one-dimensional) probability distributions. In the one-dimensional setting the class of extreme value distributions has a finite parametrization, so it suffices to apply parametric estimation methods. By contrast, no finite parametrization of the multidimensional extreme-value distributions exists. This leads to more complicated estimation methods.

A key result in this section is a new non-parametric estimator of the tail dependence coefficient which utilizes only the extreme observations. We start with some classical methods.

(A) **Parametric tail dependence estimators based on the whole available sample.**
In this class of methods, the d.f. of the vector (Y_1, Y_2) is assumed to belong to a certain parametric family. Classical methods like maximum likelihood (ML) provide estimators for the parameters of the entire d.f. An estimator of the tail dependence coefficient in this case follows as a by-product. Under the usual regularity conditions for the ML theory (see e.g. Casella and Berger [21], p.516), this tail dependence estimator can be shown to be consistent and asymptotically normal. However, recall from Proposition 3.1.1 that in our credit risk application we are interested in the copula of a random vector, rather than in its full d.f. Since the tail dependence coefficient itself is essentially a copula property, it would be more natural to estimate it without making assumptions on the marginals. The next method (B) is a step in this directlion.

(B) **Semi-parametric tail dependence estimators based on the whole available sample.**
In this class of methods, the copula of the vector (Y_1, Y_2) is assumed to belong to a certain parametric family, and the marginals are left arbitrary (but continuous, so that the copula of the random vector is uniquely determined, see Theorem 2.1.2). In this case a two-step method can be applied. In the first step, one has to transform the marginals Y_1 and Y_2 into uniform r.v.s, using one of the methods described in Section 4.3.1. In the second step, one estimates the parameters of the copula, using for instance the ML methods described in Section 4.3.2. As in method (A), an estimator of the tail dependence follows as a by-product from the parameric copula assumptions. Genest et al. [65] show that such estimators are consistent and asymptotically normal.

The main drawback of the above approaches (A) and (B) is that they infer the tail dependence coefficient, which is a measure for the dependence in the extremes, without putting more weight on the extreme observations. The methods that follow focus on this issue.

(C) **Tail dependence estimator for elliptically distributed vector, which utilizes the extreme observations.**
Recall from Proposition 2.3.12 that the tail dependence coefficient of a bivariate elliptical

random vector with represenation $Y = RAU + \mu$ as in (2.13) is given by:

$$\lambda(Y_1, Y_2) = \frac{\int_{g(\rho_{1,2})}^{\frac{\pi}{2}} \cos^\alpha t\, dt}{\int_0^{\frac{\pi}{2}} \cos^\alpha t\, dt}\,,$$

where $g(t) = \frac{\pi}{4} - \frac{\arcsin t}{2}$, $t \in (-1, 1)$ and α is the tail index of the regularly varying r.v. R. Therefore, it is sufficient to estimate the correlation coefficient $\rho_{1,2}$ and the tail index α of the spectral variable R. The estimation of the correlation $\rho_{1,2}$ can be done by one of the methods applicable to heavy-tailed elliptical distributions described in Section 4.2.1. For the estimation of the tail index α we note that by Proposition 2.2.15 the r.v. $|Y|$ ($|\cdot|$ being an arbitrary norm) is regularly varying with the same tail index as the random vector Y. Therefore it is sufficient to apply standard one-dimensional extreme value theory to estimate the tail index of $|Y|$. Klüppelberg et al. [91] apply the Hill estimator (2.8) and prove consistency and asymptotic normality of the resulting tail dependence estimator. Alternatively, Frahm et al. [58] apply the POT method (end of Section 2.2.1). In both cases $|\cdot|$ is taken to be the L^2-norm, althought theoretically this is not a significant requirement.

An advantage of the approach (C) is that it utilizes the extreme observations by applying the Hill estimator or the POT method in the estimation of the tail index α of the elliptical random vector. However, there are certain drawbacks which are similar to the drawback of approach (A). The marginal tails in fact play a crucial role in the estimation of the tail index α, and hence in the estimation of the tail dependence. The method cannot be applied to distributions with elliptical copula and arbitrary marginals, which are of central interest in our credit risk model. In many practical applications one can find that whereas the elliptical copula is a reasonable model for the dependence structure of the financial assets under consideration, they exhibit a different marginal behaviour, and in particular do not have heavy tails, see Example 4.5.5 or Schwarz [131]. In such cases method (C) breaks down.

Parametric estimation methods have the advantage of being efficient given that the model is true whereas nonparametric estimation avoids model-misspecification. The next methods are non-parametric, and are based on the extreme observations. We start with a method based on the assumption that the random vector under consideration is a general regularly varying vector.

(D) **Tail dependence estimator for regularly varying random vectors.**
Let (Y_1, Y_2) be a regularly varying random vector with tail index α and continuous marginals G_1 and G_2 resp. Recall the estimator of the spectral measure μ as in (2.12). Of particular interest in the context of tail dependence are probabilities like

$$\pi(p, q) = P(Y_1 > G_1^{-1}(1 - p) \,|\, Y_2 > G_2^{-1}(1 - q))\,,$$

where p and q are some small probabilities. The key idea is to use the scaling property of the Radon measure μ as in Proposition 2.2.16. Making use of (2.11), Wendin [135] suggests to approximate the spill-over probability $\pi(p, q)$, for p, q being small, by

$$\widetilde{\pi}(p, q) = \frac{\mu(\{(s,t) \,:\, s > (q/\sqrt{q^2 + p^2})^{1/\alpha} \,,\, t > (p/\sqrt{q^2 + p^2})^{1/\alpha}\})}{\mu(\{(s,t) \,:\, s > (q/\sqrt{q^2 + p^2})^{1/\alpha}\})} \,. \tag{4.7}$$

Invoking the estimator $\widehat{\mu}$ as in (2.12) one obtains the corresponding estimator $\widehat{\pi}(p, q)$ for $\pi(p, q)$, i.e.

$$\widehat{\pi}(p, q) = \frac{\widehat{\mu}_n(\{(s,t) \,:\, s > (q/\sqrt{q^2 + p^2})^{1/\alpha} \,,\, t > (p/\sqrt{q^2 + p^2})^{1/\alpha}\})}{\widehat{\mu}_n(\{(s,t) \,:\, s > (q/\sqrt{q^2 + p^2})^{1/\alpha}\})}$$

Setting $p = q$ in (4.7) and making use of the scaling property of the Radon measure μ as in Proposition 2.2.16 we obtain an estimator for the upper tail dependence coefficient

$$\widehat{\lambda}_\mu = \frac{\widehat{\mu}_n(\{(s,t) \,:\, s > 1/\sqrt{2} \,,\, t > 1/\sqrt{2}\})}{\widehat{\mu}_n(\{(s,t) \,:\, s > 1/\sqrt{2}\})} \tag{4.8}$$

where $\widehat{\mu}_n$ is the estimator of the Radon measure given in (2.12). Note that, unlike the estimator presented in (C), (4.8) does not depend on the tail index α. However, it is still based on the assumption that the marginal distributions are regularly varying with the same tail index α as the vector, see e.g. the conlcusion in Wendin [135]. Similarly to the method (C), the method (D) is not applicable to the more general case of a tail dependent random vector with arbitrary marginals.

The next methods which we consider are completely non-parametric, and in particular nothing more than continuity is assumed for the marginals.

(E) **Empirical tail dependence estimator.**

Let $Y = (Y_1, Y_2)$ be a bivariate random vector with continuous marginals, copula C and tail dependence $\lambda(Y_1, Y_2)$. Let $Y^{(i)}$, $i = 1, \ldots, n$, be i.i.d. copies of Y. Theorem 2.1.2 suggests an estimate of $C(p, p)$, for $p \in (0, 1)$:

$$C^E(p, p) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{F_1^E(Y_1^{(i)}) < p, \, F_2^E(Y_2^{(i)}) < p\}} \,, \tag{4.9}$$

where $F_j^E(p)$ denotes the empirical d.f. of Y_j, for $j = 1, 2$. Invoking an auxiliary sequence $k = k(n)$ with the properties $k \to \infty$ and $\frac{k}{n} \to 0$ as $n \to \infty$ we obtain

$$\widehat{\lambda}^{n,k} = \frac{n}{k} C^E(\frac{k}{n}, \frac{k}{n}) \xrightarrow{P} \lambda(Y_1, Y_2), \quad n \to \infty \,, \tag{4.10}$$

see Schmidt and Stadtmüller [127] for a proof of the above fact and for asymptotic normality of the suggested estimator.

The main drawback of the above approach is in the high variance of the simple empirical estimation of rare event probabilities. The next method is expected to perform better.

(F) **Log-transform method.**

Let Y be a tail-dependent bivariate random vector with continuous marginals and a differentiable copula C. Then for its tail dependence coefficient we have

$$\lambda(Y_1, Y_2) = \lim_{u \to 0} \frac{C(u, u)}{u} = 2 - \lim_{u \to 0} \frac{\log(1 - 2u - C(u, u))}{\log(1 - u)}.$$

Recall that $1 - 2u + C(u, u) = C(1 - u, 1 - u)$ by definition. Invoking the empirical copula estimator (4.9) and an auxiliary sequence $k = k(n)$ with the properties $k \to \infty$ and $\frac{k}{n} \to 0$ as $n \to \infty$ we obtain

$$2 - \frac{\log C^E(1 - \frac{k}{n}, 1 - \frac{k}{n})}{\log(1 - \frac{k}{n})} \xrightarrow{P} \lambda(Y_1, Y_2), \quad n \to \infty, \tag{4.11}$$

see Coles et al. [22] for a proof and for asymptotic properties of the estimator.

We conclude with a new tail dependence estimator, which combines ideas from methods (D), (E) and (F).

(G) **Alternative tail dependence estimator.**

Let Y be a tail-dependent bivariate random vector with continuous marginals, differentiable copula C and tail dependence coefficient $\lambda(Y_1, Y_2)$. Let $Y^{(i)}$, $i = 1 \ldots, n$, be i.i.d. copies of Y. We start by noting that the weakest point of the tail dependence estimator (4.10) is that it is a simple empirical estimator based on the extreme observations. More precisely, denote

$$U_j^{(i)} = F_j^E(Y_j^{(i)}), \ \ j = 1, 2, \ i = 1, \ldots, n, \tag{4.12}$$

where F_j^E is the empirical d.f. of Y_j. Then the estimator in (E) can be rewritten as

$$\widehat{\lambda}^{n, u} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{u} 1_{\{U_1^{(i)} < u, U_2^{(i)} < u\}},$$

where $u \in (0, 1)$ is some small threshold. We consider the extremes in a different set and use a weighted empirical estimator with weights proportional to the distance to the diagonal $U_1 = U_2$. To this end, we transform further $(U_1^{(i)}, U_2^{(i)})$, $i = 1, \ldots, n$, into polar coordinates

$$U_1^{(i)} = Q^{(i)} \sin \phi^{(i)}, \ U_2^{(i)} = Q^{(i)} \cos \phi^{(i)}, \ \ i = 1, \ldots, n, \tag{4.13}$$

where the r.v.s $(Q^{(i)}, \phi^{(i)})$ satisfy $0 \leq Q^{(i)} \leq \sqrt{2}$ and $0 \leq \phi^{(i)} \leq \frac{\pi}{2}$. Then we select a small $r \in (0, \sqrt{2})$ and we suggest the following estimator

$$\widetilde{\lambda}^{n, r} = \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{2}}{r} 1_{\{Q^{(i)} < r\}} \sin(2\phi^{(i)}). \tag{4.14}$$

In the following proposition we prove consistency of the suggested estimator.

Proposition 4.2.1. *For $n \in \mathbb{N}$ let $Y^{(i)} = (Y_1^{(i)}, Y_2^{(i)})$, $i = 1, \ldots, n$, be a sequence of i.i.d random vectors with differentiable copula C and arbitrary continuous marginals F_1, F_2. Let also $(Q^{(i)}, \phi^{(i)})$, $i = 1, \ldots, n$, be the r.v.s in (4.13). Let $k = k(n)$ be a sequence of positive constants such that $k(n) \to \infty$ and $\frac{k(n)}{n} \to 0$, $n \to \infty$. Then*

$$\widetilde{\lambda}^{n,k/n} \xrightarrow{P} \lambda^*, \ n \to \infty, \tag{4.15}$$

where λ^* is the tail dependence coefficient of (Y_1, Y_2).

Proof. Denote by

$$\overline{U}_j^{(i)} \ = \ F_j(Y_j^{(k)}), \ j = 1, 2, \ i = 1, \ldots, n,$$

where F_j is the true d.f. of Y_j, and the corresponding $(\overline{Q}^{(i)}, \overline{\phi}^{(i)})_{i=1,\ldots,n}$ as in (4.13) and note that this is an i.i.d sequence, since $(Y_1^{(i)}, Y_2^{(i)})_{i=1,\ldots,n}$ is an i.i.d. sequence. Denote also by $\overline{\lambda}^{n,k/n}$ the corresponding r.v. as in (4.14). Since F_j, $j = 1, 2$, is a monotone function, we have for the copula $C_{\overline{U}_1^{(i)}, \overline{U}_2^{(i)}} = C$, $i = 1, \ldots, n$. Fix $i \in \{1, \ldots, n\}$ and denote $Q = \overline{Q}^{(i)}$, $\phi = \overline{\phi}^{(i)}$, $U_j = \overline{U}_j^{(i)}$, $j = 1, 2$. We have

$$
\begin{aligned}
1_{\{Q<r\}} \sin(2\phi) \ &= \ 1_{\{Q<r\}} 2 \sin\phi \cos\phi \\
&= \ 1_{\{Q<r\}} \frac{2(Q \sin\phi)(Q \cos\phi)}{Q^2} \\
&\overset{d}{=} \ 1_{\{\sqrt{U_1^2 + U_2^2} < r\}} \frac{2 U_1 U_2}{U_1^2 + U_2^2}.
\end{aligned}
$$

Therefore

$$E[1_{\{Q<r\}} \sin(2\phi)] \ = \ \int\int_{D=\{\sqrt{u_1^2 + u_2^2} < r\}} \frac{2u_1 u_2}{u_1^2 + u_2^2} dC(u_1, u_2).$$

Changing variables by $z = \sqrt{u_1^2 + u_2^2} \in (0, r)$, $t = \sqrt{2u_1 u_2} \in (0, z)$ we obtain

$$E[1_{\{Q<r\}} \sin(2\phi)] = \int_0^r \int_0^z \frac{t^2}{z^2} dC \left(\frac{\sqrt{z^2 + t^2} + \sqrt{z^2 - t^2}}{2}, \frac{\sqrt{z^2 + t^2} - \sqrt{z^2 - t^2}}{2} \right)$$

Therefore, applying L'Hopital's rule

$$\lim_{r\to 0} \frac{\sqrt{2}}{r} E[1_{\{Q<r\}} \sin(2\phi)] = \lim_{r\to 0} \frac{\partial}{\partial u_1} C(u_1, u_2)_{|u_1=u_2=r} + \lim_{r\to 0} \frac{\partial}{\partial u_2} C(u_1, u_2)_{|u_1=u_2=r}. \tag{4.16}$$

However, by definition,

$$\lambda^* = \lim_{u\to 0} \frac{C(u, u)}{u} = \lim_{r\to 0} \frac{\partial}{\partial u_1} C(u_1, u_2)_{|u_1=u_2=r} + \lim_{r\to 0} \frac{\partial}{\partial u_2} C(u_1, u_2)_{|u_1=u_2=r}.$$

Then (4.15) follows from Chebishev's inequality. More precisely, we have for every $\epsilon > 0$,

$$P\left(\left|\widetilde{\lambda}^{n,k/n} - \lambda^*\right| > \epsilon\right) \leq \frac{1}{\epsilon}\left(E\left|\widetilde{\lambda}^{n,k/n} - \overline{\lambda}^{n,k/n}\right| + E\left|\overline{\lambda}^{n,k/n} - E\overline{\lambda}^{n,k/n}\right| + \left|E\overline{\lambda}^{n,k/n} - \lambda^*\right|\right).$$

From (4.16) we have that $E\overline{\lambda}^{n,k/n} \to \lambda^*$, $n \to \infty$.
By the SLLN we have $\overline{\lambda}^{n,k/n} \stackrel{\text{a.s.}}{\to} E\overline{\lambda}^{n,k/n}$, $n \to \infty$.
Finally, since $F_j^E(Y^{(i)}) \stackrel{\text{a.s.}}{\to} F_j(Y^{(i)})$ we have also

$$(Q^{(i)}, \phi^{(i)}) \stackrel{\text{a.s.}}{\to} (\overline{Q}^{(i)}, \overline{\phi}^{(i)}), \ i = 1, \ldots, n. \tag{4.17}$$

Furthermore

$$\begin{aligned}
E\left|\widetilde{\lambda}^{n,k/n} - \overline{\lambda}^{n,k/n}\right| &\leq \frac{\sqrt{2}}{k/n} E\left|1_{\{Q^{(i)}<r\}}\sin(2\phi^{(i)}) - 1_{\{\overline{Q}^{(i)}<r\}}\sin(2\overline{\phi}^{(i)})\right| \\
&\leq \frac{\sqrt{2}}{k/n} E\left|1_{\{Q^{(i)}<r\}} - 1_{\{\overline{Q}^{(i)}<r\}}\right| \\
&\leq \frac{\sqrt{2}}{k/n}\left(P(Q^{(i)} < \frac{k}{n})P(Q^{(i)} - \overline{Q}^{(i)} < 0) + P(\overline{Q}^{(i)} < \frac{k}{n})P(Q^{(i)} - \overline{Q}^{(i)} > 0)\right)
\end{aligned}$$

However, similarly as above,

$$P(\overline{Q}^{(i)} < r) = \int_0^r \int_0^z dC\left(\frac{\sqrt{z^2 + t^2} + \sqrt{z^2 - t^2}}{2}, \frac{\sqrt{z^2 + t^2} - \sqrt{z^2 - t^2}}{2}\right)$$

and, applying L'Hopital's rule, $\frac{\sqrt{2}}{r}P(\overline{Q}^{(i)} < r) \to \lambda^* < \infty$ as $r \to 0$. By (4.17) we obtain also $\frac{\sqrt{2}}{r}P(Q^{(i)} < r) \to \lambda^* < \infty$ and $P(|Q^{(i)} - \overline{Q}^{(i)}| \neq 0) \to 0$ as $n \to \infty$. This leads to the required result. □

Note that the estimators in (E), (F) and (G) may not necessarily be applied by using the empirical d.f. to transform the data to the copula scale as suggested in (4.12). One may potentially replace the empirical d.f. F_j^E, $j = 1, 2$, in that equation by another consistent estimate, see Section 4.3.1. However, this will presumably have a small impact on the estimators in (E) and (D), as they are based essentially on a direct approximation of the copula function at a single point, i.e. $C(k/n, k/n)$ is approximated by e.g. $C^E(k/n, k/n)$ in (4.10) and $C(1 - k/n, 1 - k/n)$ by e.g. $C^E(1 - k/n, 1 - k/n)$ in (4.11). In contrast to that, the new estimator (4.14) is intended to capture more information from the joint extreme observation by assigning different weights to each one of them. Consequently, replacing the empirical d.f. in (4.12) by another consistent estimate of the marginal d.f. (which is advisable e.g. when we are dealing with marginal heavy tails) will have a direct impact on the weights and thus on the tail dependence estimate.

In the next proposition we show asymptotic normality of the suggested estimator (4.14), provided the marginal d.f.s are known. Since (4.14) is just a weighted estimator,

the asymptotic variance is the same as the variance of the empirical estimator (E), see
Schmidt and Stadtmüller [127], Theorem 5.

For $n \in \mathbb{N}$ let $Y^{(i)} = (Y_1^{(i)}, Y_2^{(i)})$, $i = 1, \ldots, n$, be a sequence of i.i.d random vectors with
differentiable copula C, tail dependence coefficient λ^* and arbitrary continuous marginals
F_1, F_2. Assume that there exists a functon $A(p) : \mathbb{R}^+ \to \mathbb{R}^+$ such that $A(p) \to 0$, $p \to 0$
and

$$\frac{C(p,p)/p - \lambda^*}{A(p)} = g < \infty, \quad p \to 0. \tag{4.18}$$

Denote by

$$\overline{U}_j^{(i)} = F_j(Y_j^{(k)}), \; j = 1, 2, \; i = 1, \ldots, n,$$

and the corresponding $(\overline{Q}^{(i)}, \overline{\phi}^{(i)})_{i=1,\ldots,n}$ as in (4.13). Denote also by $\overline{\lambda}^{n,k/n}$ the correspond-
ing r.v. as in (4.14).

Proposition 4.2.2. *With the above notations, let $k = k(n)$ be a sequence of positive
constants such that $k(n) \to \infty$ and $\frac{k(n)}{n} \to 0$, $n \to \infty$. If*

$$\sqrt{k} A(k/n) \to 0, \quad n \to \infty,$$

where A is the function in (4.18), then

$$\sqrt{k}(\overline{\lambda}^{n,k/n} - \lambda^*) \xrightarrow{d} N(0, \sqrt{2}\lambda^*), \quad n \to \infty.$$

Proof. We have

$$\sqrt{k}(\overline{\lambda}^{n,k/n} - \lambda^*) = \sqrt{k}(\overline{\lambda}^{n,k/n} - E[\overline{\lambda}^{n,k/n}]) + \sqrt{k}(E[\overline{\lambda}^{n,k/n}] - \lambda^*).$$

By (4.16) and (4.18) we get

$$\lim_{n\to\infty} \sqrt{k}(E[\overline{\lambda}^{n,k/n}] - \lambda^*) = \lim_{n\to\infty} \sqrt{k} g A(k/n) = 0.$$

Furthermore,

$$\sqrt{k}(\overline{\lambda}^{n,k/n} - E[\overline{\lambda}^{n,k/n}]) = \sum_{i=1}^{n} Z_{k,n},$$

where $Z_{k,n}^{(i)} = \frac{\sqrt{2}}{\sqrt{k}}(1_{\{\overline{Q}^{(i)} < k/n\}} \sin(2\overline{\phi}^{(i)}) - E[1_{\{\overline{Q}^{(i)} < k/n\}} \sin(2\overline{\phi}^{(i)})])$. By definition, $Z_{k,n}^{(i)}$ are
i.i.d. with mean $E[Z_{k,n}^{(i)}] = 0$, $i = 1, \ldots, n$. Denoting by $Z_{k,n} = \frac{\sqrt{2}}{\sqrt{k}}(1_{\{\overline{Q} < k/n\}} \sin(2\overline{\phi}) - E[1_{\{\overline{Q} < k/n\}} \sin(2\overline{\phi})])$ a generic r.v. with the same distribution as $Z_{k,n}^{(1)}$, we have

$$\text{var}(\sum_{i=1}^{n} Z_{k,n}^{(i)}) = nE[Z_{k,n}^2]. \tag{4.19}$$

Then we compute

$$nE[Z_{k,n}^2] = \sqrt{2}\frac{n}{k}E[\sqrt{2}1_{\{\overline{Q}<k/n\}}\sin^2(2\overline{\phi})] - \frac{k}{n}\left(\frac{n}{k}E[\sqrt{2}1_{\{\overline{Q}<k/n\}}\sin^2(2\overline{\phi})]\right)^2 .$$

The second term converges to 0 as n tends to infinity because of (4.16):

$$\lim_{n\to\infty}\frac{k}{n}\left(\frac{n}{k}E[\sqrt{2}1_{\{\overline{Q}<k/n\}}\sin^2(2\overline{\phi})]\right)^2 = \lim_{n\to\infty}\frac{k}{n}(\lambda^*)^2 = 0 .$$

Using the same change of variable as in the proof of Proposition 4.2.1 we get for the first part

$$\sqrt{2}\frac{n}{k}E[\sqrt{2}1_{\{\overline{Q}<k/n\}}\sin^2(2\overline{\phi})] = \sqrt{2}\frac{n}{k}\int_0^{\frac{k}{n}}\int_0^z \sqrt{2}\frac{t^4}{z^4}dC\left(\frac{\sqrt{z^2+t^2}+\sqrt{z^2-t^2}}{2}, \frac{\sqrt{z^2+t^2}-\sqrt{z^2-t^2}}{2}\right) .$$

Applying L'Hopital's rule and going back to (4.19) we obtain

$$\text{var}(\sum_{i=1}^n Z_{k,n}^{(i)}) \to \sqrt{2}\lambda^* .$$

Finally we note that $|Z_{k,n}^{(i)}| \le \sqrt{\frac{2}{k}}$ a.s. and therefore, for every $i = 1,\ldots,n$,

$$P(|Z_{k,n}^{(i)}| > \epsilon) \to 0, \quad n \to \infty$$

for every $\epsilon > 0$. Then the required result then follows from Theorem 4.1. in Petrov [123].

$\square$

We leave the question for asymptotic normality in the more general case of unknown marginal distribution for future research. It is indeed a complicated problem related to the rate of convergence of multivariate empirical tail processes, see e.g. Deheuvels [33].

4.3 Elliptical copulae

The problem under consideration in this section can be formulated as follows: given a sample $(I_1^{(i)},\ldots,I_d^{(i)})$, $i = 1,\ldots,n$, of i.i.d. copies of the random vector I with elliptical copula C and continuous marginals $F_1,\ldots,F_d$, estimate C regardless of $F_1,\ldots,F_d$. In order to achive our goal, we start with several methods for transforming the observable data into the $[0,1]^d$ copula scale. We continue with some classical or more recent methods for copula estimation, based on the ML theory. Finally, we suggest a new non-parametric method for elliptical copula calibration, which makes extensive use of the information contained in the joint extreme observations.

4.3.1 Transforming the marginals

When the estimation of a parametric copula is the primary objective, the unknown marginal distributions of the data enter the problem as nuisance parameters. The first step is usually to estimate the unknown marginals and then to use the probability-integral transform to get data on the copula scale $[0, 1]^d$.

Broadly speaking the marginal modelling can be done in three ways: fitting parametric distributions to each marginal; modelling the marginals non-parametrically using the empirical d.f.; and using a hybrid of the parametric and non-parametric methods.

The first method has been termed as the IFM or inference-functions-for-margins method by Joe [83] following the terminology used by McLeish and Small [113]. Asymptotic theory has been worked out for this approach by Joe [83]. The cruicial assumption is that each of the marginals $F_1, \ldots, F_d$ belongs to some parametric family with densities resp. $f_1(x, \psi_1), \ldots, f_d(x, \psi_d)$ (e.g. the normal densities with ψ_l being the unknown mean and variance of margin l, $l = 1, \ldots, d$). The unknown parameters $\psi_1, \ldots, \psi_d$ are estimated, for each marginal separately, by standard maximum likelihood (ML), i.e.

$$\widehat{\psi}_l = \arg \max_{\psi} \{ \sum_{i=1}^{n} \log(f(I_l^{(i)}, \psi)) \}, \, l = 1, \ldots, d \, .$$

However, in practice the success of the method is obviously dependent upon finding appropriate parametric models for the marginals, which may not always be so straightforward when these show evidence of heavy tails and/or skewness.

The second method involving estimation of the marginals by the empirical d.f. has been termed as the pseudo-likelihood method and is extensively investigated by Genest et al. [65]. The method involves estimating the l-th marginal d.f. F_l by

$$F_l^E(x) = \frac{1}{n+1} \sum_{i=1}^{n} 1_{\{I_l^{(i)} \leq x\}}, \, l = 1, \ldots, d \, .$$

The pseudo-sample from the copula is then constructed by

$$U_j^{(i)} = F_l^E(I_l^{(i)}), \, l = 1, \ldots, d, \, i = 1, \ldots, n \, .$$

Observe that, even if the original data vectors $(I_1^{(i)}, \ldots, I_d^{(i)})$, $i = 1, \ldots, n$, are i.i.d, the pseudo-sample data are dependent, because the marginal estimates F_l^E are constructed from all of the original data vectors through the univariate samples. Note also that division by n+1 in (4.3.1) keeps transformed points away from the boundary of the unit cube.

A hybrid of the parametric and nonparametric methods could be developed by modelling the tails of the marginal distributions using the GPD (see Definition 2.2.7) as suggested by extreme value theory and approximating the body of the distribution using the empirical d.f. (4.3.1). The method is known as the POT method and is used for

instance by Davison and Smith [31]. Recall from Section 2.2.1 that for each marginal F_l, $l = 1, \ldots d$, the POT method starts by selecting a high threshold u_l. Then, for $x > u_l$ we have

$$P(I_l \geq x) = P(I_l \geq u_l)P(I_l \geq x \,|\, I_l \geq u_l)\,.$$

This leads to the approximation

$$\widehat{F_l}(x) = F_l^E(x)\,, \;\; x \leq u_l\,,$$

$$\widehat{F_l}(x) = 1 - (1 - F_l^E(u_l))(1 - G_{\xi,\sigma}(x - u_l))\,, \;\; x \geq u_l\,,$$

where $G_{\xi,\sigma}$ is the GPD with parameters ξ, σ estimated as explained in Section 2.2.1. As usual, the pseudo-sample from the copula is then constructed by

$$U_j^{(k)} = \widehat{F}_j(I_j^{(k)})\,, \;\; j = 1, \ldots, d, \; k = 1, \ldots, n\,.$$

This method is particularly appealing when one or more of the marginals show evidence of heavy-tails.

4.3.2   Methods based on ML

Recall from Corrolary 2.3.10 that the elliptical copula of a random vector with representation $Y = RAU + \mu$ as in (2.13) is uniquely determined by the d.f. of the spectral variable R and the correlation matrix $[\rho_{lj}]_{l,j=1,\ldots,d}$ defined in (2.14). These are the parameters which we are interested in.

Throughout this section, we assume that we have applied one of the methods for dealing with the marginals discussed in Section 4.3.1 to the sample $(I_1^{(i)}, \ldots, I_d^{(i)})$, $i = 1, \ldots, n$, of i.i.d copies of the random vector I with elliptical copula and continuous marginals. This enables us to produce the copula "sample" by

$$\widehat{U}_l^{(i)} = \widehat{F}_l(I_l^{(i)})\,, \;\; l = 1, \ldots, d, \; i = 1, \ldots, n\,,$$

where $\widehat{F}_l$, $l = 1, \ldots, d$, is the estimated marginal d.f., for instance the empirical d.f.

Among the numerous possible statistical procedures for elliptical distributions in general, only few are designed to work on elliptical copulas regardless of the marginals (see Demarta and McNeil [34]). One of the possible methods is by means of pseudo - maximum likelihood (see Genest et al. [65]). Assume also that the d.f. of the spectral r.v. R as in (2.13) belongs to some parametric family with parameters Ψ, i.e. $P(R < x) = F(x; \Psi)$. Usually one takes the F-family with (d, ν) degrees of freedom, leading to the t-copula, see Example 3.2.1. Then one may estimate the correlation coefficients $[\rho_{lj}]_{l,j=1,\ldots,d}$ and the parameters Ψ by maximizing the pseudo-log-likelihood

$$l(\Psi, [\rho_{lj}]_{l,j=1,\ldots,d}, \widehat{U}) = \sum_{i=1}^{n} l(\Psi, [\rho_{lj}]_{l,j=1,\ldots,d}, \widehat{U}_1^{(i)}, \ldots, \widehat{U}_d^{(i)})\,,$$

where l is the log-likelihood function of the elliptical copula.

As typically the likelihood function is available in terms of d-dimensional integrals (see Demarta and McNeil [34]), in practice some numerical issues arise, in particular when $d > 2$. A different approach overcoming this problem is taken in Lindskog et al. [104]. It is based on the Kendall's tau. More precisely, using (4.6) one may estimate completely non-parametrically the correlation matrix by $[\widehat{\rho}_{lj}]_{l,j=1,\dots,d}$. Then, in order to estimate the remaining parameters Ψ in the distribution of the spectral variable R, one maximizes again the pseudo-log-likelihood

$$l(\Psi, [\widehat{\rho}_{lj}]_{l,j=1,\dots,d}, \widehat{U}) = \sum_{k=1}^{n} l(\Psi, [\widehat{\rho}_{lj}]_{l,j=1,\dots,d}, \widehat{U}_1^{(i)}, \dots, \widehat{U}_d^{(i)}) \,. \tag{4.20}$$

For an application to high-dimensional datasets see Daul et al. [30].

4.3.3 Methods based on tail dependence

As already discussed, the main reason why we are interested in copulas different from the Gaussian is that we need better models for the dependence between extreme events. In this sense both of the statistical approaches from Section 4.3.2 have the drawback that they infer the parameters using the whole sample of observations, and in particular do not put more weight on the joint extremes. Thus, they provide a good fit on the empirical copula as a whole, but they might be misleading when it comes to joint extreme events. In this section, we suggest a new method for calibration of an elliptical copula which uses extensively the information contained in the joint extreme observations and the methods for estimation of the tail dependence coefficients discussed in Section 4.2.2.

Let $I = (I_1, \dots, I_d)$ be a random vector with absolutely continuous marginals with support on the whole of $\mathbb{R}$ and an elliptical copula equal to the copula of the random vector $Y = RAU$. Assume that the spectral r.v. R is regularly varying with tail index $0 < \alpha^* < \infty$.

Let S be the linear space of $d \times d$ matrices with real components and $||\cdot||$ be the L^2 distance defined on S, i.e.

$$||A|| = ||[A_{lj}]|| = \sum_{l,j=1}^{d} A_{lj}^2, \ A \in S \,.$$

Denote $\widehat{\Lambda}^n = [\widehat{\lambda}_{lj}^n] \in S$, where $\widehat{\lambda}_{lj}^n, \ l,j = 1,\dots,d, \ n \in \mathbb{N}$, is a sequence of (weakly) consistent estimates of the pairwise tail-dependence coefficients λ_{lj}^* of I_l and I_j, i.e.

$$\widehat{\lambda}_{lj}^n \xrightarrow{P} \lambda_{lj}^*, \ n \to \infty \,. \tag{4.21}$$

All methods for tail dependence estimation from Section 4.2.2 satisfy this property; however, we are especially interested in the non-parametric methods (E) and (G), since we do not want to make any additional assumptions on the marginals. Denote also the true tail dependence coefficients matrix $\Lambda^* = [\lambda_{lj}^*] \in S$.

Denote $\widehat{\tau}^n = [\widehat{\tau}_{lj}^n] \in S$, where $\widehat{\tau}_{lj}^n$, $l, j = 1, \ldots, d$, $n \in \mathbb{N}$, is a sequence of consistent estimates of the Kendall's tau coefficients τ_{lj}^* of I_l and I_j, i.e.

$$\widehat{\tau}_{lj}^n \xrightarrow{P} \tau_{lj}^*, \ n \to \infty. \tag{4.22}$$

For example, (4.5) provides such sequence, see Lindskog et al. [104]. Denote also the true Kendall's tau matrix by $\tau^* = [\tau_{lj}^*] \in S$.

Furthermore, for $\alpha > 0$ and $\tau \in S$ with $\tau_{lj} \in (-1, 1)$ denote by

$$L(\alpha, \tau) = [\lambda(\alpha, \frac{\pi}{4}(1 - \tau_{lj}))] \in S, \tag{4.23}$$

where $\lambda(\alpha, x)$ is the function from Lemma 2.3.13. This is a family of elements of S, indexed by α.

Proposition 4.3.1. *Let $I = (I_1, \ldots, I_d)$ be a random vector with absolutely continuous marginals with support on the whole of $\mathbb{R}$ and an elliptical copula such that the spectral r.v. R in (2.13) is regularly varying with tail index $0 < \alpha^* < \infty$. Let $\widehat{\Lambda}^n = [\widehat{\lambda}_{lj}^n] \in S$ and $\widehat{\tau}^n = [\widehat{\tau}_{lj}^n] \in S$ satisfy (4.21) and (4.22). In addition, let $\widehat{\tau}_{lj}^n = \widehat{\tau}_{jl}^n \in (-1, 1)$ a.s. and $\widehat{\lambda}_{lj}^n = \widehat{\lambda}_{jl}^n \in (0, \frac{1+\widehat{\tau}_{lj}^n}{2})$ a.s. for every $n \in \mathbb{N}$, $l, j = 1, \ldots, d$ ($\widehat{\tau}_{jj}^n = \widehat{\lambda}_{jj}^n = 1$). Denote*

$$\widehat{\alpha}^n = \arg \min_{\alpha > 0} ||L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n||. \tag{4.24}$$

Then

(1) $\widehat{\alpha}^n$ exists and is unique a.s. for every $n \in \mathbb{N}$.

(2) $\widehat{\alpha}^n$ is a consistent estimate of α^*, i.e.

$$\widehat{\alpha}^n \xrightarrow{P} \alpha^*, \ n \to \infty.$$

(3) Denote by $\widehat{\theta}^n$ the vector, composed of all $\widehat{\tau}_{lj}^n, \widehat{\lambda}_{lj}^n$, $l = 1, \ldots, d$, $j = l + 1, \ldots, d$ and by θ^* the corresponding vector with the true Kendall's tau and tail dependence coefficients. If

$$\sqrt{n} \left(\widehat{\theta}^n - \theta^* \right) \xrightarrow{d} N(0, \Sigma), \ n \to \infty, \tag{4.25}$$

for some non-degenerate $(2d(d-1)) \times (2d(d-1))$ covariance matrix Σ, then

$$\sqrt{n} \left(\widehat{\alpha}^n - \alpha^* \right) \xrightarrow{d} N(0, \sigma), \ n \to \infty, \tag{4.26}$$

where $\sigma > 0$ is explicitly specified in (4.31).

Proof. (1) The following arguements are valid a.s. We note that by means of Lemma 2.3.13 (2), for every $\alpha > 0$ we have

$$0 \leq ||L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n|| \leq 4d^2.$$

Also, using Lemma 2.3.13 (3) and the fact that

$$\lim_{\alpha \to 0} \lambda(\alpha, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) = \frac{1 + \widehat{\tau}_{lj}^n}{2} > \widehat{\lambda}_{lj}^n$$

and

$$\lim_{\alpha \to \infty} \lambda(\alpha, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) = 0 < \widehat{\lambda}_{lj}^n$$

we have for $l, j = 1, \ldots, d$ and for every $n \in \mathbb{N}$ a unique solution of the equation

$$\lambda(\alpha, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) = \widehat{\lambda}_{lj}^n,$$

which we denote by $\alpha_{lj} \geq 0$. Let

$$\alpha_{max} = \max_{l,j=1,\ldots,d} \alpha_{lj}, \; \alpha_{min} = \min_{l,j=1,\ldots,d} \alpha_{lj}.$$

Due to the monotonicity of $\lambda(\alpha, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n))$ as a function of α (Lemma 2.3.13 (3)), for every $\alpha > \alpha_{max}$ we have

$$
\begin{aligned}
\lambda(\alpha, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) - \widehat{\lambda}_{lj}^n \; &< \; \lambda(\alpha_{max}, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) - \widehat{\lambda}_{lj}^n \\
&\leq \; \lambda(\alpha_{lj}, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) - \widehat{\lambda}_{lj}^n = 0 \,,
\end{aligned}
$$

therefore $|\lambda(\alpha, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) - \widehat{\lambda}_{lj}^n| > |\lambda(\alpha_{max}, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) - \widehat{\lambda}_{lj}^n|$ and hence

$$||L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n|| > ||L(\alpha_{max}, \widehat{\tau}^n) - \widehat{\Lambda}^n|| \,.$$

By analogy for any $\alpha < \alpha_{min}$

$$||L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n|| > ||L(\alpha_{min}, \widehat{\tau}^n) - \widehat{\Lambda}^n|| \,.$$

Therefore, either $\widehat{\alpha}^n = \alpha_{min} = \alpha_{max}$ or $||L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n||$ is bounded on the compact interval $[\alpha_{min}, \alpha_{max}]$, hence $\widehat{\alpha}^n$ exists.

To prove uniqueness, assume for some $n \in \mathbb{N}$ that there are $\alpha_1 \neq \alpha_2$ which are both minimizers of $||L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n||$. As $L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n$ is a symmetric matrix, we may concentrate on the upper triangle of the matrix, i.e. the same α_1, α_2 minimize also

$$G(\alpha) = \sum_{k=1}^{d(d-1)} (g_k(\alpha))^2 \,, \tag{4.27}$$

where

$$g_k(\alpha) = \lambda(\alpha, \frac{\pi}{4}(1 - \widehat{\tau}_{lj}^n)) - \widehat{\lambda}_{lj}^n, \ l = 1, \ldots, d, j = l + 1, \ldots, d,$$

i.e. $k = 1, \ldots, d(d-1)$. Next define

$$H(\alpha, w) = \sum_{k=1}^{d(d-1)} w_k g_k(\alpha)$$

where w is a $d(d-1)$-dimensional non-random vector with non-negative components.

Without loss of generality assume that $\alpha_1 < \alpha_2$, which implies by 2.3.13 (3) that $g_k(\alpha_1) > g_k(\alpha_2)$, $k = 1, \ldots, d(d-1)$. From the fact that α_2 is a minimizer of $G(\alpha)$ we obtain

$$\sum_{k=1}^{d(d-1)} \left(\frac{\partial}{\partial \alpha} g_k(\alpha_2) \right) g_k(\alpha_2) = 0$$

Since $\frac{\partial}{\partial \alpha} g_k(\alpha_2)$ are all negative (Lemma 2.3.13 (3)), there are only two cases (a) and (b).

(a) $g_k(\alpha_2) = 0$, $k = 1, \ldots, d(d-1)$. From this we obtain immediately that $G(\alpha_2) = 0$. However, as $g_k(\alpha_1) > 0$, $k = 1, \ldots, d(d-1)$, we have $G(\alpha_1) > 0 = G(\alpha_2)$, which is a contradiction.

(b) There exists some index j for which $g_j(\alpha_2) > 0$. Therefore we have also $g_j(\alpha_1) > g_j(\alpha_2) > 0$, and, for every $w_j \geq 0$, $R(w_j) = w_j(g_j(\alpha_1) - g_j(\alpha_2))$ is a positive and increasing function in w_j. Therefore we may always find a vector w such that

$$H(\alpha_1, w) > H(\alpha_2, w)$$

by selecting its j-th component sufficiently large.

Fix an $\epsilon > 0$ such that $\alpha_2 - \alpha_1 > \epsilon$. As $H(\alpha, w)$ is continuous in α (Lemma 2.3.13 (1)), we may find w such that

$$H(\alpha, w) > H(\alpha_2, w) \tag{4.28}$$

for every $\alpha_1 - \epsilon < \alpha < \alpha_1 + \epsilon$. We define a function

$$F(\alpha) = \begin{cases} (1 - Q(\alpha)G(\alpha) + Q(\alpha)H(\alpha, w) & \alpha_1 - \epsilon < \alpha < \alpha_1 + \epsilon, \\ H(\alpha, w) & \alpha > \alpha_1 + \epsilon, \end{cases}$$

where

$$Q(\alpha) = 2 \left(\frac{\alpha - \alpha_1}{\epsilon} \right)^4 - \left(\frac{\alpha - \alpha_1}{\epsilon} \right)^8.$$

Note that $F(\alpha)$ is continuous on its domain, as $Q(\alpha_1 + \epsilon) = 1$. The derivative of $F(\alpha)$ for $\alpha < \alpha_1 + \epsilon$ is given by

$$F'(\alpha) = Q'(\alpha)(H(\alpha, w) - G(\alpha)) + (1 - Q(\alpha))G'(\alpha) + Q(\alpha)H'(\alpha, w).$$

Since $Q'(\alpha_1 + \epsilon) = 0$, $F(\alpha)$ is also differentiable on its domain. Note also that α_1 is a strict local minimizer of F, because $F'(\alpha_1) = G'(\alpha_1) = 0$ and

$$
\begin{aligned}
F''(\alpha) &= Q''(\alpha)(H(\alpha, w) - G(\alpha)) + 2Q'(\alpha)(H'(\alpha, w) - G'(\alpha)) + \\
&\quad + Q(\alpha)H''(\alpha, w) + (1 - Q(\alpha))G''(\alpha),
\end{aligned}
$$

therefore $F''(\alpha_1) = G''(\alpha_1) > 0$. Therefore, for a sufficiently small ϵ we have $F(\alpha)$ is increasing in $\alpha \in (\alpha_1, \alpha_1 + \epsilon]$. On the other hand, by (4.28) we have $F(\alpha_1 + \epsilon) > F(\alpha_2)$, which implies that there exists a point $\alpha_3 \in [\alpha_1 + \epsilon, \alpha_2)$ such that $F'(\alpha_3) = 0$, therefore we have $\sum_{k=1}^{d(d-1)} w_k \frac{\partial}{\partial \alpha} g_k(\alpha_3) = 0$, which is a contradiction to the monotonicity of g_k, $k = 1, \ldots, d(d-1)$ (Lemma 2.3.13 (3)). This proves (1).

(2) As $\lambda(\alpha, \frac{\pi}{4}(1 - \tau))$ is continuous in τ (Lemma 2.3.13 (1)) by the continuous mapping theorem we have for every $\alpha > 0$

$$
L(\alpha, \widehat{\tau}^n) \xrightarrow{P} L(\alpha, \tau^*), \ n \to \infty.
$$

On the other hand $\widehat{\Lambda}^n \xrightarrow{P} \Lambda^*$, hence, by Proposition 2.3.12 we have

$$
L(\alpha^*, \widehat{\tau}^n) - \widehat{\Lambda}^n \xrightarrow{P} \overline{0}, \ n \to \infty,
$$

where $\overline{0} \in \mathbb{R}^d \times \mathbb{R}^d$ is the zero matrix. By monotonicity of $\lambda(\alpha, x)$ in α (Lemma 2.3.13 (3)), for every $\alpha \neq \alpha^*$ we have

$$
L(\alpha, \tau_n) - \Lambda_n \xrightarrow{P} \overline{A}(\alpha) \in \mathbb{R}^d \times \mathbb{R}^d, \ n \to \infty,
$$

where $\overline{A}(\alpha) \neq \overline{0}$, therefore $||\overline{A}(\alpha)|| > 0$. Hence we have

$$
\widehat{\alpha}^n = \arg \min_{\alpha > 0} ||L(\alpha, \widehat{\tau}^n) - \widehat{\Lambda}^n|| \xrightarrow{P} \alpha^*, \ n \to \infty.
$$

(3) To prove asymptotic normality, we use the delta method. We consider the function $G(\alpha) = G(\alpha, \widehat{\theta}^n)$ defined in (4.27). By the definition of $\widehat{\alpha}^n$, we have

$$
0 = \frac{\partial}{\partial \alpha} G(\widehat{\alpha}^n, \widehat{\theta}^n).
$$

By Taylor expansion of $\frac{\partial}{\partial \alpha} G(\widehat{\alpha}^n, \widehat{\theta}^n)$ around α^* and we get

$$
0 = \frac{\partial}{\partial \alpha} G(\alpha^*, \widehat{\theta}^n) + \frac{\partial^2}{\partial \alpha^2} G(\widetilde{\alpha}^n, \widehat{\theta}^n)(\widehat{\alpha}^n - \alpha^*)
$$

where $\widetilde{\alpha}^n$ lies between $\widehat{\alpha}^n$ and α^* a.s. for every $n \in \mathbb{N}$. Therefore

$$
\widehat{\alpha}^n - \alpha^* = -\left(\frac{\partial^2}{\partial \alpha^2} G(\widetilde{\alpha}^n, \widehat{\theta}^n)\right)^{-1} \frac{\partial}{\partial \alpha} G(\alpha^*, \widehat{\theta}^n). \tag{4.29}
$$

As $\widetilde{\alpha}^n$ lies between $\widehat{\alpha}^n$ and α^*, by Lemma 2.3.13 (1), the definition of G and the continuous mapping theorem we have

$$\frac{\partial^2}{\partial\alpha^2}G(\widetilde{\alpha}^n,\widehat{\theta}^n) \xrightarrow{P} \frac{\partial^2}{\partial\alpha^2}G(\alpha^*,\theta^*), \ n \to \infty.$$

Since $\alpha^* = \arg\min_\alpha G(\alpha,\theta^*)$, we obtain $\frac{\partial^2}{\partial\alpha^2}G(\alpha^*,\theta^*) < 0$.

Next we use a Taylor expansion around θ^* of the function $\frac{\partial}{\partial\alpha}G(\alpha^*,\theta)$:

$$\frac{\partial}{\partial\alpha}G(\alpha^*,\widehat{\theta}^n) = \frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*) + \nabla_\theta\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\widetilde{\theta}^n)\right)(\widehat{\theta}^n - \theta^*)^\mathrm{T},$$

where $\widetilde{\theta}^n$ lies componentwise between $\widehat{\theta}^n$ and θ^*. Since $\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*) = 0$ and

$$\nabla_\theta\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\widetilde{\theta}^n)\right) \xrightarrow{P} \nabla_\theta\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*)\right), \ n \to \infty,$$

by the continuous mapping theorem we have

$$\sqrt{n}\frac{\partial}{\partial\alpha}G(\alpha^*,\widehat{\theta}^n) \xrightarrow{d} N(0,\nabla_\theta\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*)\right)\Sigma\nabla_\theta\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*)\right)^\mathrm{T}). \qquad (4.30)$$

Going back to (4.29) we obtain

$$\sqrt{n}\,(\widehat{\alpha}^n - \alpha^*) \xrightarrow{d} N(0,\sigma), \ n \to \infty,$$

where

$$\sigma = \frac{\nabla_\theta\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*)\right)\Sigma\nabla_\theta\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*)\right)^\mathrm{T}}{\frac{\partial^2}{\partial\alpha^2}G(\alpha^*,\theta^*)}. \qquad (4.31)$$

$\square$

Remark 4.3.2. Denote by $\widehat{\theta}^n_\lambda$ the vector, composed of all $\widehat{\lambda}^n_{lj}$, $l = 1,\ldots,d$, $j = l+1,\ldots,d$, by θ^*_λ the corresponding vector with the true tail dependence coefficients and by $\widehat{\theta}^n_\tau$, θ^*_τ the corresponding Kendall's tau vectors. In practice, it is unlikely to find estimates $\widehat{\theta}^n_\lambda$ such that (4.25) holds. Most offen, under additional technical conditions, we are able to find a sequence $k = k(n)$ with the properties $k \to \infty$ and $\frac{k}{n} \to 0$ as $n \to \infty$ such that

$$\sqrt{k}(\widehat{\theta}^n_\lambda - \theta^*_\lambda) \xrightarrow{d} N(0,\Sigma_\lambda), \ n \to \infty,$$

for some non-degenerate covaariance matrix Σ_λ, see e.g. Schmidt and Stadtmüller [127]. Then, whenever

$$\sqrt{\frac{k}{n}}\sqrt{n}(\widehat{\theta}^n_\tau - \theta^*_\tau) \xrightarrow{d} 0, \ n \to \infty,$$

we may modify (4.30) to get

$$\sqrt{k}\frac{\partial}{\partial\alpha}G(\alpha^*,\widehat{\theta}^n) \xrightarrow{d} N(0,\nabla_{\theta_\lambda}\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*)\right)\Sigma_\lambda\nabla_{\theta_\lambda}\left(\frac{\partial}{\partial\alpha}G(\alpha^*,\theta^*)\right)^\mathrm{T}).$$

This leads to asymptotic normality of the estimator for the tail index α^* with the (much slower) rate $\sqrt{k(n)}$ and asymptotic variance depending on the variance of the tail dependence estimates only. $\square$

We return to our original problem for copula parameter estimation of a random vector $I = (I_1, \ldots, I_d)$ with arbitrary continuous marginals and elliptical copula, equal to the copula of the elliptical vector $Y = RAU$ as in (2.13). Recall that the parameters of the copula which we are interested in are the correlation matrix $[\rho_{lj}]_{l,j=1,\ldots,d}$ and the distribution of the spectral r.v. R. Due to Proposition 2.3.12, the only parameter in the distribution of R which has significant influence on the joint extremes is the tail index α. Therefore we focus on $[\rho_{lj}]_{l,j=1,\ldots,d}$ and α only. Proposition 4.3.1 suggests to estimate the tail index α by minimizing the distance in L^2 sense between the matrix of estimated from the extreme data tail dependence coefficients and the matrix of the implied by Proposition 2.3.12 and estimated from the entire sample coefficients. Note that the method can be applied to the lower, as well as to the upper tail dependence coefficients according to which are the extremes of interest in the specific application.

We suggest the following algorithm for estimation of the parameters of an elliptical copula, which we intend to apply in our credit risk model.

Algorithm 4.3.3. (1) Estimate the Kendall's tau matrix by $\widehat{\tau}^n = [\widehat{\tau}_{lj}^n]_{l,j=1,\ldots,d}$ as in (4.5).

(2) Estimate the correlation matrix by $[\widehat{\rho}_{lj}^n]_{l,j=1,\ldots,d}$ using (4.6) and the Kendall's tau estimates $\widehat{\tau}^n$.

(3) Estimate non-parametrically the lower tail dependence coefficients by $\widehat{\Lambda}^n = [\widehat{\lambda}_{lj}^n]_{l,j=1,\ldots,d}$ as in (4.10), i.e. using only the extreme observations.

(4) Estimate the tail index of the spectral r.v. R by $\widehat{\alpha}^n$ as in (4.24).

(5) In order to quantify the extremal dependence implied by the estimated $\widehat{\tau}^n$ and α, compute the implied tail dependence matrix $L(\widehat{\alpha}^n, \widehat{\tau}^n)$, where $L \in S$ is defined in (4.23).

Note that Proposition 4.3.1 is applicable to any consistent estimates of the Kendall's tau and the tail dependence coefficients. Furthermore, note that due to (4.31), the estimator of the tail index α directly inherits the variance in the estimation of these coefficients, and in particular the variance of the tail dependence coefficients estimates, see Remark 4.3.2. Bearing in mind the lack of long time-series which is typical in credit risk, it is extremely important to look for alternatives which utilize more information about the joint extreme observations than (4.10) does by simple counting. For this reason we consider a modification of Algorithm 4.3.3, where instead of (4.10) we use (4.14) at step 3.

Algorithm 4.3.4. (1) Estimate the Kendall's tau matrix by $\widehat{\tau}^n = [\widehat{\tau}_{lj}^n]_{l,j=1,\ldots,d}$ as in (4.5).

(2) Estimate the correlation matrix by $[\widehat{\rho}^n_{lj}]_{l,j=1,\ldots,d}$ using (4.6) and the Kendall's tau estimates $\widehat{\tau}^n$.

(3) Estimate non-parametrically the lower tail dependence coefficients by $\widetilde{\Lambda}^n = [\widetilde{\lambda}^n_{lj}]_{l,j=1,\ldots,d}$ as in (4.14), i.e. using the information from the joint extreme observations.

(4) Estimate the tail index of the spectral r.v. R by $\widehat{\alpha}^n$ as in (4.24).

(5) In order to quantify the extremal dependence implied by the estimated $\widehat{\tau}^n$ and α, compute the implied tail dependence matrix $L(\widehat{\alpha}^n, \widehat{\tau}^n)$, where $L \in S$ is defined in (4.23).

4.4 Implementation

Recall our model (3.5) for the asset (log-)returns $Y = (Y_1, \ldots, Y_m)$ of the obligors in the portfolio. Recall also that by Proposition 3.1.1 we are interested in the copula of Y. Since it is an elliptical copula, provided that we have a sample of i.i.d. copies of the assets Y, we could potentially apply Algorithms 4.3.3 or 4.3.4 or any other method presented in the previous two sections to calibrate this copula. Apart from the curse of dimensionality, such a straightforward approach is impossible as the assets in practice are not observable quantities, see e.g. Bluhm et al. [16], Section 1.2.3. Instead, we assume that on an obligor-specific level we observe only $S^{(k,s)} = (S_1^{(k,s)}, \ldots, S_m^{(k,s)})$, $k, s = 1, \ldots, n$, where

$$S_j^{(k,s)} = \text{sign}[Y_j^{(k)} - Y_j^{(s)}], j = 1, \ldots, m, k, s = 1, \ldots, n, \tag{4.32}$$

i.e. we observe only whether the asset returns in a given period are higher or lower than the returns in the other periods.

Remark 4.4.1. There are various ways how one may construct such a sample in practice. For obligors listed on a stock exchange, the KMV model [89, 90] produces estimates $\widehat{Y}^{(k)} = \left(\widehat{Y}_1^{(k)}, \ldots, \widehat{Y}_m^{(k)}\right)$, $k = 1, \ldots, n$, by a combination of qualitative analysis, econometric research and structural modelling in the framework of the Merton [115] model. The accuracy of these estimates is often questionable (see Duan et al. [37]), however, one may expect that at least $\widehat{S}_j^{(k,s)} = \text{sign}[\widehat{Y}_j^{(k)} - \widehat{Y}_j^{(s)}], j = 1, \ldots, m, k, s = 1, \ldots, n$, are a good approximation of the desired quantities in (4.32).

Alternatively, a more straightforward approach similar to the one taken in CreditMetrics [74] would be to use directly $\widetilde{S}_j^{(k,s)} = \text{sign}[E_j^{(k)} - E_j^{(s)}]$, $j = 1, \ldots, m, k, s = 1, \ldots, n$, where $E^{(k)} = (E_1^{(k)}, \ldots, E_m^{(k)})$, $k = 1, \ldots, n$, are the observable stock returns of the respective obligors as an approximation of (4.32). Note that assuming $\widetilde{S}_j^{(k,s)} = S_j^{(k,s)}$, $j = 1, \ldots, m, k, s = 1, \ldots, n$, is weaker than the assumption $Y_j = a_j E_j + b_j$, $a_j > 0, b_j \in \mathbb{R}$, $j = 1, \ldots, m$, which is essentially taken in CreditMetrics [74].

If the obligors in the portfolio of interest are not listed on a stock exchange, one may potentially use as an approximation of (4.32) $P_j^{(k,s)} = -\text{sign}[p_j^{(k)} - p_j^{(s)}]$, $j = 1, \ldots, m, k, s =$

$1, \ldots, n$, where $p_j^{(k)}$ denotes the average number of defaults of obligors of the same type as obligor j (e.g. in the same initial rating category) in period k. Another opportunity would be to use for $p_j^{(k)}$ the ratio between downgrades and upgrades in period k. Any credit-scoring models (e.g. Altman's Z-score [2]) could also be used for this purpose. □

Additionally we have the problem with multidimensionality (m is large in the contemporary credit portfolios). Here the use of the factor model (3.5) is crucial. Recall that each obligor in the portfolio is influenced by a systematic part $\sum_{l=1}^{p} \alpha_{j,l} W Z_l$, $j = 1, \ldots, m$. For this systematic part, we assume that

$$W Z_l = H_l(I_l), \ l = 1, \ldots, p, \tag{4.33}$$

where $H_l : \mathbb{R} \to \mathbb{R}$ are some continuous and strictly increasing functions and $I = (I_1, \ldots, I_p)$ are observable macroeconomic factors. In practice, $I = (I_1, \ldots, I_p)$ are taken to be the log-returns of regional / industry stock indices.

Remark 4.4.2. A similar assumption is taken in CreditMetrics [74], see also Daul et al. [30]. Note that, again, we do not impose any further restrictions on the mapping functions H_l, $l = 1, \ldots, p$ (like linearity in the CreditMetrics model), hence $I = (I_1, \ldots, I_p)$ is a random vector with elliptical copula and arbitrary marginals $F_1, \ldots, F_p$. □

To summarize, recall the original problem for estimation of the parameters in model (3.5), i.e.
  (a) the distribution of the common shock W;
  (b) the correlation matrix Σ of the common factors $Z_1, \ldots, Z_p$;
  (c) the factor loadings $\alpha_{j,l}, \sigma_j$, $j = 1, \ldots, m$, $l = 1, \ldots, p$.
The available data consist of:
  (1) A sample of i.i.d copies of $I = (I_1, \ldots, I_p)$, which satisfies (4.33). This sample reflects the development of the market and is typically composed of stock index returns.
  (2) A sample $S^{(k,s)} = (S_1^{(k,s)}, \ldots, S_m^{(k,s)})$, $k, s = 1, \ldots, n$, which satisfies (4.32). For $j = 1, \ldots, m$, $S_j^{(k,s)}$ is a r.v. with range $\{-1, 1\}$ which reflects the knowledge of whether the asset return of the specific obligor j is higher in period k than in period s.

To estimate the papameters (a) and (b), it is sufficient to apply Algorithm 4.3.4 on the sample of the stock indices $I = (I_1, \ldots, I_p)$. Note that in this way we obtain only the tail index α of the r.v. W, whereas the d.f. of W may well have additional parameters. However, in our application the tail index is the only parameter of significant interest. First of all, by Proposition 2.3.14, the random vector of the assets Y inherits the tail index from W. Then recall from Proposition 2.3.12 and from Example 3.2.5 that this is the only parameter of W which plays a role in the extremal dependence of the assets and thus has a significant influence on the tail of the portfolio loss distribution. Finally, note

that due to Proposition 3.1.1 we are interested only in the copula of Y, which is invariant under, say, a multiplication of W with some positive constant.

Thus, we are left with the problem of estimating the factor loadings $\alpha_{j,l}$ and σ_j, $l = 1, \ldots, p$, $j = 1, \ldots, m$ in (c). Note that, due to the assumption that $\mathrm{var}[Y_j \,|\, W] = W$, for $j = 1, \ldots, m$ we have the following relation:

$$\sum_{l,d=1}^{p} \alpha_{j,l} \Sigma_{ld} \alpha_{j,d} = 1 - \sigma_j^2 \,. \tag{4.34}$$

Again, because of Proposition 3.1.1, this assumption is made without loss of generality.

We suggest two principle approaches to estimate the factor loadings:

Method I: Estimate $\alpha_{j,l}$, $j = 1, \ldots, m$, $l = 1, \ldots, p$ and then use (4.34) to obtain σ_j, $j = 1, \ldots, m$.

Method II: Assume that

$$\alpha_{j,l} = \frac{\sqrt{1 - \sigma_j^2}}{\sqrt{\sum_{d=1}^{p} w_{jl} w_{jd} \Sigma_{ld}}} w_{jl} \,, \quad j = 1, \ldots, m, \; l = 1, \ldots, p, \tag{4.35}$$

where $w_{jl} \geq 0$, $j = 1, \ldots, m$, $l = 1, \ldots, p$, are known quantities. Then we are left only with the parameters σ_j, $j = 1, \ldots, m$, to specify. Note that this assumption is consistent with (4.34). This method is similar to the approach in CreditMetrics [74], where $w_{j,l} \geq 0$, $j = 1, \ldots, m$, $l = 1, \ldots, p$, are called country / industry participations and are required as an input of that model. These quantities give the relative weight of the country / industry common factors for each obligor, which can be "estimated" by purely qualitative methods.

We need the following proposition, parts (a) and (b) are for Method I and parts (c) and (d) are for Method II. In part (a) we show that $\alpha_{j,l}$, $j = 1, \ldots, m$, $l = 1, \ldots, p$, satisfy a system of linear equations. In part (b) we suggest an estimate for the unknown coefficients of the system. In part (c) we show that σ_j, $j = 1, \ldots, m$, can be expresses as a function of the Kendall's tau of the marginal asset return Y_j and a particular transformation of the observable market risk factors $I = (I_1, \ldots, I_p)$. Based on that, in part (d) we suggest an estimate for σ_j, $j = 1, \ldots, m$.

Proposition 4.4.3. *Let* $Y = (Y_1, \ldots, Y_m)$, $S = (S_1, \ldots, S_m)$ *and* $I = (I_1, \ldots, I_p)$ *satisfy (3.5), (4.32) and (4.33) resp. and* $\alpha_{j,l}, \sigma_j$, $j = 1, \ldots, m$, $l = 1, \ldots, p$ *satisfy (4.34). Then, for* $j = 1, \ldots, m$:

   (a) The random vector $(Y_j, \alpha_{j,1} W Z_1, \ldots, \alpha_{j,p} W Z_p, \sigma_j W \epsilon_j)$ *is elliptical and we have*

$$\sum_{l=1}^{p} \alpha_{j,d} \alpha_{j,l} \Sigma_{ld} = \sin\left(\frac{\pi}{2} \tau(Y_j, I_d)\right), \quad d = 1, \ldots, p.$$

(b) Denoting

$$\widehat{\tau}_{j,d}^n = \binom{n}{2}^{-1} \sum_{k>s} S_j^{(k,s)} \mathrm{sign}\left[I_d^{(k)} - I_d^{(s)}\right], \ d = 1, \ldots, p,$$

where $I^{(k)} = (I_1^{(k)}, \ldots, I_p^{(k)})$, $k = 1, \ldots, n$, are i.i.d copies of I, we have

$$\widehat{\tau}_{j,d}^n \xrightarrow{P} \tau(Y_j, I_d), \ n \to \infty.$$

(c) If additionally (4.35) holds, then we have

$$\sqrt{1 - \sigma_j^2} = \sin\left(\frac{\pi}{2}\tau(Y_j, A_j)\right), \tag{4.36}$$

where

$$A_j = \sum_{l=1}^p \gamma_{j,l} F_\alpha^{-1}(F_l(I_l)),$$

$$\gamma_{j,l} = \frac{w_{jl}}{\sqrt{\sum_{l,d=1}^p w_{jl}w_{jd}\Sigma_{ld}}},$$

$F_l(x)$ is the d.f. of I_l, $l = 1, \ldots, p$, and F_α is the d.f. of WZ_0, $Z_0 \in N(0,1)$, independent of W, where W is from (3.5).

(d) Furthermore, denoting

$$\widehat{\tau}_j^n = \binom{n}{2}^{-1} \sum_{k>s} S_j^{(k,s)} \mathrm{sign}\left[\widehat{A}_j^{(k)}(n) - \widehat{A}_j^{(s)}(n)\right],$$

where

$$\widehat{A}_j^{(k)}(n) = \sum_{l=1}^p \gamma_{j,l} F_\alpha^{-1}(F_l^E(I_l^{(k)})), \ k = 1, \ldots, n,$$

and $F_l^E(x)$ is the empirical d.f. of I_l, $l = 1, \ldots, p$, we have

$$\widehat{\tau}_j^n \xrightarrow{P} \tau(Y_j, A_j), \ n \to \infty. \tag{4.37}$$

Proof. Fix $j \in \{1, \ldots, m\}$. By (3.5), the random vector $(Y_j, \alpha_{j,1}WZ_1, \ldots, \alpha_{j,p}WZ_p, \sigma_j W\epsilon_j)$ can be obtained by a linear transformation of the elliptical random vector $(\alpha_{j,1}WZ_1, \ldots, \alpha_{j,p}WZ_p, \sigma_j W\epsilon_j)$, and hence it is an elliptical random vector by Lemma 2.3.7. For the correlation we have

$$\rho(Y_j, \alpha_{j,d}WZ_d) = \sum_{l=1}^p \alpha_{j,d}\alpha_{j,l}\Sigma_{ld}, \ d = 1, \ldots, p.$$

Applying (2.15) we obtain

$$\sum_{l=1}^p \alpha_{j,d}\alpha_{j,l}\Sigma_{ld} = \sin\left(\frac{\pi}{2}\tau(Y_j, \alpha_{j,d}WZ_d)\right), \ d = 1, \ldots, p.$$

As Kendall's tau is invariant under strictly increasing marginal transformations, we get by means of (4.33) $\tau(Y_j, \alpha_{j,d} W Z_d) = \tau(Y_j, I_d)$, $d = 1, \ldots, p$, i.e. (a).

(b) follows directly from (a) and (4.5).

To prove (c), we note that the random vector $(Y_j, \sum_{l=1}^{p} \alpha_{j,l} W Z_l) = (Y_j, \sum_{l=1}^{p} \frac{\gamma_{j,l}}{\sqrt{1-\sigma_j^2}} W Z_l)$ has elliptical distribution (see Lemma 2.3.7) and correlation

$$\rho(Y_j, \sum_{l=1}^{p} \alpha_{j,l} W Z_l) = \sqrt{1 - \sigma_j^2}.$$

By (2.15),

$$\sqrt{1 - \sigma_j^2} = \sin\left(\frac{\pi}{2} \tau(Y_j, \sum_{l=1}^{p} \frac{\gamma_{j,l}}{\sqrt{1 - \sigma_j^2}} W Z_l)\right).$$

As Kendall's tau is invariant under strictly increasing transformation of the marginals, we have

$$\tau(Y_j, \sum_{l=1}^{p} \frac{\gamma_{j,l}}{\sqrt{1 - \sigma_j^2}} W Z_l) = \tau(Y_j, \sum_{l=1}^{p} \gamma_{j,l} W Z_l).$$

By Corrolary 2.1.8, as $W Z_l$ is comonotone with I_l, $l = 1, \ldots, p$, we have

$$(W Z_1, \ldots, W Z_p) \stackrel{d}{=} (F_\alpha^{-1}(F_1(I_1)), \ldots, F_\alpha^{-1}(F_p(I_p))),$$

hence we obtain (4.36).

To prove (4.37) note that for $k, s = 1, \ldots, n$

$$E\left[S_j^{(k,s)} \text{sign}\left[\widehat{A}_j^{(k)}(n) - \widehat{A}_j^{(s)}(n)\right]\right] = P((Y_j^{(k)} - Y_j^{(s)})(\widehat{A}_j^{(k)}(n) - \widehat{A}_j^{(s)}(n)) > 0)$$
$$- P((Y_j^{(k)} - Y_j^{(s)})(\widehat{A}_j^{(k)}(n) - \widehat{A}_j^{(s)}(n)) < 0).$$

By the continuous mapping theorem we have for $k = 1, \ldots, n$

$$\widehat{A}_j^{(k)}(n) \stackrel{a.s.}{\to} A_j^{(k)}, n \to \infty,$$

where

$$A_j^{(k)} = \sum_{l=1}^{p} \gamma_{j,l} F_\alpha^{-1}(F_l(I_l^{(k)})), \quad k = 1, \ldots, n,$$

and $F_l(x)$ is the true d.f. of I_l, $l = 1, \ldots, p$. Therefore

$$\lim_{n \to \infty} E\left[S_j^{(k,s)} \text{sign}\left[(\widehat{A}_j^{(k)}(n) - \widehat{A}_j^{(s)}(n))\right]\right] = \tau(Y_j, A_j),$$

i.e. the estimator $\widehat{\tau}_j^n$ is asymptotically unbiased. Furthermore we have for $\epsilon > 0$

$$P\left(|\widehat{\tau}_j^n - \tau(Y_j, A_j)| > \epsilon\right) \leq P\left(|\widehat{\tau}_j^n - \overline{\tau}_j^n| > \epsilon/2\right) + P\left(|\overline{\tau}_j^n - \tau(Y_j, A_j)| > \epsilon/2\right), \quad (4.38)$$

where

$$\overline{\tau}_j^n = \binom{n}{2}^{-1} \sum_{k>s} S_j^{(k,s)} \text{sign} \left[A_j^{(k)} - A_j^{(s)} \right],$$

The second summand in (4.38) converges to 0 as $n \to \infty$ by (c) and (4.5). Furthermore, by Chebishev's inequality,

$$
\begin{aligned}
P(|\widehat{\tau}_j^n - \overline{\tau}_j^n| > \epsilon/2) \quad &\leq \quad \frac{2}{\epsilon} E \left| S_j^{(k,s)} \text{sign} \left[\widehat{A}_j^{(k)}(n) - \widehat{A}_j^{(s)}(n) \right] - S_j^{(k,s)} \text{sign} \left[A_j^{(k)} - A_j^{(s)} \right] \right| \\
&\leq \quad P \left((A_j^{(k)} - A_j^{(s)})(\widehat{A}_j^{(k)}(n) - \widehat{A}_j^{(s)}(n)) < 0 \right)
\end{aligned}
$$

Since $\widehat{A}_j^{(k)}(n) \overset{\text{a.s.}}{\to} A_j^{(k)}$, $n \to \infty$, we obtain (d). $\qquad\qquad\qquad\square$

4.5   Numerical examples

We illustrate the properties of the suggested estimators by numerical examples. All histograms in this section are scaled and smoothed using the same normal kernel smoothing function and the same default bandwidth in Matlab [110].

In a simulation study we examine the accuracy of the copula estimation when the marginals are heavy-tailed and / or non-symmetric.

Example 4.5.1. [T-model, heavy-tailed marginals, emprirical tail dependence estimator] We consider the following model: The random vector (I_1, I_2, I_3) has t-copula with $\nu = 5$ degrees of freedom and a correlation matrix $\rho_{12} = 0.3$, $\rho_{13} = 0.4$, $\rho_{23} = 0.6$. The tail index for this copula is $\alpha = \nu$, see Example 3.2.1. For the marginals we use $I_1 \in N(2,4)$ (normal with mean 2 and variance 4), $I_2 \in GPD(0.3, 10)$ (see Definition 2.2.7), and $I_3 \in t(1)$ (t-distribution with $\nu = 1$ degrees of freedom (Cauchy distribution)). Note that the d.f. of I_2 is not symmetric (not elliptical) and the covariance matrix does not exist as $\text{var}(I_3) = \infty$.

Our goal is to assess the performance of Algorithm 4.3.3 when applied to this model. For this reason we simulate $n = 1000$ i.i.d copies of the vector (I_1, I_2, I_3) and estimate the copula parameters using the algorithm. The simulation is repeated 1000 times. In Table 4.1 we summarize the results.

In Figure 4.1 we present the results on Kendall's tau matrix estimated by (4.5). We observe that the estimation is accurate even in the cases of asymmetric / heavy-tailed marginals, i.e. we obtain narrow confidence bounds and the histograms of the estimates are symmetric around the true values. This accuracy transfers also to the estimated correlation matrix, see Figure 4.1.

In Figure 4.2 we present the results on the tail dependence matrix estimated by (4.10). We observe that the estimation is not very accurate (high empirical variance, wide confidence bounds), which is due to the small sample of observations (1000) on which it
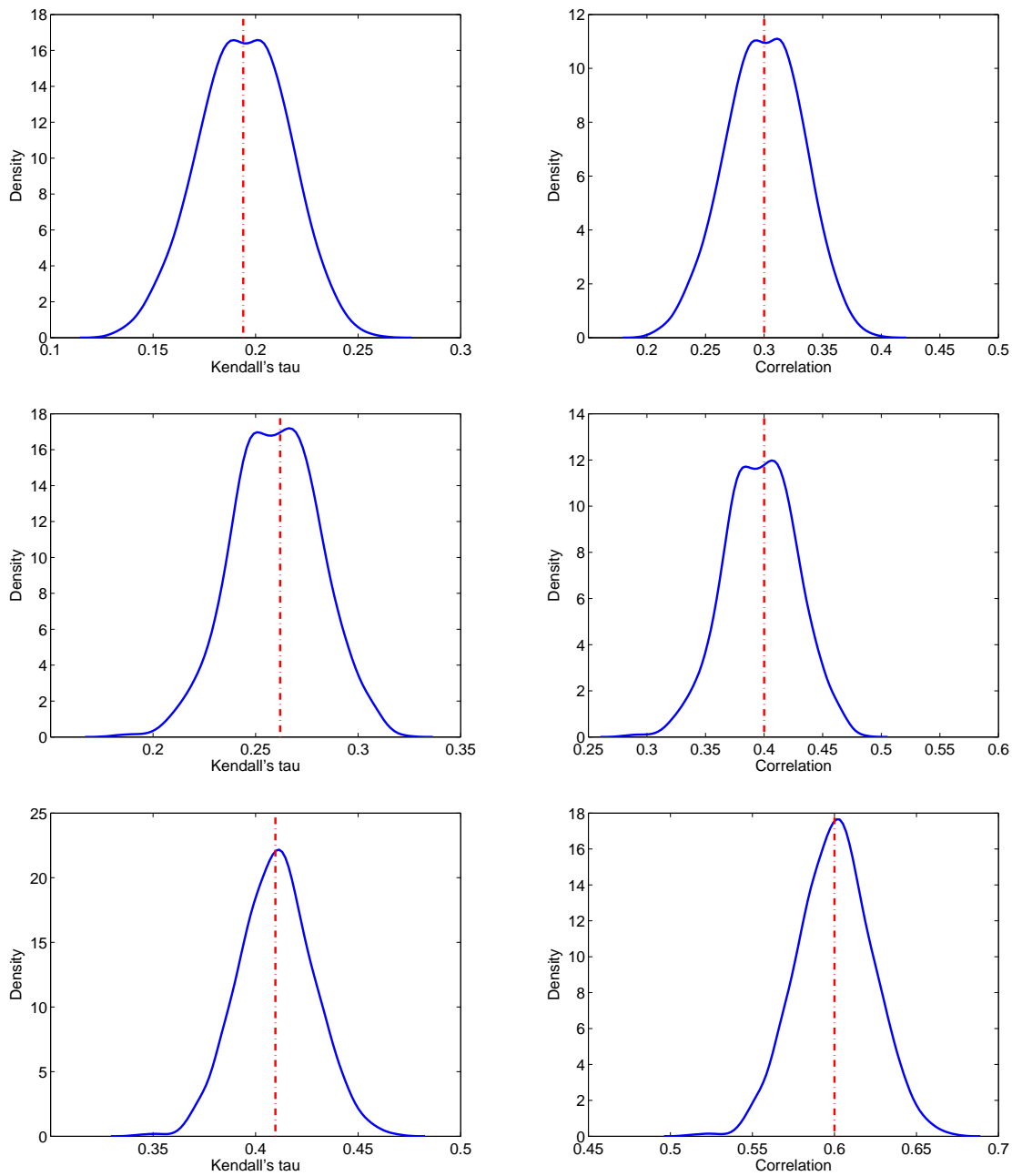
Figure 4.1: Left column (τ_{12}, τ_{13}, τ_{23}): the histograms of the Kendall's tau estimates compared to the true values for Example 4.5.1 with sample size $n = 1000$. The histograms are symmetric around the true values and the empirical variance is low. Right column (ρ_{12}, ρ_{13}, ρ_{23}): the histograms of correlation estimates compared to the true values for Example 4.5.1 with sample size $n = 1000$. Again, histograms are symmetric around the true values and the estimation is accurate even in cases when the covariance does not exist (ρ_{13}, ρ_{23}).
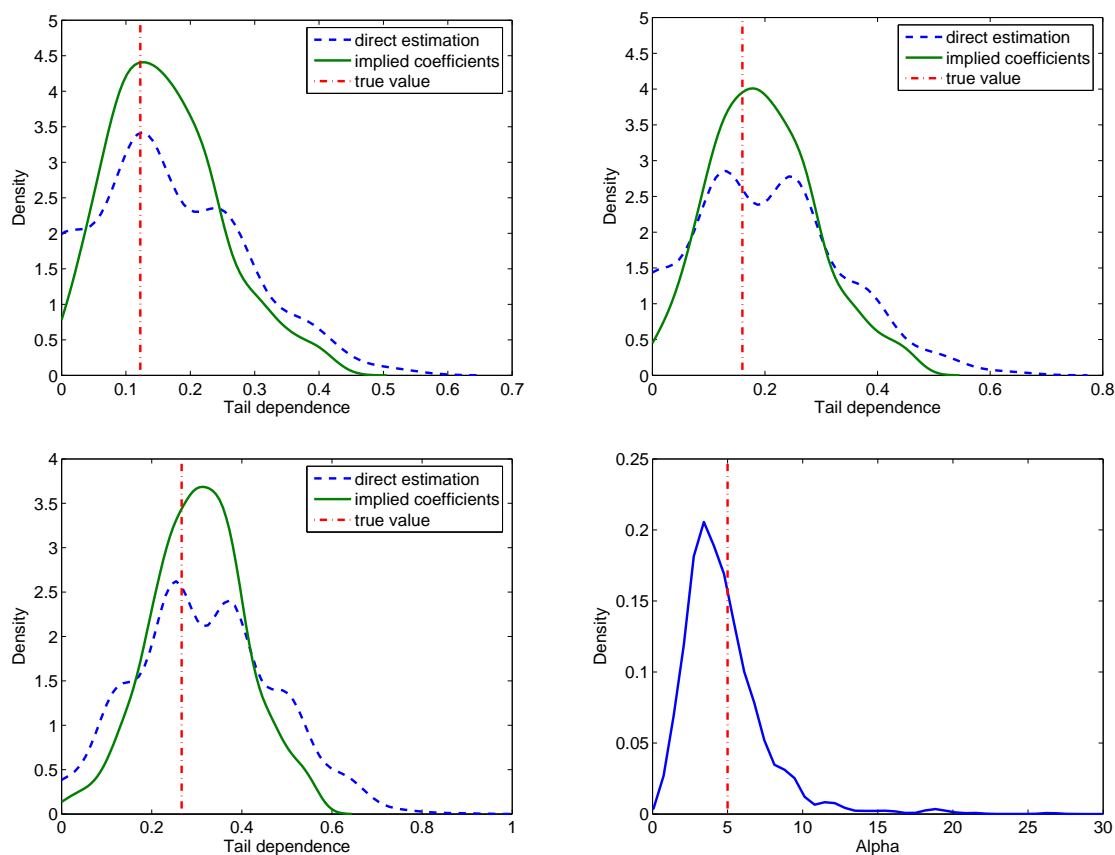
Figure 4.2: Upper row and bottom row, left (λ_{12}, λ_{13}, λ_{23}): The histograms of the direct estimates for tail dependence by (4.10), compared to the implied (by the estimated α and Kendall's tau) estimates for tail dependence and to the true values for Example 4.5.1 with sample size $n = 1000$. The implied estimates are more centered around the true values, and have less empirical variance. Bottom row, right: the histogram of the estimates for the tail index α of the copula compared to the true value $\alpha^* = \nu = 5$ for Example 4.5.1 with sample size $n = 1000$. The histogram is not symmetric around the true value 5. On the other hand, most of its mass is in the region [2,8], which includes the true value 5, and the empirical mean is 5.04.

	mean estimate	true value	m.s.e.	std
α	5.0366	5	22.1086	4.7042
ρ_{12}	0.3001	0.3	0.0010	0.0321
ρ_{13}	0.3969	0.4	0.0009	0.0301
ρ_{23}	0.6002	0.6	0.0005	0.0226
λ_{12}(implied by α)	0.1618	0.1224	0.0092	0.0877
λ_{13}(implied by α)	0.1983	0.1559	0.0108	0.0950
λ_{23}(implied by α)	0.3031	0.2666	0.0125	0.1059

Table 4.1: Estimation of the *t*-copula for the model from Example 4.5.1 with sample size $n = 1000$. The estimators of correlation and tail dependence perform equally well across the different marginals. The correlation estimates are accurate (low empirical standard deviation (std) and mean square errors (m.s.e.)). The estimators of the tail index α of the copula and the tail dependence coefficents have high empirical variance and m.s.e.

is based. Note that (4.10) uses only the extreme observations, which further increases the variance of the estimators (see e.g. Frahm et al. [58] or Section 4.2.2 for a detailed discussion).

In Figure 4.2 we present the results on the tail index α estimated by (4.24). Most of the realizations are in the region [2,8] which includes the true value 5. The empirical mean of the estimate is equal to the true value. Furthermore, taking the estimated α and the estimated Kendall's tau matrix, we compute also the implied tail dependence coefficients as in step (5) of Algorithm 4.3.3. We observe that in this way the estimates are improved as compared to the direct estimates using (4.10). However, the histogram of the estimated α is not centered around its mean, it is very skewed and there are cases when the estimate $\widehat{\alpha}^n$ takes very large values compared to the true one. This makes the asymptotic confidence bounds derived through (4.26) mainly of theoretical interest, at least for smaller samples.

In order to assess the accuracy of the method when applied to larger samples, we increase gradually the sample size to $n = 10000$ and at each step apply Algorithm 4.3.3. In Figure 4.3 we observe that the empirical standard deviation of the estimator (4.24) of the tail index α is quite satisfactory at sample size $n = 10000$.   □

In the next example we use simulated data to examine the robustness of the estimation when the copula is not elliptical.

Example 4.5.2. [Clayton copula]
We consider the following model: The marginals of the random vector (I_1, I_2) are standard
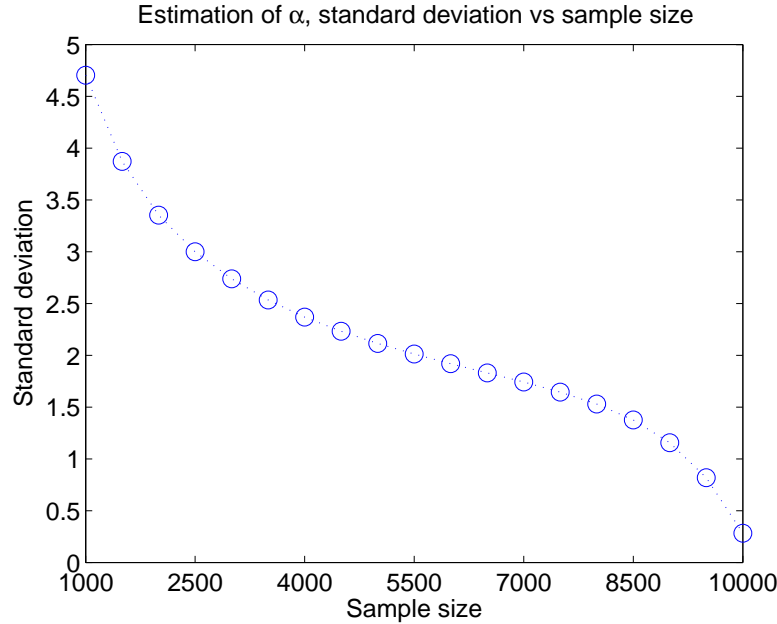
Figure 4.3: The empirical standard deviation of the estimator (4.24) for the tail index α of the copula in Example 4.5.1, as a function of the sample size.

normal $N(0,1)$. The copula of (I_1, I_2) is the 2-dimensional Clayton copula

$$C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$$

with $\theta = 1$.

This copula is not elliptical, and in particular it is not radially symmetric, see Figure 4.4. It has a tail dependence coefficient $\lambda = 2^{-1/\theta} = 0.5$ and Kendall's tau $\tau = \frac{\theta}{\theta+2} = \frac{1}{3}$. However, all elliptical copulas with Kendall's tau $\tau = \frac{1}{3}$ and tail index $\alpha = 1$ have the same tail dependence coefficient. In Figure 4.4 we compare $C(p,p)$ with $C_t(p,p)$, a 2-dimensional t-copula with $\nu = \alpha = 1$ degrees of freedom and Kendall's tau $\tau = \frac{1}{3}$. We focus on the small values of p (in the region $[0.0001, 0.025]$). We observe that $C(p,p)$ and $C_t(p,p)$ are practically indistinguishable. Therefore, even if we assume a wrong elliptical copula model, we would obtain comparatively similar results with respect to joint extreme event probabilities, provided that the parameters of the elliptical copula are selected appropriately.

Our goal is to assess the performance of Algorithm 4.3.3 when applied to data, which is coming from a non-elliptical copula model. For this reason we simulate $n = 1000$ i.i.d copies of (I_1, I_2). We apply Algorithm 4.3.3 to estimate the Kendall's tau, the tail dependence and the tail index α. We repeat the simulation 1000 times. In Table 4.2 we summarize the results.

In Figure 4.5 we present the results on the Kendall's tau estimated as in (4.5). We observe that the estimation is accurate (low empirical variance, histogram symmetric around
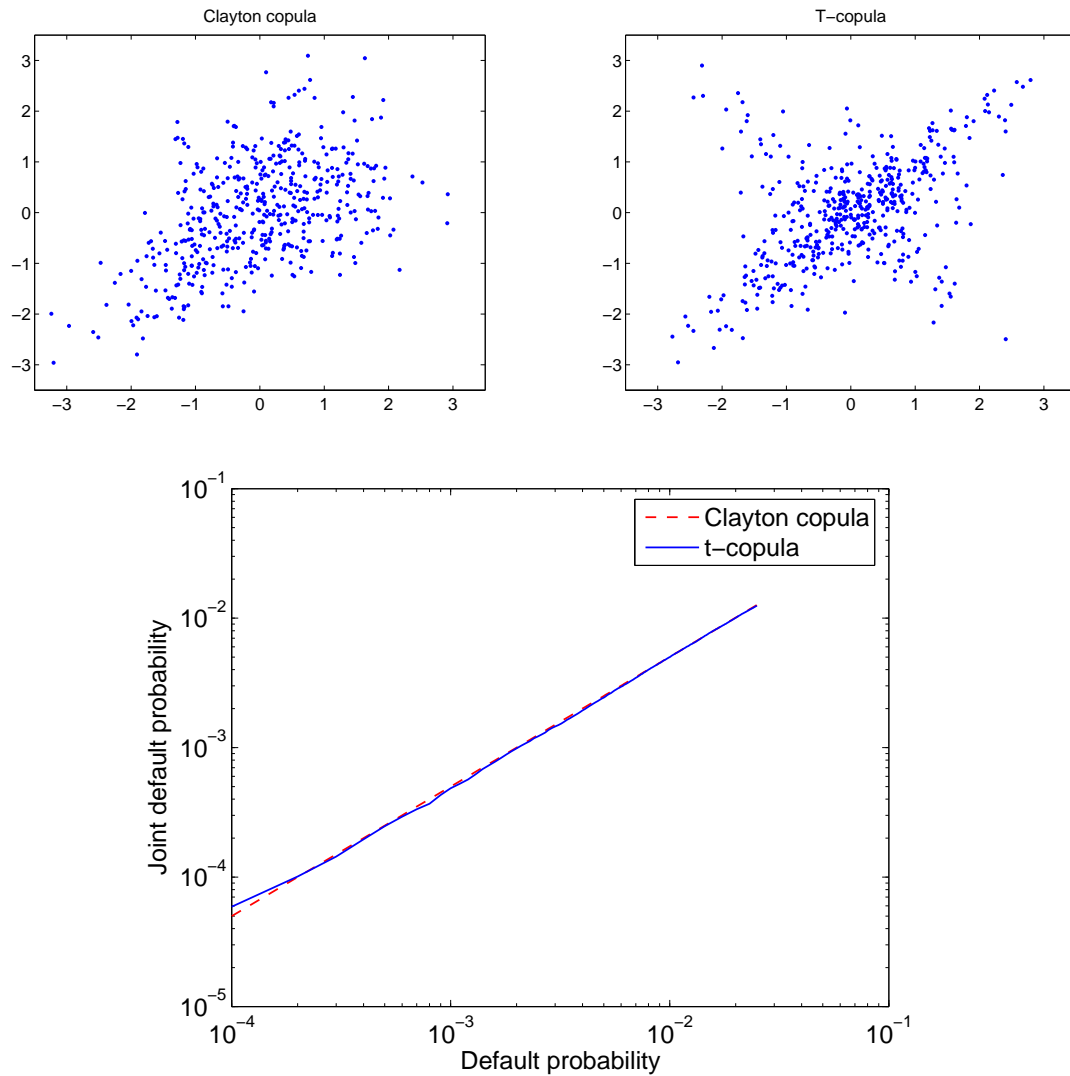
Figure 4.4: Upper row, left figure: 1000 realizations of the random vector (I_1, I_2) from Example 4.5.2. Upper row, right figure: 1000 realizations of the random vector (Y_1, Y_2) with standard normal $N(0,1)$ marginals and 2-dimensional t-copula with $\nu = \alpha = 1$ degrees of freedom and Kendall's tau $\tau = \frac{1}{3}$. Both vectors have the same marginals, same Kendall's tau and same tail dependence coefficients. The plots are quite different in the center, i.e. the two copulas are different close to the mean values. Bottom row: $C(p, p) = P(I_1 < \Phi^{-1}(p), I_2 < \Phi^{-1}(p))$ compared to $C_t(p, p) = P(Y_1 < \Phi^{-1}(p), Y_2 < \Phi^{-1}(p))$. The two joint probabilities are practically indistinguishable. This is particularly important in view of our credit risk model, as $C(p, p)$ is interpreted as joint default probabiliy of two credits, see formula (3.4).
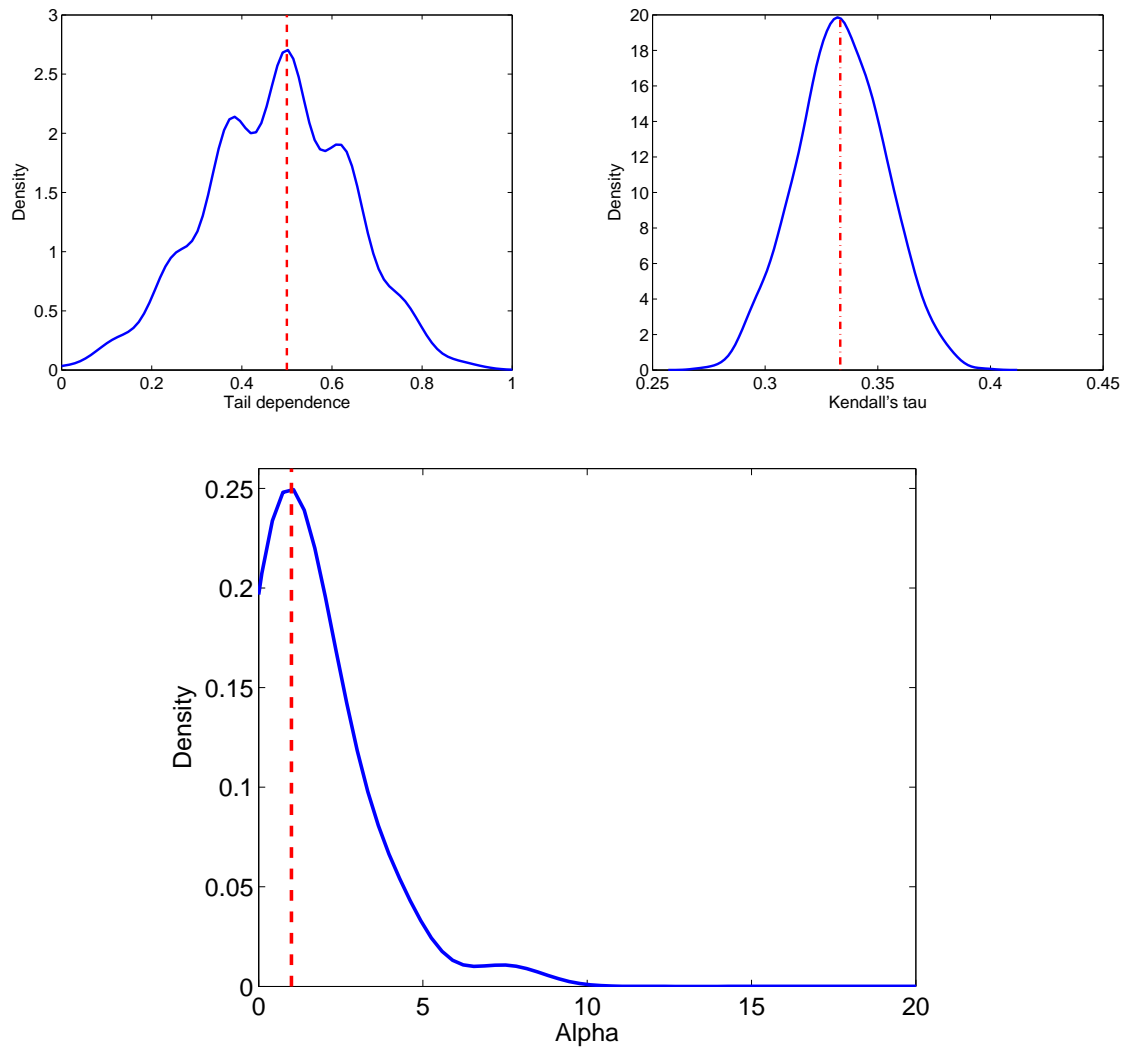
Figure 4.5: Upper row, left figure: The histogram of the tail dependence estimates for a sample of $n = 1000$ copies of the random vector (I_1, I_2) from Example 4.5.2. The estimate has high empirical variance. Upper row, right figure: The histogram of Kendall's tau estimates for the same sample. The estimate has low empirical variance. Bottom row: The histogram of the estimated tail index α of the (incorreclty assumed) elliptical copula. The mode is exacly equal to 1, which is the value for which the elliptical copula is closest to the true copula with respect to joint extremes, see Figure 4.4. Most of the mass is in the region [0,8], however, the histogram is not symmetric.

	mean estimate	true value	m.s.e.	std
τ	0.3336	1/3	0.0003	0.0196
λ	0.4765	0.5	0.0235	0.1549
α	1.6482	1	9.5512	1.6306

Table 4.2: Application of Algorithm 4.3.3 to a sample of $n = 1000$ i.i.d. vectors with Clayton copula and standard normal margins as in Example 4.5.2. The estimates for Kendall's tau are accurate (low empirical standard deviation (std) and mean square error (m.s.e)). The estimates for the tail dependence have high empirical variance. This results in high variance in the estimation of the tail index α. The 'true value' of the copula parameter α is chosen such that the elliptical copula has the same tail dependence as the Clayton copula.

the true value). We present also the results on the tail dependence coeffient estimated as in (4.10). As in Example 4.5.1, we observe that the estimation has a rather high empirical variance. Furthermore, the estimator (4.24) of the tail index α inherits the errors from (4.10). The histogram of the estimates of α is not centered around its mean, it is skewed and there are cases when the estimate $\widehat{\alpha}^n$ takes very large values. On the other hand, its mode is exactly equal to 1, and most of the mass is in the region [0,5]. Hence the estimation method seems robust with respect to joint extreme event probabilities.   □

The two previous examples show that the main problem in the proposed method comes from the high variance in the estimation of the tail dependence coefficients by (4.10). This is expected to happen and is due to the small number of observations on which the estimation is based (only the extreme observations are taken into account). However, there is a room for improvement. Note that Proposition 4.3.1 works for any consistent estimates of the tail dependence coefficients, and hence Algorithm 4.3.4 can also be applied.

Example 4.5.3. [Example 4.5.1 continued, alternative tail dependence estimator]
Using the simulated data from Example 4.5.1, we apply Algorithm 4.3.4. Recall that we consider samples of size $n = 1000$ i.i.d copies of the vector (I_1, I_2, I_3) with t-copula and various marginals. Since steps (1) and (2) of Algorithms 4.3.3 and 4.3.4 are the same, we consider the differences only with respect to the estimated tail dependence coefficients and tail index α.

In Table 4.3 we summarize the results, see also Figure 4.6. We observe that the new tail dependence estimator has smaller empirical variance than (4.10). This results immediately in improved estimation of the tail index α by (4.24). Furthermore, the histogram of the estimated α by Algorithm 4.3.4 is more centered around the true value, i.e. applying (4.26) in order to obtain confidence bounds is now possible. Besides, the implied (by the
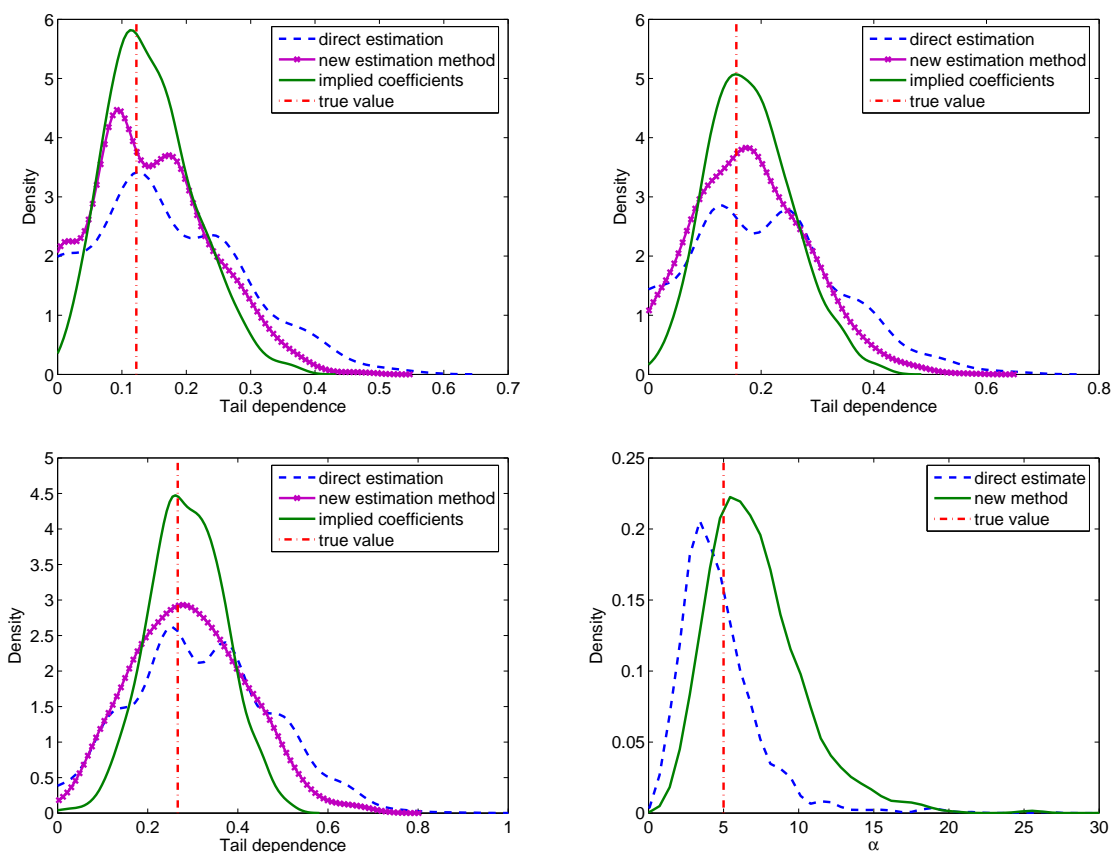
Figure 4.6: Upper row and bottom row, left (λ_{12}, λ_{13}, λ_{23}): The histograms of the direct estimates of tail dependence by (4.10), compared to the estimates by the new method (4.14) and to the implied (by the estimated Kendall's tau and tail index α by algorithm 4.3.4) estimates for tail dependence and to the true values. The direct estimates have the highest empirical variance. The new method (4.14) reduces the variance. The implied estimates improve further (4.14). Bottom row, right: the histograms of the estimates for the index α using the direct tail dependece estimates (4.10) as in Algorithm 4.3.3 and the new method (4.14) as in Algorithm 4.3.4. By the new method, the histogram is more centered around the true value.

	mean estimate	true value	m.s.e	std
α	4.9488	5	5.8749	2.4293
λ_{12} direct	0.1560	0.1224	0.0151	0.1213
λ_{12} new	0.1440	0.1224	0.0090	0.1022
λ_{12} implied	0.1454	0.1224	0.0052	0.0771
λ_{13} direct	0.1955	0.1559	0.0192	0.1357
λ_{13} new	0.1808	0.1559	0.0108	0.1268
λ_{13} implied	0.1817	0.1559	0.0063	0.0991
λ_{23} direct	0.3145	0.2666	0.0259	0.1536
λ_{23} new	0.2919	0.2666	0.0166	0.1352
λ_{23} implied	0.2874	0.2666	0.0074	0.1238

Table 4.3: Estimation of the t-copula with different marginals model from Example 4.5.1, sample size $n = 1000$. We observe that the new method (4.14) for estimation of the tail dependence coefficients improves the empirical variance and mean square error (m.s.e.), as compared to the direct estimator (4.10). This results immediately through Algorithm 4.3.4 in improved estimates of the tail index α (see Table 4.1 for comparison). Then the implied tail dependence coefficients estimates (step (5) of the Algorithm 4.3.4) have also quite satisfactory empirical variance and m.s.e.

estimated α and Kendall's tau matrix as in step (5) of Algorithm 4.3.4) tail dependence coefficients have also quite a low empirical variance.

In order to assess the accuracy of the method when applied to larger samples, we increase gradually the sample size to $n = 10000$ and at each step apply Algorithm 4.3.4 and compare it to Algorithm 4.3.3. In Figure 4.7 we observe that the new method reduces the empirical standard deviation of the estimator of the tail index α also for larger samples.

$\square$

In the next example we consider a real data sample, consisting of weekly log-returns of 8 German stock indices. In view of (4.33), this data represents a sample of i.i.d. copies of the market index vector $I = (I_1, \ldots, I_8)$ as in (4.33). We apply Algorithm 4.3.4 to estimate the parameters of copula of I.

Example 4.5.4. [German stock index data]
Our data consists of weekly log-returns of the stock indices: I_1 Automobiles (CXKAX), I_2 Banking (CXKBX), I_3 Chemicals (CXKCX), I_4 Construction (CXKOX), I_5 Insurance (CXPIX), I_6 Media (CXKDX), I_7 Software (CXKSX) and I_8 Transport (CXKTX). By standard time series analysis we conclude that the data are marginally i.i.d. More precisely, we analyse the autocorrelations of the values, the absolute values and the squared values in
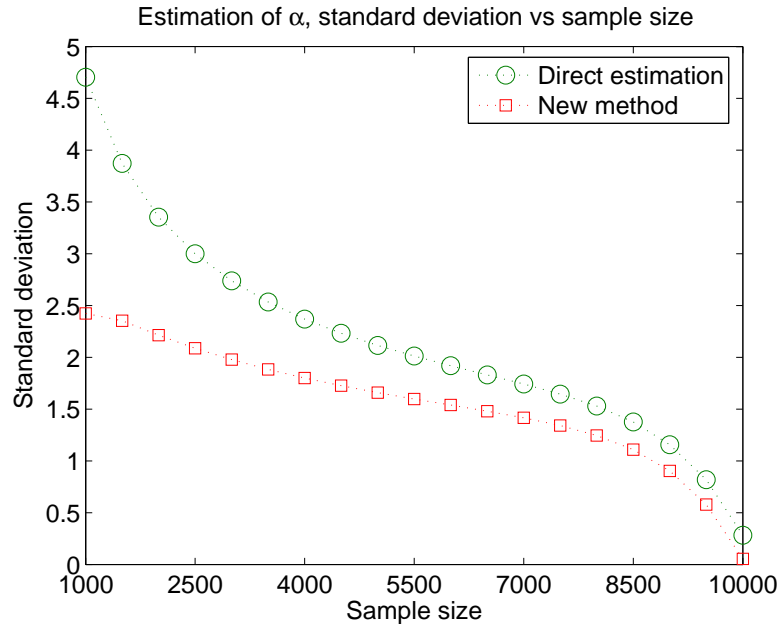
Figure 4.7: The empirical standard deviation of the estimator (4.24) for the tail index α, as a function of the sample size, using the direct tail dependence estimator (4.10) as in Algorithm 4.3.3 and the new method (4.14) as in Algorithm 4.3.4 for the model in Example 4.5.1. In all cases the new method provides lower empirical variance.

the marginal time-series, and check if they do not exceed the 0.95 confidence interval. We plot also the moving averages of the marginal time-series, and note that in all cases they are almost straight lines. Performing the Kolmogorov-Smirnov test for identical distribution over different subsamples at level 0.05, we confirm the i.i.d. hypothesis.

Plotting the bivariate marginals (see Figure 4.8) we detect that there is a significant dependence in the extremes. In order to quantify this dependence we apply the Algorithm 4.3.4.

In Table 4.4 we present results on the estimation of Kendall's tau matrix and the correlation matrix. The estimated correlations are positive, and the hypothesis for zero or negative correlation can be rejected with high significance. In Table 4.5 we present results on the estimation of the tail dependence coefficients. We obtain positive tail dependence estimates. However, due to the small sample size ($n = 300$), we cannot reject the hypothesis for tail-independence at confidence levels higher than 90%. Finally, we estimate the tail index $\widehat{\alpha} = 4.05$. By means of (4.26), Remark 4.3.2 and using the empirical covariance of the tail dependence estimates we obtain also a 90% confidence interval $\alpha \in [2.04, 6.02]$. Since α is a key parameter in our model, and the sample size is rather small, we provide also an alternative method to assess the accuracy of the estimation. More precisely, we apply a parametric bootstrap. We simulate $n = 300$ i.i.d. copies of an elliptical random vector $Y = (Y_1, \ldots, Y_8)$ with tail index $\widehat{\alpha}$ and correlation matrix $\widehat{\rho}$ as the estimated ones.

	CXKAX	CXKBX	CXKCX	CXKOX	CXPIX	CXKDX	CXKSX	CXKTX
I_1	τ_{ij}	0.47	0.44	0.26	0.36	0.18	0.34	0.44
	90%CI	[.39 .55]	[.34 .54]	[.11 .41]	[.24 .48]	[.04 .32]	[.22 .48]	[.34 .55]
	ρ_{ij}	0.67	0.64	0.40	0.54	0.28	0.52	0.64
	90%CI	[.57 .76]	[.51 .75]	[.18 .60]	[.37 .69]	[.06 .49]	[.34 .68]	[.50 .76]
I_2	0.47	τ_{ij}	0.46	0.27	0.51	0.23	0.38	0.48
	[.39 .55]	90%CI	[.36 .56]	[.11 .44]	[.41 .60]	[.10 .36]	[.26 .50]	[.39 .57]
	0.67	ρ_{ij}	0.66	0.42	0.72	0.35	0.56	0.68
	[.57 .76]	90%CI	[.53 .77]	[.17 .63]	[.60 .81]	[.15 .54]	[.39 .71]	[.57 .78]
I_3	0.44	0.46	τ_{ij}	0.24	0.43	0.17	0.28	0.47
	[.34 .54]	[.36 .56]	90%CI	[.08 .40]	[.32 .53]	[.03 .32]	[.12 .43]	[.38 .57]
	0.64	0.66	ρ_{ij}	0.37	0.62	0.27	0.42	0.69
	[.51 .75]	[.53 .77]	90%CI	[.13 .59]	[.49 .74]	[.05 .48]	[.19 .62]	[.56 .78]
I_4	0.26	0.27	0.24	τ_{ij}	0.28	0.17	0.17	0.30
	[.11 .41]	[.11 .44]	[.08 .40]	90%CI	[.13 .43]	[.02 .32]	[.00 .34]	[.14 .46]
	0.40	0.42	0.37	ρ_{ij}	0.43	0.27	0.26	0.46
	[.18 .60]	[.17 .63]	[.13 .59]	90%CI	[.20 .63]	[.03 .49]	[.00 .50]	[.22 .66]
I_5	0.36	0.51	0.43	0.28	τ_{ij}	0.21	0.35	0.42
	[.24 .48]	[.41 .60]	[.32 .53]	[.13 .43]	90%CI	[.08 .34]	[.22 .48]	[.29 .54]
	0.54	0.72	0.62	0.43	ρ_{ij}	0.33	0.52	0.61
	[.37 .69]	[.60 .81]	[.49 .74]	[.20 .63]	90%CI	[.13 .51]	[.34 .68]	[.45 .75]
I_6	0.18	0.23	0.17	0.17	0.21	τ_{ij}	0.31	0.27
	[.04 .32]	[.10 .36]	[.03 .32]	[.02 .32]	[.08 .34]	90%CI	[.17 .44]	[.15 .40]
	0.28	0.35	0.27	0.27	0.33	ρ_{ij}	0.46	0.42
	[.06 .49]	[.15 .54]	[.05 .48]	[.03 .49]	[.13 .51]	90%CI	[.27 .64]	[.23 .59]
I_7	0.34	0.38	0.28	0.17	0.35	0.31	τ_{ij}	0.40
	[.22 .48]	[.26 .50]	[.12 .43]	[.00 .34]	[.22 .48]	[.17 .44]	90%CI	[.28 0.53]
	0.52	0.56	0.42	0.26	0.52	0.46	ρ_{ij}	0.59
	[.34 .68]	[.39 .71]	[.19 .62]	[.00 .50]	[.34 .68]	[.27 .64]	90%CI	[.42 .74]
I_8	0.44	0.48	0.47	0.30	0.42	0.27	0.40	τ_{ij}
	[.34 .55]	[.39 .57]	[.38 .57]	[.14 .46]	[.29 .54]	[.15 .40]	[.28 .53]	90%CI
	0.64	0.68	0.69	0.46	0.61	0.42	0.59	ρ_{ij}
	[.50 .76]	[.57 .78]	[.56 .78]	[.22 .66]	[.45 .75]	[.23 .59]	[.42 .74]	90%CI

Table 4.4: Estimation of Kendall's tau and correlation for the German stock index data from Example 4.5.4
. In brackets are given the 90% confidence intervals, based on the empirical variance.

		CXKAX	CXKBX	CXKCX	CXKOX	CXPIX	CXKDX	CXKSX	CXKTX
I_1	λ_{ij}	λ_{ij}	.28	.16	.16	.13	.25	.27	.25
	90%UB	90%UB	.64	.36	.35	.29	.57	.60	.56
	λ_{ij}^I	λ_{ij}^I	.36	.34	.20	.27	.15	.26	.34
I_2		.28	λ_{ij}	.16	.46	.42	.16	.16	.40
		.64	90%UB	.36	.99	.95	.36	.36	.90
		.36	λ_{ij}^I	.35	.20	.40	.18	.28	.37
I_3		.16	.16	λ_{ij}	.16	.13	.16	.25	.25
		.36	.36	90%UB	.36	.29	.36	.58	.58
		.34	.35	λ_{ij}^I	.19	.32	.15	.20	.37
I_4		.16	.46	.16	λ_{ij}	.32	.16	.16	.43
		.35	.99	.36	90%UB	.72	.35	.37	.98
		.20	.20	.19	λ_{ij}^I	.21	.15	.15	.23
I_5		.13	.42	.13	.32	λ_{ij}	.12	.21	.57
		.29	.95	.29	.72	90%UB	.29	.48	.99
		.27	.40	.32	.21	λ_{ij}^I	.17	.26	.31
I_6		.25	.16	.16	.16	.12	λ_{ij}	.31	.31
		.57	.36	.36	.35	.29	90%UB	.71	.72
		.15	.18	.15	.15	.17	λ_{ij}^I	.23	.20
I_7		.27	.16	.25	.16	.21	.31	λ_{ij}	.47
		.60	.36	.58	.37	.48	.71	90%UB	.99
		.26	.28	.20	.15	.26	.23	λ_{ij}^I	.30
I_8		.25	.40	.25	.43	.57	.31	.47	λ_{ij}
		.56	.90	.58	.98	.99	.72	.99	90%UB
		.34	.37	.37	.23	.31	.20	.30	λ_{ij}^I

Table 4.5: Estimated tail dependence matrix and implied tail dependence matrix (step (5) of Algorithm 4.3.4) for the German stock indices from Example 4.5.4. In the second rows are given the 90% upper confidence bounds, based on the empirical variance.
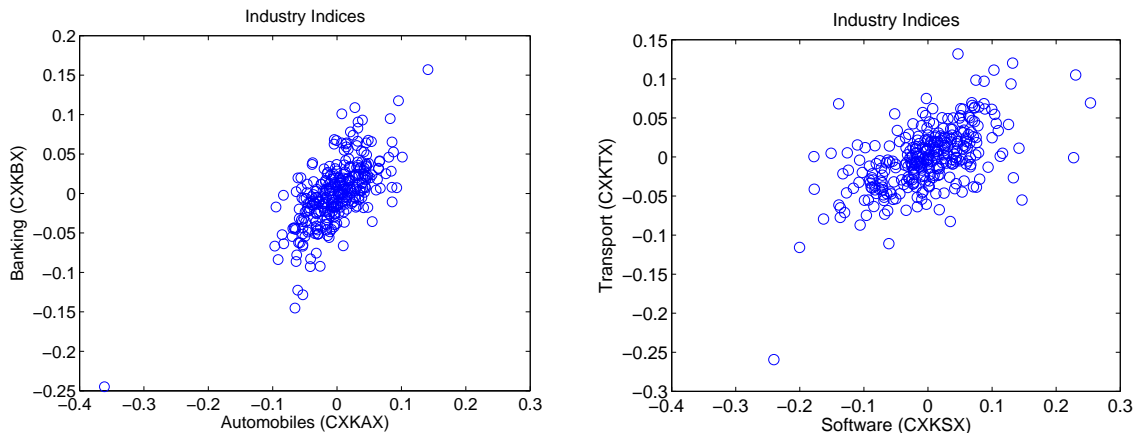
Figure 4.8: Log-returns of the indices Banking / Automobiles and Software / Transport. In both plots the point in the lower left quadrant represents the returns in the week of the 11/9'th terrorist attacks. However, even if we ignore this extreme point, we observe significant dependence in the extremely small values.

We apply Algorithm 4.3.4 and repeat the simulation 1000 times. In Figure 4.9 we compare the histogram of the resulting estimates of the tail index α to $\widehat{\alpha}$. We observe that in this way we may obtain narrower confidence bounds than using (4.26). In particular, the 90% empirical bootstrap confidence interval is given by $[2.41, 5.74]$ $\square$

We conclude this chapter with a study of the stock index data used in Daul et al. [30].

Example 4.5.5. [Monthly returns, worldwide]
We consider a sample of $n = 120$ log-returns of 92 stock indices from different industries and parts of the world. The sample comes from a period of 10 years (1992-2002). The indices are organized in 8 groups of different sizes according to the country they apply to.

By standard time-series analysis on the marginals as done in Example 4.5.4 we conclude that in almost all cases the hypothesis that the data are i.i.d. cannot be rejected with a high significance. Standard univariate analysis (chi-square goodness-of-fit test at level 0.05) shows also that in most cases the marginal distributions can be reasonably well modelled by the normal distribution. This is in fact not surprising as the period under consideration (1 month) is quite long and the 'aggregational Gaussianity' property which is typically observed in the returns of various financial assets becomes more pronounced, see e.g. Cont [24]. Similar conclusions for the same dataset have been made by Daul et al. [30] and Schwarz [131]. However, testing the bivariate marginals for tail dependence (the method of Ledford and Tawn [99]) indicates positive tail dependence, in particular for indices from the same country group. Therefore the multivariate normal distribution is not an appropriate model in this case. Instead we assume an elliptical copula model with normal marginals for each of the 8 country groups. Fix a country group with d indices

Country	Group size	ML $\widehat{\alpha}$	Tail dep. $\widehat{\alpha}$	90% conf. bounds
Australia	9	12	6	[3.7,7.6]
Canada	12	16	4.5	[3.9,6.7]
Germany	10	20	3.7	[3.0,5.5]
France	5	60	5.3	[3.2,11.3]
Japan	15	7	2.3	[1.7,2.5]
Switzerland	4	8	5.2	[2.6,8.1]
UK	15	14	3.9	[3.2,4.5]
USA	20	13	6.3	[5.2,7.1]

Table 4.6: Summary results from the estimation of the tail index α for the worldwide indices from Example 4.5.5. First column: the country of the indices. Second column: the number of indices in the country group. Third column: the estimated tail index α by the ML method (Example 4.5.5, step (5a)). Forth column: the estimated tail index α by the tail dependence method (Example 4.5.5, step (5b2)). Fifth column: the estimated 90% confidence bounds for α by the bootstrap method (Example 4.5.5, step (5b3)). The bounds are based on the empirical 0.1 and 0.9 quantiles of the simulated bootstrap copies.

within. We take the following steps.

(1) Estimate the marginal mean and variance vectors by their empirical counterparts.

(2) Estimate the Kendall's tau matrix by $\widehat{\tau}^n = [\widehat{\tau}^n_{lj}]_{l,j=1,...,d}$ as in (4.5).

(3) Estimate the correlation matrix by $[\widehat{\rho}^n_{lj}]_{l,j=1,...,d}$ using (4.6) and the Kendall's tau estimates $\widehat{\tau}^n$.

(4) Transfer the marginals to the copula scale using the normal probability integral transform and the parameters form step (1).

(5a) Assume in addition that we have the t-copula (Example 3.2.1) with α degrees of freedom as a model. Then estimate α by the ML method in (4.20) using the data from step (4), i.e. without putting more weight on the extremes.

(5b1) Using the data from step (4), estimate the lower tail dependence coefficients by $\widetilde{\Lambda}^n = [\widetilde{\lambda}^n_{lj}]_{l,j=1,...,d}$ as in (4.14).

(5b2) Estimate the tail index α by $\widehat{\alpha}^n$ as in (4.24) with the estimates for the Kendall's tau from step (2) and the tail dependence coefficients from step (5b1), i.e. using the information from the joint extremes.

(5bc) Estimate confidence bounds for the estimator from step (5b2) by the parametric bootstrap method described in Example 4.5.4.

The results are summarized in Table 4.6. We make the following observations.

(1) The tail dependence method from steps (5b) leads to lower estimates of the tail index α than the ML method form step (5a). Consequently, it is a conservative statistical method, taking care of the information in the joint extremes in the data, and in particular

in the extremes in the 'dangerous' direction (in our case these are the large in modulus negative log-returns).

(2) The higher the dimension of the random vector under consideration, the more accurate are the results of the tail dependence method, see the confidence bounds for the groups 'USA' and 'Japan' and compare to 'Switzerland' and 'France'. The price to pay is obviously the higher model risk.

(3) The higher the tail dependence in the data (the lower the α), the more accurate are the results of the tail dependence method, see the confidence bounds for the group 'Japan' and compare to 'Australia'. This is presumably due to the fact that whenever the tail dependence coefficients are high, they are estimated more accurately in step (5b1). On the contrary, for data with high correlation and low tail dependence, (4.24) overestimates the true coefficient.   □
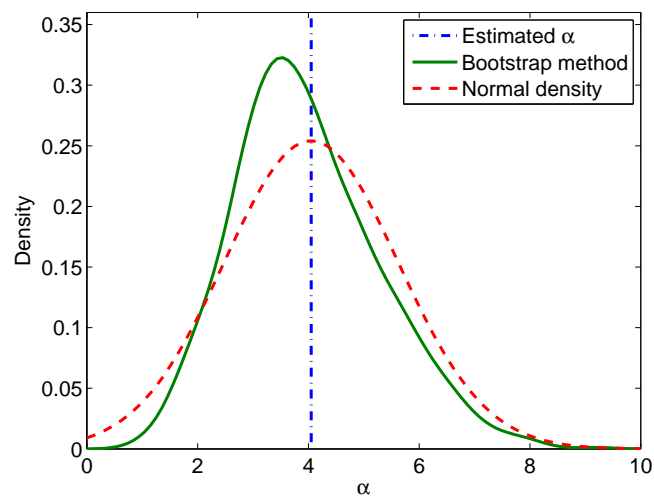
Figure 4.9: The bootstrap method applied to the German stock index data from Example 4.5.4. The estimated $\widehat{\alpha}$ compared to the histogram of the boostrap replications of $\widehat{\alpha}$ and to a normal density with mean $\widehat{\alpha} = 4.05$ and variance 2.47 equal to the variance of the estimator using the empirical covariance matrix of the tail dependence estimators and (4.26). We observe that the histogram of the bootstrap replications is more centered around $\widehat{\alpha}$ then the normal density. However, the boostrap distribution is skewed towards 0.

Chapter 5

Simulation

In this chapter we begin the analysis of the portfolio loss distribution in model (3.1) with (3.4) and (3.5). In Section 5.1 we present an algorithm for straightforward simulation of the loss L. As its application to practical problems is computationally intensive, in Section 5.2.2 we provide an improved algorithm, using a classical variance reduction technique – importance sampling. We pay particular attention to its application in the framework of heavy-tailed risk factors. Numerical support for our method is given in Section 5.3.1. Using Monte Carlo simulation, we illustrate the impact of the heavy-tailed model for the risk factors on the tail of the portfolio loss distribution in Section 5.3.2.

5.1 Direct Monte Carlo simulation

Monte Carlo simulation is among the most widely used computational tools in risk management. As in other application areas, it has the advantage of being very general and the disadvantage of being rather slow. In model (3.1) with (3.4) and (3.5) this procedure has essentially three steps.

Algorithm 5.1.1. Monte Carlo simulation of L.
    (1) Simulate the asset returns $Y = (Y_1, \ldots, Y_m)$.
    (2) Determine the ratings $X = (X_1, \ldots, X_m)$ by means of (3.4).
    (3) Given $X = (X_1, \ldots, X_m)$, simulate $(L_1, \ldots, L_m)$ and compute $L = \sum_{j=1}^{m} e_j L_j$.   □

    Simulation of $Y = (Y_1, \ldots, Y_m)$ is relatively easy, as this vector has a NVM distribution. We apply the following algorithm as in Embrechts et al. [47].

Algorithm 5.1.2. Monte Carlo simulation of a NVM distribution.
    (1) Find the Choleski decomposition A of the covariance matrix Σ_Y (i.e. $AA' = \Sigma_Y$).
    (2) Simulate m independent standard normals $Z = (Z_1, \ldots, Z_m)$.

(3) Simulate an independent W.

(4) Compute $Y = WBZ$. $\square$

However, several features of the credit risk setting pose a particular challenge to this approach.

(1) It requires accurate estimation of low-probability events, e.g. the probability of an extremely high loss. Hence, only a small part of the simulated data is used.

(2) Portfolios are multidimensional, i.e. m is large, and a large quantity of random numbers has to be drawn.

(3) Straightforward simulation of heavy-tailed r.v.s like $Y_1, \ldots, Y_m$ produces notoriously poor results, see Glasserman [69].

Therefore alternative approaches are needed.

5.2   Importance sampling (IS)

5.2.1   General IS algorithms

Importance sampling is a classical variance reduction technique. Consider for a moment a general problem for estimating $c = P(F(Y) \geq x) = E\left[1_{\{F(Y)>x\}}\right]$ for some random vector Y with density g and range $D \subseteq \mathbb{R}^p$ and some function F. The most simple way is to sample independent $Y^{(1)}, Y^{(2)}, \ldots, Y^{(N)}$ and take the average

$$\widehat{c}_{MC} = \frac{1}{N} \sum_{i=1}^{N} 1_{\left\{F(Y^{(i)}) \geq x\right\}}$$

as an estimate. However,

$$
\begin{aligned}
E\left[1_{\{F(Y) \geq x\}}\right] &= \int_D 1_{\{F(Y) \geq x\}} g(y) dy \\
&= \int_D 1_{\{F(Y) \geq x\}} g(y) \frac{f(y)}{f(y)} dy \\
&= E\left[1_{\left\{F(\widetilde{Y}) \geq x\right\}} \frac{g(\widetilde{Y})}{f(\widetilde{Y})}\right],
\end{aligned}
$$

where f is another density with the property $g(y) > 0 \Rightarrow f(y) > 0$, $y \in D$, and $\widetilde{Y}$ is a r.v. with density f. Then

$$\widehat{c}_{IS} = \frac{1}{N} \sum_{i=1}^{N} 1_{\left\{F(\widetilde{Y}^{(i)}) \geq x\right\}} \frac{g(\widetilde{Y}^{(i)})}{f(\widetilde{Y}^{(i)})},$$

where $\widetilde{Y}^{(1)}, \widetilde{Y}^{(2)}, \ldots \widetilde{Y}^{(N)}$ are i.i.d copies of $\widetilde{Y}$ is again an estimator for c. The quantity g/f is called *likelihood ratio*.

The goal in an importance sampling procedure is to select the density f such that the variance (or equivalently the second moment) of the estimator $\widehat{c}_{IS}$ is minimized. As shown for instance in Glasserman et al. [68], deriving the optimal (zero variance) density is equivalent to knowing the desired quantity c. Hence, some approximation of the optimal density is needed.

A classical method for selection of the density f is **exponential twisting**. Suppose that the components of $Y = (Y_1, \ldots, Y_p)$ are independent with densities $g_1, \ldots, g_p$, hence

$$g(y) = \prod_{l=1}^{p} g_l(y), \quad y \in D.$$

Then we select f_l, $l = 1, \ldots, p$, from the parametric family

$$f_l(y) = \frac{\exp(\mu_l y)}{E\left[\exp(\mu_l Y_l)\right]} g_l(y), \quad y \in D,$$

for some $\mu = (\mu_1, \ldots, \mu_p)$, provided that $E\left[\exp(\mu_l Y_l)\right]$ exists. The likelihood ratio becomes

$$\frac{g(y)}{f(y)} = \prod_{l=1}^{p} \frac{E\left[\exp(\mu_l Y_l)\right]}{\exp(\mu_l y)}, \quad y \in D.$$

The optimal $\widehat{\mu} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_p)$ minimizes the second moment of $\widehat{c}_{IS}$, i.e.

$$\widehat{\mu} = \arg\min_{\mu} E\left[1_{\{F(\widetilde{Y}) > x\}} \left(\prod_{l=1}^{p} \frac{E\left[\exp(\mu_l Y_l)\right]}{\exp(\mu_l \widetilde{Y}_l)} \right)^2 \right] = \arg\min_{\mu} E\left[\widehat{c}_{IS}^2\right], \qquad (5.1)$$

where $\widetilde{Y} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_p)$ has a density $f(y) = \prod_{l=1}^{p} f_l(y_l)$ Unfortunately, in most cases this problem is also infeasible (see Glasserman et al. [68]), and a further approximation is necessary. The typically used approximation is based on the classical Laplace method for integrals, see e.g. Dupois and Ellis [43], Chapter 3. Given that the function F is increasing componentwisely, one may replace (5.1) by

$$\widetilde{\mu} = \arg\min_{\mu} \left(\prod_{l=1}^{p} \frac{E\left[\exp(\mu_l Y_l)\right]}{\exp(\mu_l \widetilde{y}_l(\mu_l))} \right), \qquad (5.2)$$

where

$$\widetilde{y}(\mu) = (\widetilde{y}_1(\mu)), \ldots, \widetilde{y}_p(\mu)) = \arg\max_{y: F(y) = x} \sum_{l=1}^{p} \mu_l y_l,$$

see e.g. Glasserman et al. [68] for a detailed reasoning behind this method.

5.2.2 IS for portfolio credit risk

To motivate the approach we take in the credit risk model (3.1) with (3.4) and (3.5), observe that for every estimator $\widehat{c}$ of $P(L > x)$ and for every random vector ξ on the same probability space we have the variance decomposition

$$\text{var}[\widehat{c}] = E[\text{var}[\widehat{c}\,|\,\xi]] + \text{var}[E[\widehat{c}\,|\,\xi]]. \tag{5.3}$$

We try to minimize the two parts of (5.3) separately. To do this we need a proper random vector ξ. Bearing in mind that, given the global shock W and the common factors Z, the individual credits in the portfolio are independent (see Section 3.3), we select $\xi = (W, Z)$.

First we try to minimize the conditional (on W, Z) variance of the estimator. i.e. the first term in (5.3). Let $\theta > 0$ be a fixed constant. Let $X^\theta = \left(X_1^\theta, \ldots, X_m^\theta\right)$ be a random vector with discrete marginals with range $\{1, 2, \ldots, K\}$ defined on the conditional probability space $(\Omega, \mathcal{F}, P_{W,Z} = P(\cdot|W, Z))$. Let also $X_1^\theta, \ldots, X_m^\theta$ be independent with

$$P_{W,Z}\left(X_j^\theta = s\right) = \frac{g_{j,s}(W, Z)\varphi_{j,s}(e_j\theta)}{\sum_{k=1}^K g_{j,k}(W, Z)\varphi_{j,k}(e_j\theta)} = q_{j,s}(W, Z), \quad j = 1, \ldots, m, \, s = 1, \ldots, K \tag{5.4}$$

where $g_{j,k}$ and $\varphi_{j,k}$ are defined as in (3.11) and (3.12) respectively.

Consider the r.v.

$$L^\theta = \sum_{j=1}^m e_j L_j^\theta \tag{5.5}$$

where for $j = 1, \ldots, m$:

- e_j is the known positive constant from (3.1);
- L_j^θ is a real-valued r.v., defined on the conditional probability space $\left(\Omega, \mathcal{F}, P(\cdot|X_j^\theta)\right)$ and its density is given by

$$dP\left(L_j^\theta < x \,|\, X_j^\theta\right) = \frac{\exp(e_j\theta x)}{\varphi_{j,X_j}(e_j\theta)} dP\left(L_j < x \,|\, X_j = X_j^\theta\right), \tag{5.6}$$

where L_j is defined as in (3.1).

Under these assumptions, we have:

(A1) From Assumption (A) in (3.1), we have that L_j^θ are conditionally independent, given X^θ.

(B1) From Assumption (B) in (3.1), we have that given X_j^θ, L_j^θ is independent of X_s^θ for $s = 1, \ldots, m$, $s \neq j$.

(C1) From Assumption (C) in (3.1), we have that $C_j \leq L_j^\theta \leq 1$ a.s.

Lemma 5.2.1. *Under the assumptions of model (3.1) with (3.4) and (3.5), the density of the r.v. L^θ defined in (5.5) is given by:*

$$dP_{W,Z}\left(L^\theta < x\right) = \frac{\exp(\theta x)}{\exp(H(W,Z,\theta))}dP_{W,Z}\left(L < x\right),$$

where $H(W,Z,\theta)$ is defined in (3.8).

Proof. Denote $\widetilde{E}[\cdot] = E[\cdot \mid W, Z]$ and $\widetilde{\varphi}(\tau) = \widetilde{E}[\exp(\tau L^\theta)]$. We have

$$
\begin{aligned}
\widetilde{\varphi}(\tau) &= \widetilde{E}[E[\exp(\tau L^\theta) \mid X^\theta]] \\
&= \widetilde{E}[\prod_{j=1}^{m} E[\exp(\tau e_j L_j^\theta) \mid X_j^\theta]]
\end{aligned}
$$

because of (A1) and (B1). By (5.4) we get

$$
\begin{aligned}
\widetilde{\varphi}(\tau) &= \widetilde{E}[\prod_{j=1}^{m}\sum_{s=1}^{K} q_{j,s}(W,Z)E[\exp(\tau e_j L_j^\theta) \mid X_j^\theta = s]] \\
&= \widetilde{E}[\prod_{j=1}^{m}\sum_{s=1}^{K} q_{j,s}(W,Z)\frac{E[\exp((\tau+\theta)e_j L_j) \mid X_j = s]}{\varphi_{j,s}(e_j\theta)}]
\end{aligned}
$$

due to (5.6). Therefore

$$
\begin{aligned}
\widetilde{\varphi}(\tau) &= \widetilde{E}[\prod_{j=1}^{m}\sum_{s=1}^{K}\frac{g_{j,s}(W,Z)\varphi_{j,s}(e_j\theta)}{\sum_{k=1}^{K}g_{j,k}(W,Z)\varphi_{j,k}(e_j\theta)}\frac{\varphi_{j,s}(e_j(\theta+\tau))}{\varphi_{j,s}(e_j\theta)}] \\
&= \widetilde{E}[\prod_{j=1}^{m}\frac{\sum_{s=1}^{K}g_{j,s}(W,Z)\varphi_{j,s}(e_j(\theta+\tau))}{\sum_{k=1}^{K}g_{j,k}(W,Z)\varphi_{j,k}(e_j\theta)}] \\
&= \frac{\exp(H(W,Z,\theta+\tau))}{\exp(H(W,Z,\theta))}
\end{aligned}
$$

by the definition of H in (3.8). Therefore

$$
\begin{aligned}
\widetilde{E}[\exp(\tau L^\theta)] &= \int_{\mathbb{R}} \exp(\tau x)dP_{W,Z}(L^\theta < x) \\
&= \int_{\mathbb{R}} \exp(\tau x)\frac{\exp(\theta x)dP_{W,Z}(L < x)}{\exp(H(W,Z,\theta))}
\end{aligned}
$$

which proves the lemma.   □

By means of the above lemma,

$$P(L > x \mid W, Z) = \widetilde{E}[1_{\{L>x\}}] = \widetilde{E}[1_{\{L^\theta>x\}}\exp(H(W,Z,\theta) - \theta L^\theta)].$$

This result suggests the following simulation algorithm for estimation of $P(L > x)$:

Algorithm 5.2.2. (1) Simulate the global shock W and the common factors $Z = (Z_1, \ldots, Z_p)$.

(2a) If $E[L \,|\, W, Z] \geq x$, then the event $\{L > x\}$ is not rare. In this case we draw $(X_1^0, \ldots, X_m^0) = (X_1, \ldots, X_m)$ as in (5.4). Then we draw $(L_1^0, \ldots, L_m^0) = (L_1, \ldots, L_m)$ as in (5.6), compute $L^0 = L$ as (5.5) and return

$$1_{\{L^0 > x\}} = 1_{\{L > x\}}$$

as an estimator for $P(L > x)$.

(2b) If $E[L \,|\, W, Z] < x$, we select some $\theta = \theta(W, Z) > 0$ and we draw $(X_1^\theta, \ldots, X_m^\theta)$ as in (5.4). Then we draw $(L_1^\theta, \ldots, L_m^\theta)$ as in (5.6), compute L^θ as (5.5) and return

$$1_{\{L^\theta > x\}} \exp(H(W, Z, \theta) - \theta L^\theta)]$$

as an estimator for $P(L > x)$.

(3) Average over N independent random draws.

Step (1) of the above algorithm is trivial, as it requires a simulation of a one-dimensional r.v. W and the multivariate normal Z. We compute immediately

$$\widetilde{E}[L] = \sum_{j=1}^m e_j g_{j,k}(W, Z)(X_j = k)\mu_{j,k} \,,$$

where $\mu_{j,k}$ are as in Proposition 3.3.2 and $g_{j,k}$ are given in (3.11), $j = 1, \ldots, m$, $k = 1, \ldots, K$. Therefore we can directly determine whether we are in regime (2a) or (2b). Step (2a) is also straightforward. The hardest point in step (2b) is the selection of an appropriate $\theta = \theta(W, Z)$. Theoretically, we have to minimize the second moment of the estimator (in other words, the first summand in (5.3)), i.e. to choose

$$\widehat{\theta} = \arg\min_\theta \widetilde{E}[1_{\{L^\theta > x\}} \exp(2H(W, Z, \theta) - 2\theta L^\theta)] \,.$$

As this problem is infeasible, we use the bound

$$\widetilde{E}[1_{\{L^\theta > x\}} \exp(2H(W, Z, \theta) - 2\theta L^\theta)] \leq \exp(2H(W, Z, \theta) - 2\theta x) \,.$$

Instead of minimizing the integral, we minimize its upper bound and get

$$\widetilde{\theta} = \arg\min_\theta H(W, Z, \theta) - \theta x \,.$$

Various considerations lead to such a choice, see Glasserman [68] or Glasserman and Li [70] for a detailed discussion. In particular, with this choice we obtain from Lemma 5.2.1 that $\widetilde{E}[L^{\widetilde{\theta}}] = x$, which makes the event $\{L^{\widetilde{\theta}} > x\}$ common, rather than rare.

Algorithm 5.2.2 is intended to miminize only the first summand in (5.3). However, for high dimensional portfolios (when m is large), Proposition 6.1.1 suggests that in particular the global shock W and the common factors $Z = (Z_1, \ldots, Z_p)$ are important for the tail of the portfolio loss distribution. Therefore we need to minimize also the second term in the variance decomposition (5.3). To do that we need an approximation of $P(L > x)$ with an explicit function of W and Z. Proposition 6.1.1 suggests such an approximation:

$$P(L > x) \approx P(E[L|W, Z] > x).$$

Our goal is to apply the importance sampling techniques discussed in Section 5.2.1 to the probability $c = P(E[L|W, Z] > x)$. When attempting to apply directly an exponential change of measure as described in Section 5.2.1 to W and Z we face the following problems:
   (1) The global shock W is heavy-tailed, hence $E[\exp(\lambda W)]$ does not exist for $\lambda > 0$.
   (2) The common factors $Z = (Z_1, \ldots, Z_p) \in N_p(0, \Sigma)$ are not independent.
   However, we notice that, due to Proposition 3.3.2,

$$E[L|W, Z] = \sum_{j=1}^{m} \sum_{s=1}^{K} g_{j,s}(W, Z)\mu_{j,s} \stackrel{d}{=} \sum_{j=1}^{m} \sum_{s=1}^{K} \widetilde{g}_{j,s}(S, Q)\mu_{j,s}, \tag{5.7}$$

where $S \stackrel{d}{=} \frac{1}{W}$ has a well defined moment generating function, $Q = (Q_1, \ldots, Q_p)$ is a vector of independent standard normal r.v.s, such that for some matrix $B = [b_{l,d}]_{l,d=1,\ldots,p}$ we have $BQ \stackrel{d}{=} Z$ ($BB' = \Sigma$, e.g. the Choleski decomposition), and finally

$$\widetilde{g}_{j,s}(S, Q) = \Phi\left(\frac{G_j^{-1}(p_j^k)}{\sigma_i}S - \sum_{l=1}^{p}\beta_{j,l}Q_l\right) - \Phi\left(\frac{G_j^{-1}(p_j^{k-1})}{\sigma_i}S - \sum_{l=1}^{p}\beta_{j,l}Q_l\right)$$

with $\beta_{j,l} = \sum_{d=1}^{p}\frac{\alpha_{j,d}}{\sigma_j}b_{d,l}$. Hence, instead of applying an exponential change of measure to W, Z, we apply it to the transformed r.v.s. S, Q.
   We obtain the following algorithm.

Algorithm 5.2.3. (1) Select $\lambda > 0$ and $\mu = (\mu_1, \ldots, \mu_p) \in \mathbb{R}$.
   (2) Draw $\widetilde{S}$ and $\widetilde{Q} = (\widetilde{Q}_1, \ldots, \widetilde{Q}_p)$ with densities, resp.

$$f(y) = \frac{\exp(\lambda y)}{E[\exp(\lambda S)]}g(y) \text{ and } f_l(y) = \frac{\exp(\mu_l y)}{E[\exp(\mu_l Q_l)]}g_l(y), \ l = 1, \ldots, p, \tag{5.8}$$

where $g(y)$ is the density of S and $g_l(y)$, $l = 1, \ldots, p$ is the density of Q_l.
   (3) Compute $\widetilde{g}_{j,k}(\widetilde{S}, \widetilde{Q})$, $j = 1, \ldots, m$, $k = 1, \ldots, K$. If $\sum_{j=1}^{m} \sum_{s=1}^{K} \widetilde{g}_{j,s}(\widetilde{S}, \widetilde{Q})\mu_{j,s} \geq x$, repeat step (2a) of Algorithm 5.2.2. Otherwise repeat step (2b) of Algorithm 5.2.2.
   (4) Multiply the result from step (3) to the likelihood ratio

$$\frac{E[\exp(\lambda S)]}{\exp(\lambda \widetilde{S})} \prod_{l=1}^{p} \frac{E[\exp(\mu_l Q_l)]}{\exp(\mu_l \widetilde{Q}_l)}$$

to obtain an estimate for $P(L > x)$

(5) Average over N independent draws.

Step (2) of the above algorithm is comparatively easy to perform. In particular, by (5.8) $\widetilde{Q}$ is a vector of independent normals with mean μ. In the most frequently used in practice heavy-tailed t-model (see Example 3.2.1), $\widetilde{S}^2$ has a gamma d.f., see e.g. Glasserman et al. [69], who apply a similar techique to the simulation of a market risk model with heavy-tailed risk factors. For the selection of the design parameters λ and $\mu_1, \ldots, \mu_p$ in step (1) we apply the method in (5.2), i.e. we select

$$\widetilde{\mu} = \arg \min_{(\lambda, \mu)} \left(\frac{E\left[\exp(\lambda S)\right]}{\exp(\lambda \widetilde{q}_0(\lambda, \mu))} \prod_{l=1}^{p} \frac{E\left[\exp(\mu_l Q_l)\right]}{\exp(\mu_l \widetilde{q}_l(\lambda, \mu))} \right) ,$$

where $\widetilde{q}(\lambda, \mu) = (\widetilde{q}_0(\lambda, \mu), \ldots, \widetilde{q}_p(\lambda, \mu))$ satisfy

$$\widetilde{q}(\lambda, \mu)) = \arg \max_q (\lambda q_0 + \sum_{l=1}^{p} \mu_l q_l \mid E[L \mid S = q_0, Q = (q_1, \ldots, q_p)] = x) .$$

Due to (5.7) and (5.8) this choice guarantees, for instance, that

$$E[L \mid W = 1/E[\widetilde{S}], Z = BE[\widetilde{Q}]] = x ,$$

i.e. losses equal or exceeding x are no longer rare under the twisted probability measure.

5.3   Numerical examples

5.3.1   Comparison of the methods

We consider an example to demonstrate the suggested importance sampling method for a credit portfolio with heavy-tailed risk factors.

Example 5.3.1. [IS for 10 factors and global shock portfolio]
The parameters of the considered model (as in (3.1) with (3.4) and (3.5)) are as follows:
- $m = 100$ credits in the portfolio;
- the exposures e $= (e_1, \ldots, e_m)$ are generated uniformly on the interval $(1, 25)$;
- rating system with $K = 2$ ratings (default and non-default);
- the default probabilities $P(X_j = 1) = p_{j,1}, \ j = 1, \ldots, m$, are generated uniformly on the interval $[0.001, 0.02]$;
- the marginal loss distributions are given as $L_j = 1_{\{X_j=1\}}, \ j = 1, \ldots, m$.
For the dependence structure we use in (3.5) the t-model with $\nu = 4$ degrees of freedom, $p = 21$ common factors and $Z \in N_p(0, I)$. The factor loadings are given, for $j = 1, \ldots, m$,

by $\alpha_{j,1} = 0.7$, $\alpha_{j,l} = 0.3$, $l = j \bmod 10 + 1$ and $\alpha_{j,l} = 0.3$, $l = [j/10] + 11$ ($[x]$ denotes the closest integer smaller than x). Thus, each credit has a loading of 0.7 on the first factor and loadings 0.3 on two of the next. There are no equivalent credits with respect to the dependence structure. The first factor may be thought of as a global factor, the next ten as regional factors and the last ten as industry factors. This multifactor dependence structure is taken from Glasserman [67], where it is described as particularly hard to deal with.
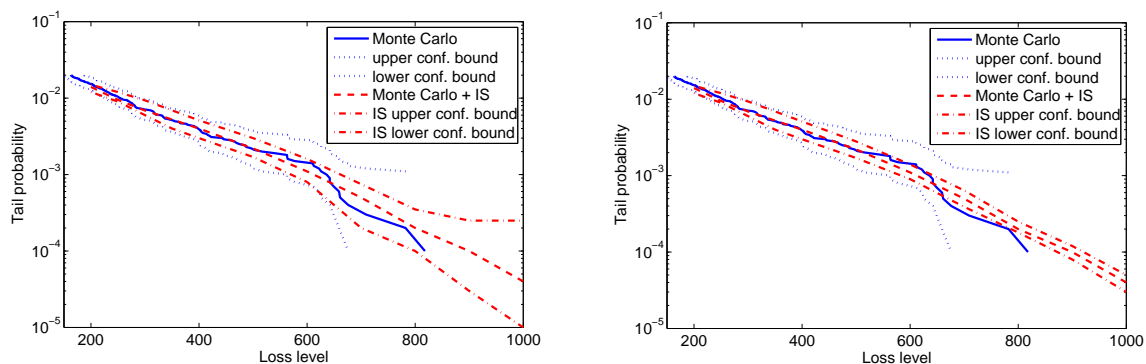
Figure 5.1: The parameters are the same as in Example 5.3.1.
Left figure: The tail of the portfolio loss distribution obtained by 10 000 Monte Carlo simulations of all random components of L compared to the tail obtained by Algorithm 5.2.2 with the same computational time budget, together with their respective 90% confidence bounds. The tightening of the confidence bounds, in particular at the extremely high loss levels (800-1000), is not sufficient enough.
Right figure: The tail of the portfolio loss distribution obtained by 10 000 Monte Carlo simulations of all random components of L compared to the tail obtained by Algorithm 5.2.3 with the same computational time budget, together with their respective 90% confidence bounds. The tightening of the confidence bounds is now quite satisfactory.

We first simulate $N = 10000$ i.i.d copies of the portfolio loss L using the standard Algorithm 5.1.1. As expected, this method produces rather broad confidence bounds at high loss levels. In order to analyse if there is a room for improvement, we fix the computational time required by Algorithm 5.1.1 and apply Algorithm 5.2.2, i.e. we use straightforward simulation of the global shock W and the common factors $Z_1, \ldots, Z_p$, combined with (conditional) IS for the marginal losses. In Figure 5.1 (left) we observe that in this way we obtain better confidence bounds, however, there is less improvement at extremely high loss levels. For this reason, with the same fixed computational time, we apply also Algorithm 5.2.3, i.e. we use combined IS for the global shock W, the common factors $Z_1, \ldots, Z_p$ and the marginal losses. In Figure 5.1 (right) we observe that in this way we may obtain narrow confidence bounds also at extremely high confidence levels. $\qquad \square$

5.3.2   Comparison of the models

From the examples and the analysis in Section 3.2, we expect that the heavy tailed risk factors Y in model (3.5) will have a significant impact on the tail of the credit portfolio loss distribution. We demonstrate this in the following example, see Frey et al. [62] or Schwarz [131] for more examples.

Example 5.3.2. [Example 3.2.4 continued, comparison of the impact on the tail]
We consider 6 different portfolios $L^{[1]}, \ldots, L^{[6]}$ with parameters, for each portfolio, as in model (as in (3.1) with (3.4) and (3.5)) as follows:
   - $m = 100$ credits in the portfolio;
   - the exposures $\mathrm{e} = (e_1, \ldots, e_m)$ are generated uniformly on the interval $(1, 25)$;
   - rating system with $K = 2$ ratings (default and non-default);
   - the default probabilities $P(X_j = 1) = p_{[k]}$, $j = 1, \ldots, m$, where $p_{[k]}$ is the default probability for the credits of group R_k, $k = 1, \ldots, 6$, as in Example 3.2.4;
   - the marginal loss distributions are given as $L_j = 1_{\{X_j=1\}}$, $j = 1, \ldots, m$.
   The dependence structure is as in Example 3.2.4. More precisely, we compare a one-factor CreditMetrics model ($W = 1$ a.s. in (3.5)) to a heavy-tailed t-model (Example 3.2.1) with $\nu = 4$ degrees of freedom and equal correlation structure.

   In Figure 5.2 we compare the tails for the 6 portfolios $L^{[1]}, \ldots, L^{[6]}$ under the Gaussian and under the heavy-tailed dependence assumptions. In all cases, the heavy-tailed model leads to higher risk as measured by, e.g. a high quantile of the loss distribution. It is particularly apparent for portfolios $L^{[1]}, L^{[2]}$ (good quality obligors), see the upper row of Figure 5.2 (recall our forecasts in Example 3.2.4).

   Finally we consider the portfolio $L = \sum_{k=1}^{6} L^{[k]}$ and its high α-quantile $\mathrm{VaR}_\alpha(L)$. It is often assumed in practice that $\sum_{k=1}^{6} \mathrm{VaR}_\alpha(L^{[k]})$ provides a conservative upper bound for $\mathrm{VaR}_\alpha(L)$, although this is theoretically wrong, see Embrechts et al. [45] for a detailed discussion on this issue. Our simple example demonstrates that the sum of the VaRs for the 6 portfolios under the normal model assumption is significanly smaller than the total portfolio VaR in the heavy-tailed case, see Figure 5.3. Clearly, it is hopeless to obtain a conservative upper bound on the portfolio risk using simplified models and theoretically unmotivated "shortcuts".   □
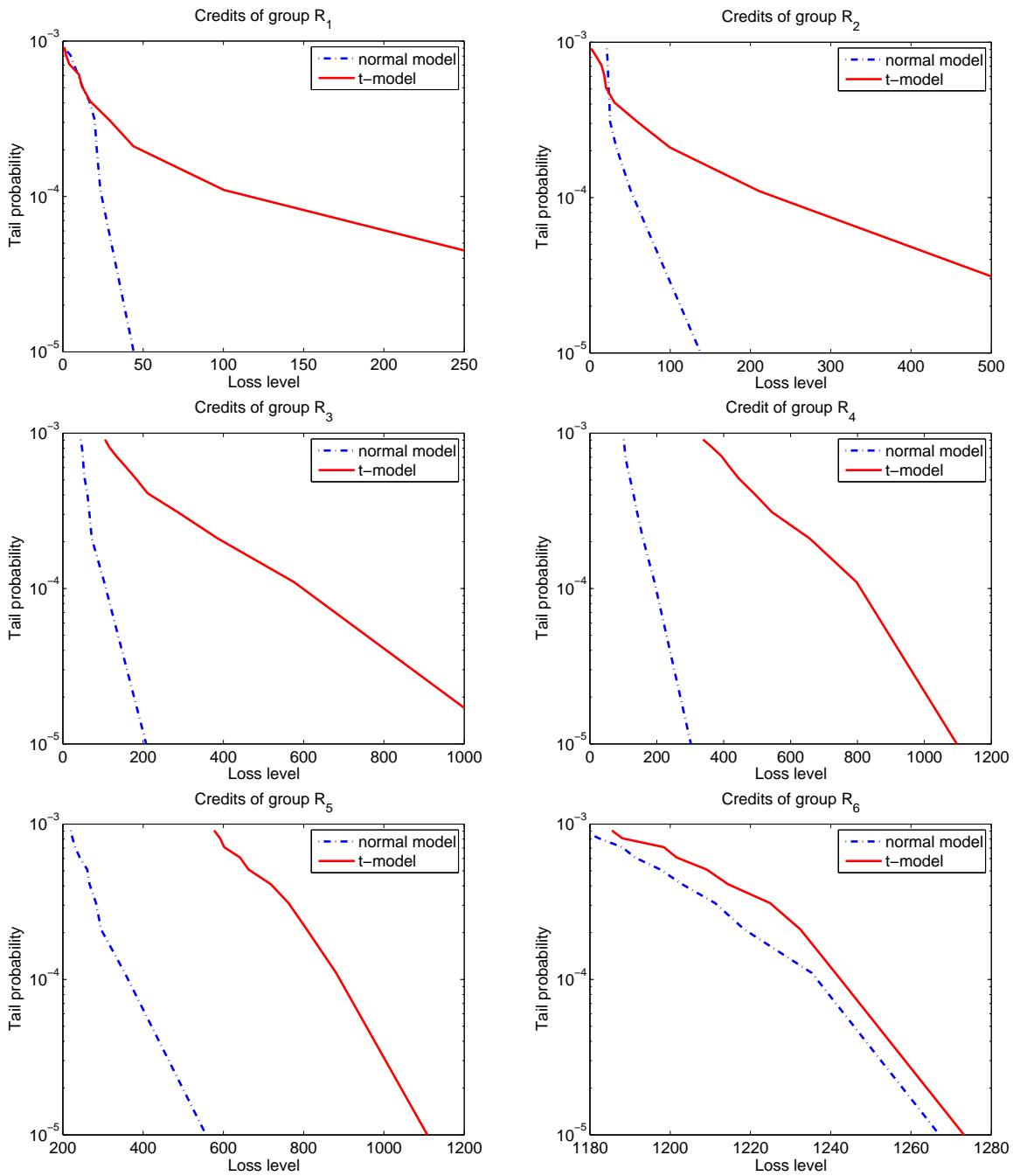
Figure 5.2: The parameters are the same as in Example 5.3.2.

The tails of the portfolio loss distributions of the good quality portfolios $L^{[1]}, L^{[2]}$ (upper row), the average quality portfolios $L^{[3]}, L^{[4]}$ (middle row) and the low quality portfolios $L^{[5]}, L^{[6]}$ under the Gaussian and the heavy-tailed t-model assumptions. In all cases, the heavy-tailed model leads to higher probabilities for large losses.
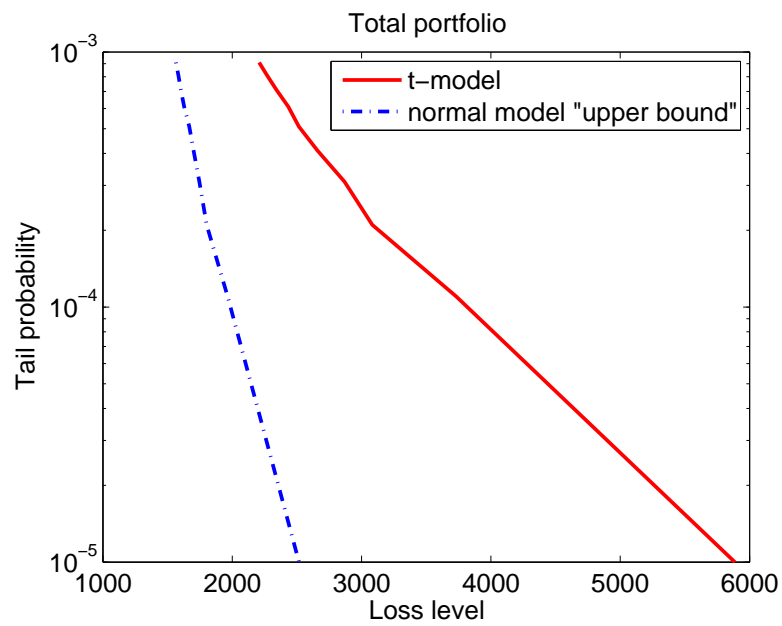
Figure 5.3: The parameters are the same as in Example 5.3.2.
The VaR$_\alpha$ of the aggregate portfolio L for different values of α. The t-model leads to higher VaR than the sum of the VaRs of the composite portfolios in the normal model.

Chapter 6

Tail approximation

For many applications in risk management, a fast evaluation of the tail of the credit portfolio loss distribution is needed. As already noted (see CreditMetrics [74]), it is not possible to obtain the distribution analytically even in the simplest Gaussian copula model. The Monte Carlo simulation methods described in Chapter 5, even when enhanced with the appropriate variance reduction techniques, are not the universal answer. In particular when the goal is to perform a sensitivity analysis for some of the model parameters, an analytic approximation is more suited.

Various approximations of the portfolio loss distribution have been developed. Most of them are based on the SLLN or on the CLT, see Section 6.1. The required granularity assumptions for these approximations (see e.g. Bluhm et al. [16], Assumption 2.5.2) are not always easy to verify in practice. Furthermore, such approximations are typically accurate at the center of the distribution (close to the mean) and could be misleading in the tail. In contrast, we consider an approximation based on the large deviations theory, which is designed to be accurate in the tail. More precisely, we apply the Markov's inequality to derive an upper bound of exponential type for the tail of L. In order to compute the best possible (the closest to the tail) upper bound, we develop a stochastic approximation algorithm and show its a.s. convergence to the optimal bound. This algorithm turns out to be a useful tool for risk management applications in Chapter 7. Before that we illustrate its accuracy and computational efficiency by a numerical example in Section 6.3.

6.1   Approximation for large portfolios

The approximations which are currently used in practice rely on two facts:

(1) Given a set of factors, the individual transactions in a credit portfolio are independent (see Gordy [71] for general credit risk models or Finger [52] for the CreditMetrics model in particular).

(2) The portfolio consists of a large number of comparatively homogeneous obligors.

More precisely, we consider the asymptotic distribution of the sequence of portfolios $L^{(m)}$ satisfying (3.1) with (3.4) and (3.5) for the number of obligors m going to infinity. We need the following additional 'granularity' assumptions on the obligors' exposures:

$$\lim_{m\to\infty} \sum_{j=1}^{m} e_j = \infty\,, \tag{6.1}$$

$$\lim_{m\to\infty} \sum_{j=1}^{m} \left(\frac{e_j}{\sum_{i=1}^{j} e_i} \right)^2 < \infty\,, \tag{6.2}$$

$$C_j \geq C,\, j \in \mathbb{N}\,, \tag{6.3}$$

where C_j are the constants from Assumption (C) in (3.1) and $C \in \mathbb{R}$ is a fixed constant.

Assumption (6.1) guarantees that the total portfolio exposure increases to infinity as the number of obligors increases to infinity. This holds, for instance, when $e_j \geq a,\, j \in \mathbb{N}$, for some constant $a > 0$. Assumption (6.2) implies that the exposure weights shrink rapidly with increasing number of obligors. This holds, for instance, when $e_j \in [a, b],\, j \in \mathbb{N}$ for some constants $0 < a \leq b < \infty$. Assumption (6.3) guarantees that no credit in the portfolio could lead to a profit larger than $|C|e_j,\, j \in \mathbb{N}$.

A less general version of the proposition below is for instance Bluhm et al. [16], Proposition 2.5.4. We prove it in our general setting.

Proposition 6.1.1. *Let $L^{(m)}$ be a sequence of rvs with distributions as in (3.1) with (3.4) and (3.5). Let also (6.1), (6.2) and (6.3) hold and denote by $\tau_m = \sum_{j=1}^{m} e_j$. Then*

$$\frac{1}{\tau_m}(L^{(m)} - E\left[L^{(m)} \,|\, W, Z\right]) \overset{\text{a.s.}}{\to} 0,\, m \to \infty. \tag{6.4}$$

Proof. We note first that, due to (3.5), given W and Z, the r.v.s $Y_j,\, j = 1, \ldots, m$, are independent (inherited by the independence of ϵ_j). Therefore $X_j,\, j = 1, \ldots, m$, are conditionally (on W and Z) independent by means of (3.4). Therefore, due to assumption (A) in (3.1), the r.v.s $\xi_j = e_j(L_j - E[L_j \,|\, W, Z])$ are (conditionally) independent. Furthermore, $E[\xi_j] = 0$ by definition. Finally, due to assumption (C) in (3.1) and (6.3), condition (6.2) leads to

$$\lim_{m\to\infty} \sum_{j=1}^{m} \frac{\widetilde{E}[\xi_j^2]}{(\tau_j)^2} \leq \lim_{m\to\infty} \sum_{j=1}^{m} \frac{4\max(1, |C|)^2 e_j^2}{(\tau_j)^2} < \infty\,,$$

where $\widetilde{E}$ denotes the expectation under the conditional on W and Z probability measure. Applying the SLLN we obtain

$$P(\lim_{m\to\infty} \frac{1}{\tau_m}(L^{(m)} - E[L^{(m)} \,|\, W, Z]) = 0 \,|\, W = w, Z = z) = 1\,,$$

for every $w > 0$ and $z \in \mathbb{R}^d$. Integrating w.r.t. W and Z we obtain the required result. $\quad\square$

6.2 Approximation by an upper bound

In order to obtain an upper bound of the portfolio loss distribution we apply Markov's inequality: for every $\theta \geq 0$ and $x \in [E[L], L_{max})$

$$P(L \geq x) \leq E[\exp(\theta(L - x))] = \varphi(\theta)\exp(-\theta x) =: F(\theta, x). \tag{6.5}$$

In the next lemma we summarize some of the important properties of $F(\theta, x)$.

Lemma 6.2.1. *Let $x \in (E[L], L_{max})$ be fixed. Then the function $F(\theta, x)$ satisfies the following properties.*

(1) $F(\theta, x)$ is analytic in θ;

(2) there exists a unique positive point

$$\widehat{\theta} = \widehat{\theta}(x) = \arg\min_\theta F(\theta, x), \tag{6.6}$$

which is the unique positive solution of the equation $\frac{\partial}{\partial\theta}F(\theta, x) = 0$;

(3) $F(\theta, x)$ is strictly decreasing for $\theta < \widehat{\theta}(x)$ and strictly increasing for $\theta > \widehat{\theta}(x)$;

(4) $F(\theta, x) \to \infty$, $\theta \to \infty$;

(5) $\widehat{\theta}(x) \in (0, \theta_{max}(x))$, where $\theta_{max}(x) < \infty$ satisfies $F(\theta_{max}(x), x) = 1$;

(6) there exist constants $D(x) \geq C(x) > 0$ such that for all θ, $0 < \theta < \infty$, $\theta \neq \widehat{\theta}$

$$C(x) \leq \frac{\frac{\partial}{\partial\theta}F(\theta, x)}{\theta - \widehat{\theta}} \leq D(x). \tag{6.7}$$

Proof. Properties (1), (2) and (3) are standard (see e.g. Jensen [82], section 1.2 and references therein).

As $x < L_{max}$, there exists $\epsilon > 0$ such that $x + \epsilon < L_{max}$ and hence $P(L > x + \epsilon) > 0$, therefore

$$\lim_{\theta\to\infty} F(\theta, x) \geq \lim_{\theta\to\infty} P(L > x + \epsilon)\exp(\theta\epsilon) = \infty,$$

i.e. we get (4).

By (2) we have $0 < \widehat{\theta}(x)$. Note that $F(0, x) = 1$ for every x, hence, due to (3), $F(\widehat{\theta}(x), x) < 1$. As $F(\theta, x)$ is continuous in θ, by means of (4) there exists a point $0 < \theta_{max}(x) < \infty$ which satisfies $F(\theta_{max}(x), x) = 1$. Because of (3), we get $\widehat{\theta}(x) < \theta_{max}(x)$, i.e. (5).

To prove (6) we use a Taylor expansion of $\frac{\partial}{\partial\theta}F(\theta, x)$ around $\widehat{\theta}$ and the fact that $\frac{\partial}{\partial\theta}F(\widehat{\theta}, x) = 0$; we get for $\epsilon = \epsilon(\theta) \in (\min(\widehat{\theta}, \theta), \max(\widehat{\theta}, \theta))$

$$\frac{\frac{\partial}{\partial\theta}F(\theta, x)}{\theta - \widehat{\theta}} = \frac{\partial^2}{\partial\theta^2}F(\epsilon, x).$$

For $\theta \in \mathbb{R}$ we have

$$
\begin{aligned}
\frac{\partial^2}{\partial \theta^2} F(\theta, x) &= \frac{\partial^2}{\partial \theta^2} E\left[\exp\left(\theta\left(L - x\right)\right)\right] \\
&= E\left[(L - x)^2 \exp\left(\theta\left(L - x\right)\right)\right] > 0.
\end{aligned}
$$

Also, by means of (3.7),

$$
\frac{\partial^2}{\partial \theta^2} F(\theta, x) \leq \left(\max(|L_{min}|, L_{max}) + |x|\right)^2 \exp\left(\theta\left(L_{max} - x\right)\right) < \infty.
$$

As by (1) $\frac{\partial^2}{\partial \theta^2} F(\theta, x)$ is continuous in θ, for ϵ in a compact subset of $\mathbb{R}$ the function $\frac{\partial^2}{\partial \theta^2} F(\epsilon, x)$ achieves a minimum and a maximum at some points $\widehat{\epsilon}_{min}$ and $\widehat{\epsilon}_{max}$, which are strictly positive, hence there exist positive constants $C(x)$ and $D(x)$ such that

$$
0 < C(x) \leq \frac{\partial^2}{\partial \theta^2} F(\widehat{\epsilon}_{min}, x) \leq \frac{\frac{\partial}{\partial \theta} F(\theta, x)}{\theta - \widehat{\theta}} \leq \frac{\partial^2}{\partial \theta^2} F(\widehat{\epsilon}_{max}, x) \leq D(x) < \infty.
$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

To derive a best upper bound of $P(L \geq x)$ we calculate the **saddlepoint** $\widehat{\theta}$ as defined in (6.6) and we obtain

$$
P(L \geq x) \leq F(\widehat{\theta}, x), \quad x \in (E[L], L_{max}) \tag{6.8}
$$

This classical large deviations technique has been successfully applied by Martin et al. [109] in the case of a one-factor Gaussian model ($p = 1$, $W = 1$ in (3.5)). Unfortunately in our case it is not possible to compute $\widehat{\theta}$ explicitly or by simple numerical methods, since the moment generating function $\varphi(\theta)$ is available only in terms of the $(p+1)$-dimensional integral (3.8). As a remedy we develop a Monte Carlo estimator for the saddlepoint $\widehat{\theta}$ and at the same time we obtain an estimator for the best upper bound $F(\widehat{\theta}, x)$ in (6.8).

The proposed method is in the framework of stochastic approximation algorithms (see Kushner and Yin [96]). More precisely, we approximate the saddlepoint $\widehat{\theta}$, for fixed $x \in (E[L], L_{max})$, by simulating recursively the r.v.s:

$$
\theta_{n+1} = \theta_n - a_n T_n, \ n \in \mathbb{N},
$$

where θ_1 is an arbitrary positive number,

$$
T_n = \frac{\partial}{\partial \theta} \exp\left(H\left(W^{(n)}, Z^{(n)}, \theta_n\right) - \theta_n x\right) \tag{6.9}
$$

and $W^{(n)}$ and $Z^{(n)}$ are i.i.d. copies of W and Z, respectively, $a_{n, n \in \mathbb{N}}$ is a sequence of positive constants such that

$$\sum_{n=1}^{\infty} a_n = \infty \tag{6.10}$$

$$\sum_{n=1}^{\infty} a_n^2 = A^2 < \infty \tag{6.11}$$

$$\lim_{n\to\infty} a_n = 0\,, \tag{6.12}$$

and $H(W, Z, \theta)$ is defined in (3.9).

In the next theorem we prove that (a modification of) $\theta_n \overset{\text{a.s.}}{\rightarrow} \widehat{\theta}$. The modification is taken in order to ensure 'stability' of the algorithm, i.e. to avoid θ_n growing to infinity for some $\omega \in \Omega$, see Kushner and Yin [96], Section 5.1 for details. Usually in stochastic approximation algorithms stability is achieved by assuring that each iterate θ_n belongs to some compact set which includes the true value $\widehat{\theta}$. In our case this compact set is given by $[0, \theta_{max}(x)]$, where $\theta_{max}(x)$ is the constant from Lemma 6.2.1 (5). Since the constant $\theta_{max}(x)$ is not explicitly available, we approximate it by a sequence of r.v.s converging a.s. to it.

Theorem 6.2.2. *Let $x \in (E[L], L_{max})$ be fixed. For $n \in \mathbb{N}$ let $T_n^{(i)}$, $i \in \mathbb{N}$, be i.i.d copies of the r.v. defined in (6.9). Let $a_{n,\, n\in\mathbb{N}}$ be a sequence of positive constants satisfying (6.10), (6.11) and (6.12). Define*

$$\theta_{n+1} = \min_i \left\{ \theta_{n+1}^{(i)} = \theta_n - a_n T_n^{(i)} \; : \; 0 \leq \theta_{n+1}^{(i)} \leq K_n \right\}, \tag{6.13}$$

where

$$K_n = \sup_{\theta} \left\{ \theta \; : \; F_n(\theta, x) \leq 1 \right\}$$

with $F_n(\theta, x)$ being the empirical counterpart of $F(\theta, x)$ as defined in (6.5)

$$F_n(\theta, x) = \frac{1}{n} \sum_{i=1}^{n} \exp \left(H\left(W^{(i)}, Z^{(i)}, \theta \right) - \theta x \right). \tag{6.14}$$

Then

$$\theta_n \overset{\text{a.s.}}{\rightarrow} \widehat{\theta}, \; n \to \infty.$$

Proof. Step 1: We prove that θ_n is finite a.s. for every $n \in \mathbb{N}$.

First we prove by induction that $\theta_n < \infty$ for any fixed n. We have $\theta_1 < \infty$. Assume that $\theta_n < \infty$. Note that

$$|T_n| \;\; = \;\; |\frac{\partial}{\partial \theta} H(W, Z, \theta_n) - x| \exp\left(H(W, Z, \theta_n) - \theta_n x\right).$$

However, since $\theta_n \geq 0$ by means of (6.13), $\exp\left(H\left(W, Z, \theta_n\right)\right) \leq \exp(\theta_n L_{max})$ because of (3.7); and also by (3.9)

$$
\begin{aligned}
|\frac{\partial}{\partial\theta}H\left(W, Z, \theta_n\right)| &\leq \sum_{j=1}^{m}|\frac{\partial}{\partial\theta}\log H_j\left(W, Z, \theta\right)| \\
&= \sum_{j=1}^{m}|\frac{e_j E\left[L_j \exp(\theta_n e_j L_j)\,|\,W, Z\right]}{E\left[\exp(\theta_n e_j L_j)\,|\,W, Z\right]}| \\
&\leq \sum_{j=1}^{m}\frac{e_j \max(1, |C_j|)\exp(\theta_n e_j)}{\exp(\theta_n e_j C_j)}
\end{aligned}
$$

by means of assumption (C) in (3.1). Therefore

$$
|T_n| \quad \leq \quad K(\theta_n)\,, \tag{6.15}
$$

where $K(\theta_n)$ is finite if θ_n is finite. Therefore, by means of (6.13) and (6.15), θ_{n+1} is also finite for a finite n.

Now assume that for some $\omega \in \Omega$ $\theta_n \to \infty$, $n \to \infty$. By the SLLN we have for $\theta \in \mathbb{R}$

$$
F_n\left(\theta, x\right) \overset{a.s.}{\to} E\left[\exp\left(\theta\left(L - x\right)\right)\right] = F\left(\theta, x\right)\,.
$$

Furthermore, $F\left(\theta, x\right)$ is continuous in θ and $F_n\left(\theta, x\right)$ is a.s. continuous in θ and, by means of Lemma 6.2.1 (4), $F\left(\theta, x\right) \to \infty$ when $\theta \to \infty$. So have for this ω and for some sufficiently large n that $F_n\left(\theta_n, x\right) > 1$. Such a value of θ_n is excluded by (6.13). Hence θ_n is finite a.s.

Step 2: We prove that

$$
(\theta_n - \widehat{\theta})^2 \overset{a.s.}{\to} \gamma,\ n \to \infty\,,
$$

where γ is some r.v. with finite mean.

Denote by $\mathcal{F}_n = \sigma\left(\theta_1, \ldots, \theta_n, W_{(1)}, Z_{(1)}, \ldots, W_{(n-1)}, Z_{(n-1)}\right)$, where $W_{(j)}, Z_{(j)}$ are the realizations for which $\theta_{j+1} = \theta_j - a_j T_j^{(i)}$. Denote $T_j = T_j^{(i)}$. From (6.13) we have $\theta_{n+1} - \widehat{\theta} = \theta_n - \widehat{\theta} - a_n T_n$, hence

$$
E\left[(\theta_{n+1} - \widehat{\theta})^2\,|\,\mathcal{F}_n\right] = (\theta_n - \widehat{\theta})^2 + a_n^2 E\left[T_n^2\,|\,\mathcal{F}_n\right] - 2a_n E\left[T_n(\theta_n - \widehat{\theta})\,|\,\mathcal{F}_n\right]\,. \tag{6.16}
$$

Consider first

$$
\begin{aligned}
E\left[T_n(\theta_n - \widehat{\theta})\,|\,\mathcal{F}_n\right] &= (\theta_n - \widehat{\theta})E\left[T_n\,|\,\mathcal{F}_n\right] \\
&= (\theta_n - \widehat{\theta})E\left[T_n\,|\,\theta_n\right] \\
&= (\theta_n - \widehat{\theta})\frac{\partial}{\partial\theta}E\left[\exp\left(H\left(W_{(n)}, Z_{(n)}, \theta_n\right) - \theta_n x\right)\,|\,\theta_n\right] \\
&= (\theta_n - \widehat{\theta})\frac{\partial}{\partial\theta}F\left(\theta_n, x\right)\,.
\end{aligned}
$$

Secondly,

$$
\begin{aligned}
\sum_{k=1}^{n} a_k^2 E\left[T_k^2 \mid \mathcal{F}_k\right] &= \sum_{k=1}^{n} a_k^2 E\left[T_k^2 \mid \theta_k\right] \\
&\leq \sum_{k=1}^{\infty} a_k^2 E\left[T_k^2 \mid \theta_k\right] \\
&\leq K^2 A^2,
\end{aligned}
\tag{6.17}
$$

where A^2 is the limiting constant from (6.11) and $K^2 < \infty$ is a constant, independent of θ_k, such that

$$
E\left[T_k^2 \mid \theta_k\right] \leq K^2
\tag{6.18}
$$

for every $k \in \mathbb{N}$ (the existence of such a constant follows from Step 1 and (6.15)).

Denote

$$
M_{n+1} = (\theta_{n+1} - \widehat{\theta})^2 + K^2 A^2 - \sum_{k=1}^{n} a_k^2 E\left[T_k^2 \mid \theta_k\right].
$$

From (6.17) we know that $M_n \geq 0$. On the other hand, using (6.16),

$$
\begin{aligned}
E\left[M_{n+1} \mid \mathcal{F}_n\right] &= E\left[(\theta_{n+1} - \widehat{\theta})^2 \mid \mathcal{F}_n\right] + K^2 A^2 - \sum_{k=1}^{n} a_k^2 E\left[E\left[T_k^2 \mid \theta_k\right] \mid \mathcal{F}_n\right] \\
&= M_n - 2 a_n E\left[T_n(\theta_n - \widehat{\theta}) \mid \mathcal{F}_n\right] \\
&= M_n - 2 a_n(\theta_n - \widehat{\theta}) \frac{\partial}{\partial \theta} F(\theta_n, x).
\end{aligned}
$$

By means of (6.7) $(\theta_n - \widehat{\theta}) \frac{\partial}{\partial \theta} F(\theta_n, x) > 0$ a.s., hence M_n is a non-negative supermartingale. By Doob's limit theorem

$$
M_n \xrightarrow{\text{a.s.}} M, \; n \to \infty,
\tag{6.19}
$$

where M is a r.v. with finite mean. As $K^2 A^2$ is some constant and $\sum_{k=1}^{n} a_k^2 E\left[T_k^2 \mid \theta_k\right]$ is an increasing, but bounded by means of (6.17) sequence, (6.19) implies that $(\theta_n - \widehat{\theta})^2 \xrightarrow{\text{a.s.}} \gamma$, where γ is a r.v. with finite mean.

Step 3. Denote $\eta_n = E(\theta_n - \widehat{\theta})^2$. We prove that there exists $n_1 \in \mathbb{N}$ such that for every $n > n_1$

$$
\eta_{n+1} \leq (1 - a_n C)^2 \eta_n + a_n^2 K^2,
\tag{6.20}
$$

where C is the constant from (6.7) and K is the constant from (6.18).

Denoting $\widetilde{T}_n = T_n - \frac{\partial}{\partial \theta} F(\theta_n, x)$ we obtain from (6.13)

$$
\begin{aligned}
\theta_{n+1} - \widehat{\theta} &= \theta_n - \widehat{\theta} - a_n \widetilde{T}_n - a_n \frac{\frac{\partial}{\partial \theta} F(\theta_n, x)}{\theta_n - \widehat{\theta}}(\theta_n - \widehat{\theta}) \\
&= \left(1 - a_n \frac{\frac{\partial}{\partial \theta} F(\theta_n, x)}{\theta_n - \widehat{\theta}}\right)(\theta_n - \widehat{\theta}) - a_n \widetilde{T}_n.
\end{aligned}
$$

Raising to second power and integrating we get

$$\eta_{n+1} \;\; \leq \;\; E\left[\left(1 - a_n\frac{\frac{\partial}{\partial\theta}F\left(\theta_n, x\right)}{\theta_n - \widehat{\theta}}\right)^2(\theta_n - \widehat{\theta})^2\right] - 2a_n E\left[\left(1 - a_n\frac{\frac{\partial}{\partial\theta}F\left(\theta_n, x\right)}{\theta_n - \widehat{\theta}}\right)(\theta_n - \widehat{\theta})\widetilde{T}_n\right]$$
$$+ a_n^2 E\widetilde{T}_n^2 \,.$$

Conditioning on $\mathcal{F}_n$ we have

$$
\begin{aligned}
E\left[\left(1 - a_n\frac{\frac{\partial}{\partial\theta}F(\theta_n,x)}{\theta_n-\widehat{\theta}}\right)(\theta_n - \widehat{\theta})\widetilde{T}_n\right] &= E\left[E\left[\left(1 - a_n\frac{\frac{\partial}{\partial\theta}F(\theta_n, x)}{\theta_n - \widehat{\theta}}\right)(\theta_n - \widehat{\theta})\widetilde{T}_n \mid \mathcal{F}_n\right]\right] \\
&= E\left[\left(1 - a_n\frac{\frac{\partial}{\partial\theta}F(\theta_n, x)}{\theta_n - \widehat{\theta}}\right)(\theta_n - \widehat{\theta})E\left[\widetilde{T}_n \mid \mathcal{F}_n\right]\right] \\
&= E\left[\left(1 - a_n\frac{\frac{\partial}{\partial\theta}F(\theta_n, x)}{\theta_n - \widehat{\theta}}\right)(\theta_n - \widehat{\theta})E\left[\widetilde{T}_n \mid \theta_n\right]\right] \\
&= 0 \,,
\end{aligned}
$$

because $E\left[\widetilde{T}_n \mid \theta_n\right] = E\left[T_n \mid \theta_n\right] - E\left[\frac{\partial}{\partial\theta}F(\theta_n, x) \mid \theta_n\right] = 0$ a.s..

Due to (6.7) we have

$$1 - a_n\frac{\frac{\partial}{\partial\theta}F\left(\theta_n, x\right)}{\theta_n - \widehat{\theta}} \leq 1 - a_n C.$$

Also, since a_n converges to 0 and $\frac{\frac{\partial}{\partial\theta}F(\theta_n,x)}{\theta_n-\widehat{\theta}} \geq D$ as in (6.7), we can always select an index n_1, such that $a_n < \frac{1}{D}$ for every $n > n_1$, and hence $1 - a_n\frac{\frac{\partial}{\partial\theta}F(\theta_n,x)}{\theta_n-\widehat{\theta}} > 0$. We get

$$\eta_{n+1} \;\; \leq (1 - a_n C)^2\,\eta_n + a_n^2 E\widetilde{T}_n^2 \,.$$

Since $E\widetilde{T}_n^2 = E\left(T_n - \frac{\partial}{\partial\theta}F(\theta_n, x)\right)^2 = \mathrm{var}[T_n] \leq ET_n^2 \leq K^2$ we obtain (6.20).

Step 4. Finally we prove that

$$\lim_{n\to\infty} \eta_n = 0. \tag{6.21}$$

Since a.s. convergence implies convergence in probability, by means of Step 2 we have

$$\lim_{n\to\infty} \eta_n = E\gamma < \infty.$$

On the other hand, since $\gamma = \lim_{n\to\infty}(\theta_n - \widehat{\theta})^2$ is a non-negative r.v., showing (6.21) is enough to prove the theorem.

We have by (6.20), for every sufficiently large $n > n_1$,

$$\sqrt{\eta_{n+1}} \;\; \leq (1 - a_n C)\sqrt{\eta_n} + a_n K \,.$$

We also have from Step 2 that $\lim_{n\to\infty}\sqrt{\eta_n}=\sqrt{E\gamma}$. Now assume that $\sqrt{E\gamma}>\frac{K}{C}$. This implies for some $\epsilon>0$ $\sqrt{\eta_n}\geq\frac{K}{C}+\epsilon$, for $n>n_2=n_2(\epsilon)$. Let $n_2>n_1$. We get

$$
\begin{aligned}
\sqrt{\eta_{n+1}} &\leq \sqrt{\eta_n}-a_nC(\frac{K}{C}+\epsilon)+a_nK\\
&=\sqrt{\eta_n}-a_nC\epsilon\,.
\end{aligned}
$$

Using the inequality recursively we get

$$
\sum_{j=n_2+1}^{n}\left(\sqrt{\eta_{j+1}}-\sqrt{\eta_j}\right)\leq -C\epsilon\sum_{j=n_2+1}^{n}a_j\,,
$$

which means

$$
\sqrt{\eta_{n+1}}\leq\sqrt{\eta_{n_2}}-C\epsilon\sum_{j=n_2+1}^{n}a_j\,.
$$

However, due to (6.10), there exists some index n_3, such that

$$
\frac{\sqrt{\eta_{n_2}}}{C\epsilon}<\sum_{j=n_1+1}^{n_3}a_j\,,
$$

hence $\sqrt{\eta_{n_3}}<0$, which is a contradiction. Therefore we obtain $E\gamma\leq\frac{K^2}{C^2}$. Hence there exists $n_4\in\mathbb{N}$ such that for $n>n_4$ $\eta_n\leq\frac{K^2}{C^2}$.

Going back to (6.20), we have for sufficiently large $n>\max{(n_1,n_4)}$

$$
\begin{aligned}
\eta_{n+1} &\leq(1-a_nC)^2\,\eta_n+a_n^2K^2\\
&=\eta_n+a_n^2C^2\eta_n-2a_nC\eta_n+a_n^2K^2\\
&\leq\eta_n+a_n^2C^2\frac{K^2}{C^2}-2a_nC\eta_n+a_n^2K^2\\
&=\eta_n-2a_nC\eta_n+2a_n^2K^2
\end{aligned}
$$

Assume the contrary to the hypothesis, that $\eta_n>\epsilon>0$ for every n larger than some fixed n_5. Then we get

$$
\eta_{n+1}\leq\eta_n-2a_nC\epsilon+2a_n^2K^2\,.
$$

Applying the inequality recursively we obtain

$$
\sum_{j=n_5+1}^{n}\left(\eta_{j+1}-\eta_j\right)\leq\sum_{j=n_5+1}^{n}2a_j^2K^2-2C\epsilon\sum_{j=n_5+1}^{n}a_j
$$

and therefore

$$
\begin{aligned}
\eta_{n+1} &\leq\eta_{n_5}+2K^2\sum_{j=n_5+1}^{n}a_j^2-2C\epsilon\sum_{j=n_5+1}^{n}a_j\\
&\leq\eta_{n_5}+2K^2\sum_{j=1}^{\infty}a_j^2-2C\epsilon\sum_{j=n_5+1}^{n}a_j\,.
\end{aligned}
$$

However, due to (6.10), there exists some index n_6, such that

$$\frac{\eta_{n_5} + 2K^2 A}{2C\epsilon} < \sum_{j=n_5+1}^{n_6} a_j,$$

hence $\eta_{n_6+1} < 0$, which is a contradiction. Therefore we obtain the required result. $\qquad\square$

Next we derive an approximation of the optimal upper bound $F(\widehat{\theta}, x)$ of $P(L \geq x)$ as defined in (6.8).

Proposition 6.2.3. *Let $x \in (E[L], L_{max})$ be fixed. For $\theta \in \mathbb{R}$ let $F_n(\theta, x)$ be defined as in (6.14). If $\theta_n \overset{a.s.}{\to} \widehat{\theta}$, $n \to \infty$, then*

$$F_n(\theta_n, x) \overset{a.s.}{\to} F(\widehat{\theta}, x), \ \ n \to \infty \tag{6.22}$$

Proof. By the SLLN we have

$$F_n(\theta, x) \overset{a.s.}{\to} E\left[\exp\left(\theta\left(L - x\right)\right)\right] = F(\theta, x)$$

for every $\theta \in \mathbb{R}$. As $\exp\left(H(W^{(i)}, Z^{(i)}, \theta) - \theta x\right)$ and $F(\theta, x)$ are a.s. continuous in θ and $\theta_n \overset{a.s.}{\to} \widehat{\theta}$, by the continuous mapping theorem we have

$$F_n(\theta_n, x) \overset{a.s.}{\to} F(\widehat{\theta}, x), \ \ n \to \infty.$$

$\qquad\square$

Remark 6.2.4. At each simulation step of (6.13), we check if $F_n(\theta_n, x) < 1$. Therefore, the upper bound approximation of $P(L \geq x)$ is available as by-product from the proposed algorithm. $\qquad\square$

6.3   Numerical examples

We give an example to demonstrate our method.

Example 6.3.1. [Example 5.3.1 continued, tail approximation]

We consider model (3.1) with (3.4) and (3.5) and the parameters from Example 5.3.1. In Figure 6.1 (left) we compare the tail of the portfolio loss distribution, obtained by Monte Carlo simulation as explained in Section 5.1 to the upper bound approximation as in (6.8), obtained by the new method (6.22). We observe that the upper bound approximation is quite accurate at high loss levels (i.e. 500-800), but degenerates quickly as we move in direction to the mean of the distribution. Furthermore, by the new approximation method
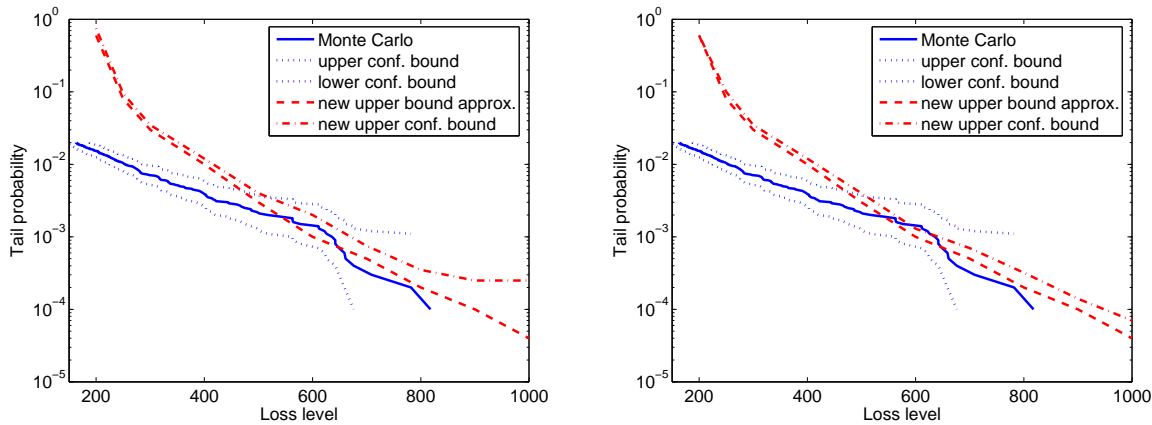
Figure 6.1: The parameters are the same as in Example 6.3.1.

Left figure: The tail of the portfolio loss distribution obtained by 10 000 Monte Carlo simulations of all random components of L compared to the upper bound approximation (6.8) obtained by the new method (6.22), together with their respective 90% confidence bounds. The upper bound approximation is accurate at high loss levels (i.e. 500-800), and stays within the 90% Monte Carlo confidence bounds. At extremely high loss levels (i.e. 800-1000), the new approximation method allows for extrapolation of estimates and confidence bounds beyond the range of the simulated data.

Right figure: The tail of the portfolio loss distribution obtained by 10 000 Monte Carlo simulations of all random components of L compared to the upper bound approximation (6.8) obtained by the new method (6.22), enhanced by importance sampling, together with their respective 90% confidence bounds. The new approximation method, enhanced by importance sampling, gives accurate information at loss levels where the standard Monte Carlo degenerates (i.e. 800-1000).

it is possible to obtain estimates and confidence bounds at extremely high loss levels (i.e. 800-1000), which are beyond the range of the simulated data in the Monte Carlo method. Note that the new method is computationally more efficient as it requires simulation only of the common factors Z and the global shock W, and not of all random components in the model.

We further improve the numerical performance of the proposed algorithm from Section 6.2 by applying importance sampling techniques in the simulation of the common factors Z and the global shock W. We use a classical variance reduction method, namely exponential change of measure, see Section 5.2.2. Note that in the framework of heavy-tailed risk factors, such a technique is not directly applicable, as $E[\exp(\theta W)] = \infty$ for every $\theta > 0$. Instead, we apply an exponential change of measure to the transformed r.v. $S = \frac{1}{W}$. With this enhanced method, the accuracy of the approximation at high loss levels can be further improved, as demonstrated in Figure 6.1 (right). $\qquad\square$

Chapter 7

Application to risk measurement

Today, virtually all major banks and financial institutions evaluate the risk-adjusted profitability within various business lines, including credits. The key concept behind this risk-adjusted performance measurement is described briefly in Section 7.1. The crucial point is to compute the contribution of the marginal risky position (e.g. the marginal credit) to the overall portfolio risk. Depending upon the choice of a risk measure, this could be a straightforward or a quite complicated task in the credit risk model under consideration, see Section 7.2.

The main result in this chapter is an application of the upper bound of the tail of the portfolio loss as derived in Section 6.2 to risk measurement. The application is given in Section 7.3. We define a new risk measure – tail bound VaR. We provide an algorithm to compute this risk measure and the contribution of the marginal credits to it. In Section 7.4.1 the risk contributions obtained by our method are compared to the Expected Shortfall (ES) contributions suggested for instance in Overbeck [122]. We find out that the two methods give similar results, however, our method has lower computational costs. This opens up a way to investigate the impact of the different model assumptions (Gaussian vs heavy-tailed) on the risk structure of the portfolio, as given by the marginal risk contributions.

7.1 Risk-adjusted performance measurement

Suppose an investor wants to place a fixed amount of capital into some asset. He has two exclusive choices: asset j, $j = 1, 2$, yields expected return m_j with risk r_j. Evidently, if the risks are equal, he will choose the asset with the higher yield. In the case of different risks (say $r_1 > r_2$) there is not such an obvious answer. No doubt that m_1 should be larger than m_2 for asset 2 to be eligible at all. But how large should be the difference in order to make asset 2 more attractive?

The Markowitz portfolio theory (see [108]) is the classical reference for a solution to this problem. It is still effective nowadays (see Becker et al. [11]). The investor should decide in favor of asset 1 if and only if $\frac{m_1}{r_1} \geq \frac{m_2}{r_2}$. This kind of computing is commonly called RORAC (return on risk-adjusted capital, see Matten [111]) and the comparison procedure is called RAPM (risk-adjusted performance measurement).

Note that the Markowitz's notion of risk is rather abstract. It is considered as a measure of the uncertainty in return, and is defined mathematically as the standard deviation of the return. Despite its computational convenience, this perspective has some drawbacks. From the risk management point of view, it is not desirable that not only the unfavorable, but also favorable fluctuations of the return around its mean have impact on the risk, as it happens with the standard deviation. The way to handle this problem is clear: one has to make use of down-sided risk measures, see Fishburn [54].

The down-sided risk measure Value-at-Risk (VaR) has nowadays become an industry standard (see BIS [9]). VaR is defined as a (small) quantile of the return distribution. In the case when the distributions of the returns are normal, the VaR method and the Markowitz theory yield identical results. However, the normal distribution assumption seems to be completely wrong for credit portfolios, see CreditMetrics [74]. Nonetheless, for the sake of RAPM, CreditMetrics [74] use the standard deviation as a risk measure.

In view of model (3.1), a portfolio is represented by the vector $\mathrm{e} = (e_1, \ldots, e_m) \in \mathbb{R}_+^m$ of the exposures of the individual credits. Further, a risk measure is simply a differentiable function $r = r(\mathrm{e}) : \mathbb{R}_+^m \to \mathbb{R}$. We present some examples in the next section. Following Tasche [134], we define marginal risk contributions as the gradient of the risk measure, i.e. $\mathrm{A} = (A_1, \ldots, A_m) = \nabla r(\mathrm{e})$. The goal of this chapter is to provide a computationally feasible alternative to the use of the standard deviation as a risk measure. More precisely, we are interested in computing the marginal risk contributions w.r.t. different risk measures.

Remark 7.1.1. Note that in the original CreditMetrics model, the risk contribution is defined differently, namely $A_j = r(\mathrm{e}) - r(\mathrm{e}^{(j)})$, where $\mathrm{e}^{(j)}$ is the vector of exposures e with 0 in its j-th component. For the drawbacks of this approach see Tasche [134].   □

7.2   Risk contributions w.r.t. standard risk measures

A classical example for risk measure is the standard deviation, i.e.

$$r_{\mathrm{var}}(\mathrm{e}) = \sqrt{\mathrm{var}(L(\mathrm{e}))},$$

where L is the portfolio loss defined in (3.1). An expression for $r_{\mathrm{var}}(\mathrm{e})$ is available in (3.14). Note that, once $r_{\mathrm{var}}(\mathrm{e})$ is computed for some $\mathrm{e} \in \mathbb{R}_+^m$, computing $r_{\mathrm{var}}(\mathrm{e}')$, $\mathrm{e} \neq \mathrm{e}' \in \mathbb{R}_+^m$

is trivial. Furthermore, by Example 5.1. in Tasche [134], we have for the marginal risk contribution of credit j, $j = 1, \ldots, m$

$$A_j^{\mathrm{var}} = \frac{\mathrm{cov}(L_j, L)}{\sqrt{\mathrm{var}(L)}} \, .$$

Furthermore,

$$\mathrm{cov}(L_j, L) = \sum_{l=1}^{m} e_l \mathrm{cov}(L_j, L_l) \, ,$$

and therefore this quantity is available as a by-product from the computation of $\sqrt{\mathrm{var}(L)}$. This makes the covariance based risk contributions particularly attractive for high-dimensional portfolios.

Another frequently used in practice risk measure is the VaR, defined as

$$\mathrm{VaR}(\mathrm{e}) = \inf_{x} \{ P(L(\mathrm{e}) > x) \leq \alpha \} \, , \tag{7.1}$$

for some small probability α. Note that, if L is normally distributed, then $\mathrm{VaR}(\mathrm{e}) = \Phi^{-1}(1 - \alpha) \sqrt{\mathrm{var}(L)}$, i.e. in such cases this quantity is available explicitly through (3.14). Unfortunately in model (3.1) with (3.4) and (3.5) the d.f. of L is far from being normal. Therefore, $\mathrm{VaR}(\mathrm{e})$ can be obtained only by means of Monte Carlo simulation, see CreditMetrics [74].

As shown in Tasche [134], under some regularity conditions, the marginal risk contributions for VaR are

$$A_j^{\mathrm{VaR}} = E \left[L_j \,|\, L = \mathrm{VaR} \right], \; j = 1, \ldots, m \, . \tag{7.2}$$

It is extremely hard to compute these quantities even by Monte Carlo methods, see Bluhm et. al. [16], Section 5.2.2.

An alternative risk measure for credit risk portfolios is used for instance in Overbeck [122]. Denoting the expected shortfall

$$\mathrm{ES}\,(\mathrm{e}) = E \left[L(\mathrm{e}) \,|\, L(\mathrm{e}) \geq \mathrm{VaR}\,(\mathrm{e}) \right] \, ,$$

one obtains for the marginal risk contributions

$$A_j^{\mathrm{ES}} = E \left[L_j \,|\, L \geq \mathrm{VaR} \right], \; j = 1, \ldots, m \, , \tag{7.3}$$

see Tasche [134]. This measure has many advantages in the credit risk framework, see e.g. Frey and McNeil [60] or Bluhm et al [16], Section 5.2.3. Furthermore, it is possible to compute marginal risk contributions A_j^{ES} by Monte Carlo methods. More precisely, let $L^{(i)}$, $i = 1, \ldots, n$, (and resp. $L_j^{(i)}$, $j = 1, \ldots, m$,) be i.i.d copies of L (and resp. of

L_j, $j = 1, \ldots, m$,) as in (3.1). Denote by $B_n = \left\{ i, \ i = 1, \ldots, n : L^{(i)} \geq L^{\lceil \alpha n \rceil} \right\}$, where $L^{\lceil \alpha n \rceil}$ is the $\lceil \alpha n \rceil$th largest order statistic among $L^{(i)}$, $i = 1, \ldots, n$. Then the SLLN provides estimates for (7.3), i.e.

$$\frac{1}{\# B_n} \sum_{i \in B_n} L_j^{(i)} \overset{\text{a.s.}}{\to} A_j^{\text{ES}}, \ n \to \infty .$$

7.3   Risk contributions w.r.t. tail bound VaR

In this section we explain how we use the upper bound approximation of the portfolio loss derived in Section 6.2 to approximate the VaR and the marginal risk contributions w.r.t. it. First we analyze the optimal upper bound $F(\widehat{\theta}(x), x)$ defined in (6.8), i.e.

$$F(\widehat{\theta}(x), x) = \varphi(\widehat{\theta}(x)) \exp(-\widehat{\theta}(x)x) ,$$

where $\varphi(\theta)$ is the moment generating function of L as in (3.8) and $\widehat{\theta}(x)$ is the saddlepoint defined in (6.6).

Lemma 7.3.1. *Let $F(\widehat{\theta}(x), x)$, $x \in (E[L], L_{max})$, be the function defined in (6.8). Then*
   (1) $\widehat{\theta}(x)$ is continuous and strictly increasing in x;
   (2) $F(\widehat{\theta}(x), x)$ is continuous and strictly decreasing in x;
   (3) the inverse function

$$\widehat{\text{VaR}}(\alpha) = \arg_x \left\{ F(\widehat{\theta}(x), x) = \alpha \right\} , \ \ \alpha \in (0, 1) \tag{7.4}$$

is a well defined and strictly decreasing function;
   (4) for every $\alpha \in (0, 1)$

$$\text{VaR}(\alpha) \leq \widehat{\text{VaR}}(\alpha) . \tag{7.5}$$

Proof. To prove (1), taking into account that $\frac{\partial}{\partial \theta} \log(\varphi(\widehat{\theta}(x))) = x$ (Lemma 6.2.1 (2)), we get

$$\frac{\partial}{\partial x} \frac{\partial}{\partial \theta} \log(\varphi(\widehat{\theta}(x))) = 1 ,$$

therefore $\frac{\partial}{\partial x} \widehat{\theta}(x) \frac{\partial^2}{\partial \theta^2} \log(\varphi(\widehat{\theta}(x))) = 1$. Hence $\frac{\partial}{\partial x} \widehat{\theta}(x) = (\frac{\partial^2}{\partial \theta^2} \log(\varphi(\widehat{\theta}(x))))^{-1} > 0$ by the strict convexity of $\varphi(\theta)$.

To prove (2) we note that $\log(F(\widehat{\theta}(x), x)) = \log(\varphi(\widehat{\theta}(x))) - \widehat{\theta}(x) x$. By differentiation we get

$$\begin{aligned}
\frac{\partial}{\partial x} \log(F(\widehat{\theta}(x), x)) &= \frac{\frac{\partial}{\partial \theta} \varphi(\widehat{\theta}(x))}{\varphi(\widehat{\theta}(x))} \frac{\partial}{\partial x} \widehat{\theta}(x) - x \frac{\partial}{\partial x} \widehat{\theta}(x) - \widehat{\theta}(x) \\
&= \frac{\partial}{\partial x} \widehat{\theta}(x) \left(\frac{\frac{\partial}{\partial \theta} \varphi(\widehat{\theta}(x))}{\varphi(\widehat{\theta}(x))} - x \right) - \widehat{\theta}(x) \\
&= -\widehat{\theta}(x) < 0
\end{aligned}$$

for every $x \in (E[L], L_{max})$, hence $F(\widehat{\theta}(x), x)$ is strictly decreasing.

Property (2) implies also the existence and the strict monotonicity of the inverse as in (7.4), i.e. (3) holds.

To prove (4) assume the contrary, i.e. that $\text{VaR}(\alpha) > \widehat{\text{VaR}}(\alpha)$. Then we have

$$
\begin{aligned}
P(L \geq \text{VaR}(\alpha)) & \leq P(L \geq \widehat{\text{VaR}}(\alpha)) \\
& \leq F(\widehat{\theta}(\widehat{\text{VaR}}(\alpha)), \widehat{\text{VaR}}(\alpha)) = \alpha,
\end{aligned}
$$

which is a contradiction to the definition of $\text{VaR}(\alpha)$ in (7.1). $\qquad \square$

Using the algorithm described in Section 6.2, we compute the optimal upper bound $F(\widehat{\theta}(x), x)$ for a sufficiently large number of points $x \in (E[L], L_{max})$ and we find the upper bound approximation of $\text{VaR}(\alpha)$, denoted by $\widehat{\text{VaR}}(\alpha)$ as in (7.4). We call the function $\widehat{\text{VaR}}$ the *tail bound VaR*. Normally we expect that $\widehat{\text{VaR}}$ is close to VaR as the upper bound derived in (6.8) is close to the tail of the portfolio loss $P(L > x)$ for large loss levels x.

We consider also the marginal risk contributions with respect to $\widehat{\text{VaR}}(\alpha)$. We fix α and define marginal risk contributions as

$$
\widehat{A}_j = \frac{\partial}{\partial e_j} \widehat{\text{VaR}}(\alpha; e) \ , \ j = 1, \ldots, m, \tag{7.6}
$$

where $e = (e_1, \ldots, e_m) \in \mathbb{R}^m_+$ are the exposures of the individual credits as in (3.1). Note that, for fixed $\alpha \in (0, 1)$, $\widehat{\text{VaR}}(\alpha; e)$ is well defined for every $e \in \mathbb{R}^m_+$. This can be seen from the following arguments.

Since the distribution of L in (3.1) depends on e, its moment generating function depends on e, i.e. we have $\varphi(\theta) = \varphi(\theta; e)$. Due to (3.8) and (3.9) we have also that $\varphi(\theta; e) = E[\exp(H(W, Z, \theta, e))]$ is well defined and

$$
H(W, Z, \theta, e) = \sum_{j=1}^{m} \log H_j(W, Z, \theta, e_j) \tag{7.7}
$$

with H_j, $j = 1, \ldots, m$, defined in (3.10). Hence, for $x \in (E[L], L_{max})$ we have also that the saddlepoint $\widehat{\theta}(x)$ defined in (6.6) is a function of e, i.e $\widehat{\theta}(x) = \widehat{\theta}(x; e)$. Hence the upper bound $F(\widehat{\theta}(x), x) = F(\widehat{\theta}(x; e), x; e)$ as in (6.8) is well defined. Therefore, due to lemma 7.3.1 (3), $\widehat{\text{VaR}}(\alpha; e)$ is well defined for every $e \in \mathbb{R}^m_+$.

Proposition 7.3.2. *With the above notations, for $\alpha \in (0, 1)$, the upper bound marginal risk contributions $(\widehat{A}_1, \ldots \widehat{A}_m)$ defined in (7.6) are given by*

$$
\widehat{A}_j = E \left[\frac{\exp\left(H(W, Z, \theta, e) - \theta \widehat{\text{VaR}}(\alpha; e)\right)}{\alpha e_j} \frac{\partial}{\partial \theta} \log H_j(W, Z, \theta, e_j) \right]_{\big| \theta = \widehat{\theta}(\widehat{\text{VaR}}(\alpha; e); e)} . \tag{7.8}
$$

Proof. By the definition of $\widehat{\text{VaR}}(\alpha;e)$ as in (7.4) we have

$$F\left(\widehat{\theta}(\widehat{\text{VaR}}\,(\alpha;e)\,;e),\widehat{\text{VaR}}\,(\alpha;e)\,;e\right)=\alpha\,, \tag{7.9}$$

hence

$$\frac{\partial}{\partial e_j}F\left(\widehat{\theta}(\widehat{\text{VaR}}\,(\alpha;e)\,;e),\widehat{\text{VaR}}\,(\alpha;e)\,;e\right)=0\,.$$

Therefore, setting $\widehat{\theta}=\widehat{\theta}(\widehat{\text{VaR}}\,(\alpha;e)\,;e)$ we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}F\left(\theta,\widehat{\text{VaR}}\,(\alpha;e)\,;e\right)_{|\theta=\widehat{\theta}}\frac{\partial}{\partial e_j}\widehat{\theta}(\widehat{\text{VaR}}\,(\alpha;e)\,;e)\\
&\quad+\left[(\frac{\partial}{\partial e_j}\varphi(\theta;e))\exp(-\theta\widehat{\text{VaR}}(\alpha;e))\right]_{|\theta=\widehat{\theta}}\\
&\quad-\left[\varphi(\theta;e)\theta(\frac{\partial}{\partial e_j}\widehat{\text{VaR}}(\alpha;e))\exp(-\theta\widehat{\text{VaR}}(\alpha;e))\right]_{|\theta=\widehat{\theta}}\,,
\end{aligned}$$

where the last two summands come from the fact that

$$F(\theta,\widehat{\text{VaR}}\,(\alpha;e)\,;e)=\varphi(\theta;e)\exp(-\theta\widehat{\text{VaR}}(\alpha;e))\,.$$

Since by (6.6) $\widehat{\theta}(x;e)$ is the point at which the minimum of $F(\theta,x;e)$ with respect to θ is achieved, we have $\frac{\partial}{\partial\theta}F(\theta,x;e)_{|\theta=\widehat{\theta}(x;e)}=0$ for every $x\in(E[L],L_{max})$. Therefore we have for $j=1,\ldots,m$

$$\left[\frac{\partial}{\partial e_j}\varphi(\theta;e)-\varphi(\theta;e)\theta\frac{\partial}{\partial e_j}\widehat{\text{VaR}}(\alpha;e)\right]_{|\theta=\widehat{\theta}}=0\,. \tag{7.10}$$

As $\varphi(\theta;e)=E\left[\exp\left(H\left(W,Z,\theta,e\right)\right)\right]$, we get by (7.7) and (3.10)

$$\frac{\partial}{\partial e_j}\varphi(\theta;e)=\frac{\theta}{e_j}E\left[\exp\left(H\left(W,Z,\theta,e\right)\right)\frac{\partial}{\partial\theta}\log H_j(W,Z,\theta,e_j)\right]\,.$$

On the other hand, by (7.9) we have

$$\varphi(\widehat{\theta}(\widehat{\text{VaR}}\,(\alpha;e)\,;e);e)=\alpha\exp\left(\widehat{\theta}(\widehat{\text{VaR}}\,(\alpha;e)\,;e)\widehat{\text{VaR}}\,(\alpha;e)\right)\,.$$

Substituting this in (7.10) and using the fact that $\widehat{\theta}>0$ we obtain the required result. $\square$

Corollary 7.3.3. *Denote by $\widehat{E}\,[\cdot]$ the expectation under the probability measure defined by*

$$d\widehat{P}\,(L<x)=\frac{\exp(\widehat{\theta}(\widehat{\text{VaR}}(\alpha;e);e)x)}{\varphi(\widehat{\theta}(\widehat{\text{VaR}}(\alpha;e));e)}dP\,(L<x)\,. \tag{7.11}$$

Then

$$\widehat{A}_j=\widehat{E}\,[L_j]\,,\quad j=1,\ldots,m\,.$$

Proof. By formula (1.2.2) in Jensen [82], $\widehat{E}[L] = \widehat{\text{VaR}}(\alpha; e)$. On the other hand,

$$\widehat{E}[L] = \sum_{j=1}^{m} e_j \widehat{E}[L_j],$$

and, therefore, $\widehat{E}[L_j] = \frac{\partial}{\partial e_j} \widehat{E}[L] = \widehat{A}_j$. $\qquad \square$

Remark 7.3.4. Risk contributions of this type have been suggested by Martin et al. [109] in the case of a one-factor Gaussian model ($W = 1$ a.s. and $p = 1$ in (3.5)). In this case it is possible to compute these quantities by simple numerical methods. Unfortunately this cannot be done for the general heavy-tailed model under consideration. As a remedy may serve the stochastoc approximation algorithm derived in Section 6.2. Note that the SLLN ensures, for $W^{(i)}, Z^{(i)}, i = 1, \ldots, n$, being i.i.d copies of W, Z, and $\widehat{\theta} = \widehat{\theta}(\widehat{\text{VaR}}(\alpha; e); e)$, that

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\exp(H(W^{(i)}, Z^{(i)}, \widehat{\theta}, e) - \widehat{\theta}\widehat{\text{VaR}}(\alpha; e))}{\alpha e_j} \frac{\partial}{\partial \theta} \log H_j(W^{(i)}, Z^{(i)}, \widehat{\theta}, e_j) \overset{\text{a.s.}}{\to} \widehat{A}_j, \ n \to \infty.$$

Therefore we obtain an estimate for the marginal risk contributions as a by-product from the recursion (6.13).

7.4   Numerical examples

7.4.1   Comparison of the methods

We consider a simple example in which we focus not on the absolute portfolio risk, but on the portfolio structure as represented by the marginal risk contributions. We are interested in whether the new method for estimation of marginal risk contributions as in (7.8) provides a good measure for the marginal risks in the portfolio.

We compare the risk contributions w.r.t. tail bound VaR with the ES(α)-contributions given by

$$\frac{\partial}{\partial e_j} \text{ES}(\alpha) = E[L_j \mid L \geq \text{VaR}(\alpha)], \ j = 1, \ldots, m, \tag{7.12}$$

Recall that for $L^{(i)}$ (and resp. $L_j^{(i)}, \ j = 1, \ldots, m$), $i = 1, \ldots, n$, being i.i.d copies of L (and resp. of $L_j, \ j = 1, \ldots, m$) as defined in (3.1), we have

$$\frac{1}{\#B_n} \sum_{i \in B_n} L_j^{(i)} \overset{\text{a.s.}}{\to} \frac{\partial}{\partial e_j} \text{ES}(\alpha), \ n \to \infty,$$

where $B_n = \left\{ i = 1, \ldots, n \ : \ L^{(i)} \geq \widetilde{\text{VaR}}(\alpha) \right\}$ is the set of simulations, where the portfolio loss exceeds the $1 - \alpha$ empirical quantile $\widetilde{\text{VaR}}(\alpha)$.

However, as $\#B_n$ increases slowly with the increase of n, extensive Monte Carlo simulation of all random components of L is necessary, see Overbeck [122], or Merino and Nyfeler [114].

Example 7.4.1. [Portfolio with concentration in 1 credit]
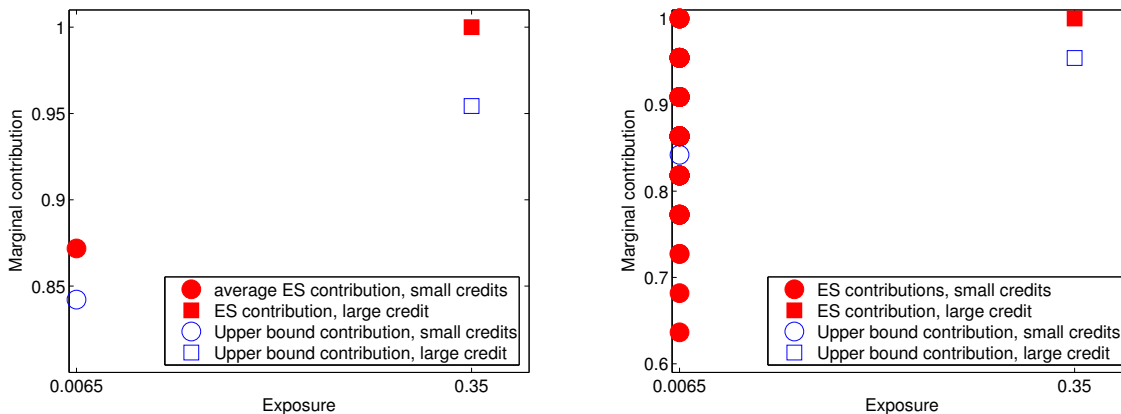The parameters of the considered model (as in (3.1) with (3.4) and (3.5)) are as follows:

Figure 7.1: The parameters are the same as in Example 7.4.1.
Left figure: the marginal risk contribution of the largest credit and the average marginal risk contribution of the small credits w.r.t the Expected Shortfall, as defined in (7.12), compared to the contributions w.r.t the upper bound approximation, as defined in (7.8). The results are similar in terms of distance between the small credits and the large one.
Right figure: the marginal risk contributions w.r.t the two risk measures $\mathrm{ES}(\alpha)$ and $\widehat{\mathrm{VaR}}(\alpha)$. We used 10 000 Monte Carlo simulations of all random components of L to compute the ES-contributions (7.12) and the same number of simulations of the common factors Z and global shock W to compute the upper bound contributions (7.8) as in Remark 7.3.4. Due to the error from the simulation, with the ES-method, equivalent credits have different contributions. The new upper bound method avoids this problem.

- $m = 101$ credits in the portfolio;
- exposures $e_1 = e_2 = \ldots = e_{m-1} = 0.0065$ and $e_m = 0.35$;
- rating system with $K = 2$ ratings (default and non-default);
- default probabilities $P(X_j = 1) = p_{j,1} = 0.02, \ j = 1, \ldots, m$;
- the marginal loss distributions are given as $L_j = \mathrm{I}_{\{X_j=1\}}, \ j = 1, \ldots, m$.

For the dependence structure we use in (3.5) the t-model with $\nu = 4$ degrees of freedom and one common factor $(p = 1)$ with factor loadings $\alpha_{j,1} = 0.8, \ j = 1, \ldots, m$.

The parameters in this example are selected in such a way that the portfolio is completely homogeneous, except for one of the credits, whose exposure is very large. A reasonable method for the computation of the marginal risk contributions should give equal contributions for all credits except the largest, which obviously contributes much more to the portfolio risk.

We fixed $1 - \alpha = 0.998$ and we computed the risk contributions w.r.t the upper bound approximation, as defined in (7.8), and w.r.t the Expected Shortfall (7.12). The overall portfolio risk measured by $ES(\alpha)$ is higher than the one measured by $\widehat{VaR}(\alpha)$, as $\widehat{VaR}(\alpha)$ is an approximation of $VaR(\alpha)$, which is by definition strictly smaller than $ES(\alpha)$. However, in Figure 7.1 (left) we observe that the ES-method and the upper bound method provide similar results in terms of the difference between the marginal risk contributions of the small credits and the large one. Note that the computation of the ES-contributions requires Monte Carlo simulation of all random components of L. Due to the error from such a simulation, equivalent credits appear to have different ES-contributions, see Figure 7.1 (right). This problem is avoided by the new upper bound approximation method, which uses Monte Carlo simulation only of the global shock W and the common factors Z, see Remark 7.3.4. □

7.4.2 Comparison of the models

Example 7.4.2. [Example 7.4.1 continued, CreditMetrics vs heavy tails]
We compare our heavy-tailed model as in Example 7.4.1 with a standard Gaussian model (i.e. $W = 1$ in (3.5)). For both models, we use the same marginal parameters as in Example 7.4.1. Further, in both cases we use $p = 1$ common factor with factor loadings $\alpha_{j,1} = 0.8$, $j = 1, \ldots, m$ in (3.5), i.e. the same correlation structure.
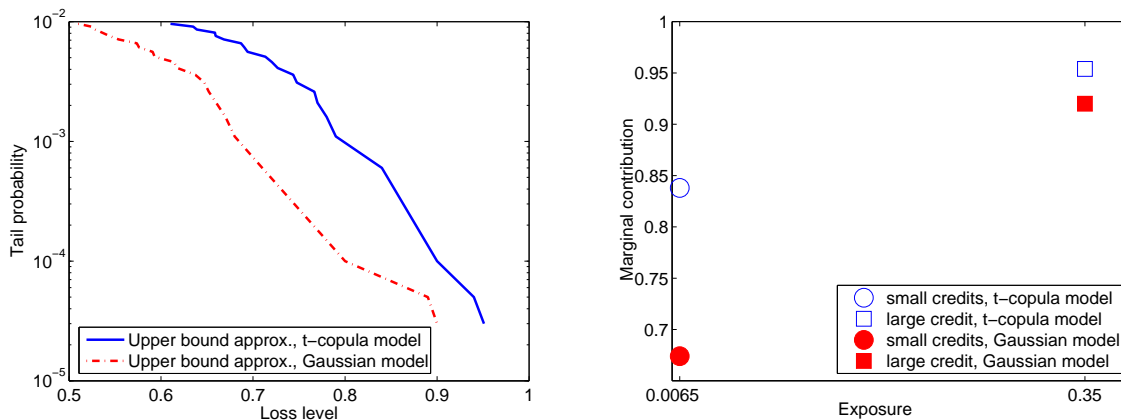

Figure 7.2: The parameters are given in Example 7.4.2.
Left figure: The upper bound approximation obtained by the new method for the Gaussian and for the t-copula models. We observe that in the t-copula model the losses are significantly higher.
Right figure: The marginal risk contributions obtained by the new method (7.8) for the Gaussian and for the t-copula models. The distance between the small credits and the largest one is larger for the Gaussian model, i.e. in this case the largest credit is relatively more risky. Since the overall portfolio risk measured by $\widehat{VaR}(\alpha)$ is higher for the t-copula model, in that case all credits are absolutely more risky than in the Gaussian case.

The upper bound approximation as in (6.8) can be applied also to the Gaussian model, since that is a special case ($W = 1$ a.s.) of our general model. Using the approximation we obtain $\widehat{\text{VaR}}(\alpha)$ and the corresponding marginal risk contributions (at level $1 - \alpha = 0.998$) for the two models. In Figure 7.2 (left) we observe that changing the dependence structure from Gaussian copula to t-copula has an important impact on the tail of the credit loss distribution, and hence on the overall portfolio risk. In view of Section 3.2, this result is not surprising; it confirms the results in e.g. Frey et. al [62]. Furthermore, we observe that the dependence model is important also for the portfolio structure. As demonstrated on Figure 7.2 right, under the Gaussian assumption, the largest credit is relatively (to the small credits) more risky than under the heavy-tailed assumption. One possible explanation for this behaviour could be that in the heavy-tailed model, large portfolio losses are caused typically by a global shock, which affects all credits simultaneously. In other words, our heavy-tailed model is less sensitive to portfolio concentrations. This may have important consequences in risk management, in particular for setting of exposure limits or for diversification analysis.   □

Chapter 8

Conclusions

In a global economy, it is no longer possible to explain the dependence between various financial assets entirely by modelling their linear correlation. In this thesis, we analysed a portfolio credit risk model with heavy-tailed risk factors to introduce non-linear (tail) dependence.

First, we looked at the model's input. With respect to the availability and the structure of the input data, we made similar or even weaker assumptions than the assumptions in the standard for the industry CreditMetrics model. We suggested new calibration methods which make use of the information contained in the joint extreme observations. The accuracy and the robustness of the estimators were investigated in simulations as well as in several real-data studies.

Second, we looked at the model's output. We developed (1) an importance sampling algorithm and (2) a semi-analytic approximation for the tail of the portfolio loss distribution. These methods allow, holding the computational time costs fixed, for tail and quantile estimation of the portfolio loss and for narrow confidence bounds beyond the range of the simulated by straightforward Monte Carlo methods scenarios. The tail approximation turned out to be quite useful in portfolio analysis, as it gives a by-product – estimates of the contributions of the marginal credits to the overall portfolio risk.

Throughout the thesis we compared various aspects of the heavy-tailed and the CreditMetrics model. We identified the key parameters which imply the difference in the tail behaviour of the credit portfolio loss in the two models. Not only does a heavy-tailed model lead to higher portfolio risk than the CreditMetrics model, and thus to e.g. higher risk capital requirements. It changes significantly the risk structure within a credit portfolio, and assuming it may lead to completely different results in an exposure or in a diversification analysis.

Bibliography

[1] Allen, L. and Saunders, A. (2002): *Credit Risk Measurement.* John Wiley & Sons, Inc., New York.

[2] Altman, E. (1989): Measuring corporate bond mortality and performance. *Journal of Finance* 9, 909-922.

[3] Altman, E. (1993): Defaulted bonds: demand, supply and performance 1987-1992. *Financial Analysts Journal* 6, 55-60.

[4] Altman, E. and Kishore, V. (1996): Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analysts Journal* 12, 57-64.

[5] Andersen, P., Borgan, O., Gill, R. and Keiding, N. (1993): *Statistical Models for Counting Processes.* Springer, New York.

[6] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999): Coherent measures of risk. *Mathematical Finance* 9, 203-228.

[7] Asmussen, S. (2000): *Ruin Probabilities.* World Scientific Piblishing, Singapore.

[8] Bank of International Settlements (1999): Credit risk modelling: current practices and applications. Basel Committee on Banking Supervision, Basel.

[9] Bank of International Settlements (2001): *The New Basel Capital Accord.* Basel Committee on Banking Supervision, Basel.

[10] Basrak, B. (2000): *The Sample Autocorrelation Function of Non-Linear Time Series.* Ph.D. thesis, Rijksuniversiteit Groningen.
Available at *www.ub.rug.nl/eldoc/dis/science/b.basrak/.*

[11] Becker, K., Isimbabi, M. Ogden, J. and Tucker, A. (1994): *Contemporary Portfolio Theory and Risk Management.* West Publishing, Minneapolis.

[12] Behar, R. and Nagpal, K. (1999): Dynamics of rating migration. Working paper, Standard & Poor's. Available at *www.standardandpoors.com.*

[13] Bielecki, T. and Rutkowski, M. (2001): *Credit Risk: Modelling, Valuation and Hedging*. Springer, Berlin.

[14] Bingham, N.H. Goldie, C.M. and Teugels, J.L. (1987): *Regular Variation*. Cambridge University press, Cambridge.

[15] Bluhm, C. (2003): CDO Modeling: techniques, examples and applications. Reseacrh paper, HypoVereinsbank Munich. Available at *www.defaultrisk.com*.

[16] Bluhm, C., Overbeck, L. and Wagner, C. (2003): *An Introduction to Credit Risk Modeling*. Chapman & Hall / CRC, Boca Raton.

[17] Breiman, L. (1965): On some limit theorems similar to the arc sin law. *Theory Probab. Appl.* 10, 323-331.

[18] Bünning, H. and Trenkel, G. (1978): *Non-parametric Statistical Methods*. Walter de Gruyter Verlag, Berlin.

[19] Cantor, R. and Hamilton, D. (2004): Rating migrations and defaults conditional on watchlist, outlook and rating history. Special comment, Moody's Investment Service. Available at *www.moodys.com*.

[20] Carty, L. and Fons, J. (1994): Measuring changes in corporate credit quality. *The Journal of Fixed Income* 6, 66-78.

[21] Casella, L. and Berger, R. (2002): *Statistical Inference*. Duxbury Press Pacific Grove, London.

[22] Coles, S., Heffernan, J. and Tawn, J. (1999): Dependence Measures for Extreme Value Analyses. *Extremes* 2, 339-365.

[23] Collett, D. (1999): *Modelling Binary Data*. Chapman & Hall, London.

[24] Cont, R. (2001): Empirical properties of asset returns. Stylized facts and statistical issues. *Quantitative Finance* 1, 223-236.

[25] Credit Suiss Financial Products (1997): *CreditRisk+ - a Credit Risk Management Framework*.

[26] Czado, C. (2000): Multivariate regression analysis of panel aata with binary outcomes applied to unemployment data. *Statistical Papers* 41, 281-304.

[27] Das, S., and Geng, G. (2002): Modeling the processes of correlated default. Working paper, Santa Clara University. Available at *www.defaultrisk.com*.

[28] Das, S., and Tufano, P. (1996): Pricing credit-sensitive debt when interest rates, credit ratings and credit spreads are stochastic. *The Journal of Financial Engineering* 5(2), 161-198.

[29] Das, S., Freed, L., Geng, G., Kapadia, N. (2002): Correlated default risk. Working paper, Santa Clara University. Available at *www.defaultrisk.com.*

[30] Daul, S., De Giorgi, E., Lindskog, F. and McNeil, A. (2003): The grouped t-copula with application to credit risk. *RISK* 16(11), 17-22.

[31] Davison, A. and Smith, R. (1990): Models for exeedances over high thresholds (with discussion). *J. R. Statistical Soc. Ser. B. Stat. Methodol.* 52, 393-442.

[32] de Haan, L. and Peng, L. (1998): Comparison of tail index estimators. *Statistica Neederlanica* 52, 60-70.

[33] Deheuvels, P. (1991): On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics and Probability Letters* 12, 429-439.

[34] Demarta S. and McNeil, A. (2004): The t-copula and related copulas. Preprint, ETHZ.

[35] Devlin, S., Gnanadesikan, R. and Kettenring, J. (1981): Robust estimation of dispersion matrices and principal components. *Journal of Amer. Statist. Assoc.* 76, 354-362.

[36] Draisma, G., Drees, H, Ferreira, A. and de Haan, L. (2004): Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* 10, 251-280.

[37] Duan, J., Gauthier, G. and Simonato, J. (2005): On the Equivalence of the KMV and Maximum Likelihood Methods for Structural Credit Risk Models. Working paper, University of Toronto. Available at *www.defaultrisk.com.*

[38] Duffie, D. (1999): Estimating the price of default risk. *Review of Financial Studies* 12, 197-226.

[39] Duffie, D. and Lando, D. (2001): Term structures of credit spreads with incomplete accounting information. *Econometrica* 69, 633-664.

[40] Duffie, D. and. Singleton, K. (2003): *Credit Risk: Pricing, Measurement and Management.* Princeton University Press, Princetion.

[41] Duffie, D., Pedersen, K. and. Singleton, K. (2003): Modelling sovereign yield spreads: a case study of Russian debt. *Journal of Finance* 52, 1287-1321.

[42] Duhamel, P. and Vetterli, M. (1990): Fast Fourier transforms: a tutorial review and a state of the art. *Signal Processing* 19, 259-299.

[43] Dupois, P. and Ellis, R. (1997): *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley, New York.

[44] Eberhart, A. and Sweeny, R. (1992): Does the bond market predict bankupcy settlements? *Journal of Finance* 3, 943-980.

[45] Embrechts, P., Hoeing, A. and Juri, A. (2003): Using Copulae to bound the Value-at-Risk for functions of dependent risks. *Finance and Stochastics* 7(2), 145-167.

[46] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997): *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[47] Embrechts, P., Lindskog, F. and McNeil, A. (2001): Modelling dependence with copulas and applications to risk management. In: *Handbook of Heavy Tailed Distributions in Finance,* 329-384, ed. S. Rachev, Elsevier.

[48] Embrechts, P., McNeil, A. and Straumann, D. (1999): Correlation and dependence in risk management: properties and pitfalls. In: *Risk Management: VaR and beyond,* Cambridge University press, Cambridge.

[49] Emmer, S. and Klüppelberg (1998): VaR - a measure for extreme risks. *Solutions* 2, 53-56.

[50] Fang, K.T., Kotz, S. and Ng, K.W. (1990): *Symmetric Multivariate and Related Distributions.* Chapman & Hall, London.

[51] Finger, C. (1998): Credit derivatives in CreditMetrics. *CreditMetrics Monitor* 3.

[52] Finger, C. (1999): Conditional approaches for CreditMetrics portfolio distribution. *CreditMetrics Monitor* 1.

[53] Finger, C. (1999): Risk-return reporting. *CreditMetrics Monitor* 1.

[54] Fishburn, P.C. (1977): Mean-risk analysis with risk associated with below-target returns. *American Economic Review* 67, 116-126.

[55] Follmer, H. and Leukert, P. (1999): Quantile hedging. *Finance and Stochastics* 3, 251-273.

[56] Fortin, I. and Kuzmics, C. (2002): Tail dependence in stock return pairs. *International Journal of Intelligent Systems in Accounting, Finance and Management* 11(2), 89-107.

[57] Frahm, G. and Junker, M. (2003): Generalized Elliptical Distributions: Models and Estimation. working paper, Research Center Caesar, Germany.

[58] Frahm, G., Junker, M. and Schmidt, R. (2004): Estimating tail dependence coefficient. Properties and pitfalls. *Insurance: Mathematics and Economics*, in print.

[59] Frey, R. and McNeil, A. (2001): Modelling dependent defaults. Preprint, ETH Zürich.

[60] Frey, R. and McNeil, A. (2002): VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights. *Journal of Banking and Finance* 26, 1317-1334.

[61] Frey, R. and McNeil, A. (2003): Dependent defaults in models of portfolio credit risk. *Journal of Risk* 6(1), 59-92.

[62] Frey, R., McNeil, A. and Nyfeler M. (2001): Copulas and credit models. *RISK* 8, 111-114.

[63] Frye, J. (2000): Collateral damage. *RISK*, Apr., 91-94.

[64] Frye, J. (2000): Collateral damage detected. Working paper, Federal Reserve Bank of Chicago.

[65] Genest, C., Ghoudi, K. and Rivest, L. (1995): A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrica* 82, 543-552.

[66] Gieseke, K. (2004): Correlated default with incomplete information. *Journal of Banking and Finance* 28, 1521-1545.

[67] Glasserman, P. (2003): Tail approximations for portfolio credit risk. Working paper, Columbia University.
Available at *www.gsb.columbia.edu/faculty/pglasserman*.

[68] Glasserman, P., Heidelberger, P. and Shahabuddin, P. (1999): Asymptotically optimal importance sampling for pricing path-dependent options. *Mathematical Finance* 9, 117-152.

[69] Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2002): Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance* 12, 239-270.

[70] Glasserman, P. and Li, J. (2004): Importance sampling for portfolio credit risk. Working paper, Columbia University.
Available at *www.gsb.columbia.edu/faculty/pglasserman*.

[71] Gordy, M. (2000): A comparative anatomy of credit risk models. *Journal of Banking and Finance* 24, 119-149.

[72] Gordy, M. (2001): A risk factor model foundation for rating-based bank capital rules. *Journal of Financial Intermediation* 12, 199-232.

[73] Gordy, M. (2002): Saddlepoint approximation of CreditRisk+. *Journal of Banking and Finance* 26, 1335-1353.

[74] Gupton, G., Bhatia, M. and Finger, C. (1997): *CreditMetrics – Technical document.* J.P.Morgan, New York.

[75] Gupton, G. and Stein, R. (2005): LossCalc v2: Dynamic prediction of LGD. Research paper, KMV. Available at *www.defaultrisk.com*.

[76] Hamilton, D., Gupton, G. and Berthault, A. (2001): Default and recovery rates of corporate bond issuers: 2000. Special comment, Moody's Investment Service. Available at *www.moodys.com*.

[77] Höffding, W. (1947): On the distribution of the rank correlation coefficient τ whern the variates are not independent. *Biometrica* 34, 183-196.

[78] Hult, H. and Lindskog, F. (2001): Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability* 34(3), 587-608.

[79] Israel, R., Rosenthal, J. and Wei, J. (2001): Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11(2), 245-265.

[80] Jarrow, R. and Turnbull, S. (1995): Pricing derivatives on financial securities subject to credit risk. *Journal of Finance* 50, 53-85.

[81] Jarrow, R., Lando, D. and Turnbull, S. (1997): A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* 10, 481-523.

[82] Jensen, J. (1995): *Saddlepoint Approximations.* Oxford University press, Oxford.

[83] Joe, H. (1997): *Multivariate Models and Dependence Concepts.* Chapman & Hall, New York.

[84] Johnson, N. and Kotz, S. (1969): *Discrete Distributions.* John Wiley & Sons, New York.

[85] Junker, M. and May, A. (2002): Measurement of aggregate risk with copulas. Preprint, Research Center CAESAR.

[86] Kallenberg, O. (1983): *Random Measures.* Akademie-Verlag, Berlin.

[87] Kavvathas, D. (2001): Estimating credit rating migration probabilities for corporate bonds. Working paper, University of Chicago. Available at *www.ssrn.com/fen.*

[88] Kealhofer, S. (1995): Managing default risk in portfolios of derivatives. In *Derivative Credit Risk, Advances in Measurement and Management*, 49-66. Risk publications, London.

[89] Kealhofer, S. (2003): Quantifying credit risk I: default prediction. Research paper, Moody's KMV. Available at *www.defaultrisk.com*

[90] Kealhofer, S. (2003): Quantifying credit risk II: default prediction. Research paper, Moody's KMV. Available at *www.defaultrisk.com*

[91] Klüppelberg, C., Kuhn, G. and Liang, P. (2005): Estimating tail Dependence of elliptical distributions. Preprint, Munich University of Technology.
Available at *www.ma.tum.de/stat/papers.*

[92] Kostadinov, K. (2005): Non-parametric estimation of elliptical copulae with application to credit risk. Preprint, Munich University of Technology.
Available at *www.ma.tum.de/stat/papers.*

[93] Kostadinov, K. (2006): Tail approximation for credit risk portfolios with heavy-tailed risk factors. *Journal of Risk* 8, 156-173.

[94] Kuhn, G. (2004): Tails of credit default portfolios. Preprint, Munich University of Technology. Available at *www.ma.tum.de/stat/papers.*

[95] Kreinin, A. and Sidelnikova, M. (2001): Regularization algorithms for transition matrices. In *Enterprise Credit Risk*, ed. by Aguais, Rosen and Smithson, 43-60. Algorithmics publications, New York.

[96] Kushner, H. and Yin, G. (1997): *Stochastic Approximation Algorithms and Applications.* Springer, New York.

[97] Lando, D. (1998): Cox processes and credit risky securities. *Review of Derivatives Research* 2, 99-120.

[98] Lando, D. and Skodeberg, T. (2000): Analyzing rating migrations and rating drift with continuous observations. *Journal of Banking and Finance* 26, 423-444.

[99] Ledford, A. and Tawn, J. (1996): Statistics for near independence in multivariate extreme values. *Biometrica* 83, 169-187.

[100] Leland, H.E. (1994): Corporate debt value, bond covenants and optimal capital structure. *Journal of Finance* 49, 1213-1252.

[101] Leland, H.E. and Toft, K. (1996): Optimal capital structure, endogeneous bankruptcy and the term structure of credit spreads. *Journal of Finance* 51, 987-1019.

[102] Li, D.X. (2000): On default correlation: a copula function approach. *The Journal of Fixed Income* 3, 43-54.

[103] Li, H. (2000): A model for pricing defaultable bonds and credit ratings. Working paper, Washington University.

[104] Lindskog, F., McNeil, A. and Schmock, U. (2003): Kendall's tau for elliptical distributions. In *Credit Risk – Measurement, Evaluation and Management*, edited by Bol, Nakhaeizadeh, Rachev, Ridder and Vollmer, Physica-Verlag, Heidelberg.

[105] Longstaff, F.A. and Schwartz, E.S. (1995): A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance* 50, 789-819.

[106] Malevergne, Y. and Sornette, D. (2004): Tail dependence of factor models. *Journal of Risk* 6, 71-116.

[107] Mandelbrot, B. (1963): The variation of certain speculative prices. *Journal of Business* 36, 394-419.

[108] Markowitz, H. (1952): Portfolio selection. *Journal of Finance* 7, 77-91.

[109] Martin, R., Thompson, K. and Browne, C. (2001): Taking to the saddle. *RISK* 14, 91-94.

[110] The MathWorks (2005): *Matlab User Guide*. Available at *www.mathworks.com*.

[111] Matten. C. (1996): *Managing Bank Capital*. Wiley, Chichester.

[112] McCulloch, J.H. (1971): Measuring the term structure of interest rates. *Journal of Business* 44, 19-31.

[113] Mcleish, D. and Small, C. (1988): *The Theory and Applications of Statistical Inference Functions*, vol. 44 of *Lecture Notes in Statistics*, Springer, New York.

[114] Merino, S. and Nyfeler, M. (2004): Applying importance sampling for estimating coherent risk contributions. *Quantitative Finance* 4, 199-207.

[115] Merton, R (1974): On the pricing of corporate debt. *Journal of Finance* 29, 449-470.

[116] Mikosch, T. (2001): Modelling dependence and tails of financial time series. In *SemStat: Seminaire Europeen de Statistique, Extreme Value Theory and Applications*, ed. by H. Rootzen and B. Finkenstadt, Chapman and Hall, London.

[117] Mina, J. and Ulmer, A. (1999): Delta-gamma four ways. Working paper, RiskMetrics group. Available at *www.riskmetrics.com*.

[118] Nelsen, R. (1999): *An Introduction to Copulas.* Springer, New York.

[119] Nelson, R. and Siegel, A. (1987): Parsimonious modelling of yield curves. *Journal of Business* 60, 473-499.

[120] Nickell, P., Perraudin, W. and Varotto, S. (2000): Stability of rating migrations. *Journal of Banking and Finance* 24, 203-228.

[121] Oda, N. and Muranaga, J. (1997): A new framework for measuring credit risk of a portfolio: The 'ExVaR' model. *Monetary and Economic Studies* 15, Bank of Japan.

[122] Overbeck, L. (2000): Allocation of economic capital in loan portfolios. In *Proceedings 'Measuring Risk In Complex Stochastic Systems'*, Springer, Berlin.

[123] Petrov, V. (1995): *Limit Theorems of Probability Theory.* Oxford Univ. Press, New York.

[124] Resnick, S.I, (1987): *Extreme Values, Regular Variation and Point Processes.* Springer, New York.

[125] Rousseeuw, P. and Molenberghs, G. (1993): Transformations of non positive semidefinite correlation matrices. *Comm. Statist. A – Theory and Methods* 4, 965-984.

[126] Schmidt, R. (2002): Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research* 55, 301-327.

[127] Schmidt, R. and Stadtmüller, U. (2003): Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, in print.

[128] Schlather, M. and Tawn, J.A. (2003): A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrica* 90 (1), 139-156.

[129] Schönbucher, P. (2003): *Credit Derivatives Pricing Models*. John Wiley & Sons, Chichester, England.

[130] Schönbucher, P. and Schubert, D. (2001): Copula dependent default risk in intensity models. Working paper, Bonn University.

[131] Schwarz, C. (2005): Eine gruppierte elliptische Copula und ihre Anwendung im Kreditrisikomanagement. Diploma thesis, Munich University of Technology.

[132] Smith, R.L (1987): Estimating tails of probability distributions. *Annals of Statistics* 15, 1174-1207.

[133] Sklar, A. (1996): Random variables, distribution functions, and copulas – a personal look backward and forward. In *Distributions with Fixed Marginals and Related Topics* 1-14, ed. by L. Rüschendorff, B. Schweizer and M. Taylor, Institute of Mathematical Statistics, Hayward, CA.

[134] Tasche, D. (2000): Risk contributions and performance measurement. Working paper, TU Munchen. Available at *www.ma.tum.de/stat*.

[135] Wendin, J. (2002): *Estimation of the Spectral Measure and the Tail Dependence Coefficient for Regularly Varying Random Vectors*. Ph.D. thesis, ETH Zürich. Available at *www.math.ethz.ch/ wendin*.

[136] Zhou, C. (1997): A jump-diffusion approach to modelling credit risk and valuing defaultable secuities. Discussion paper, Board of Governors of the Federal Reserve System. Available at *www.defaultrisk.com*.